ÉCOLE DE PHYSIQUE DES HOUCHES – UJF & INPG – GRENOBLE

a NATO Advanced Study Institute

LES HOUCHES

SESSION LXXVI

30 July – 31 August 2001

Unity from duality: Gravity, gauge theory and strings

L'unité de la physique fondamentale : gravité, théorie de jauge et cordes

Edited by

C. BACHAS, A. BILAL, M. DOUGLAS,
N. NEKRASOV and F. DAVID

EDP SCIENCES

Les Ulis, Paris, Cambridge

Springer

Berlin, Heidelberg, New York, Hong Kong, London, Milan, Paris, Tokyo

Published in cooperation with the NATO Scientific Affair Division

ISSN 0924-8099 print edition
ISSN 1610-3459 online edition

ISBN 3-540-00276-6 Springer-Verlag Berlin Heidelberg New York
ISBN 2-86883-625-9 EDP Sciences Les Ulis

Printed in France

Preface

The 76th session of the Les Houches Summer School in Theoretical Physics was devoted to recent developments in string theory, gauge theories and quantum gravity.

As frequently stated, Superstring Theory is the leading candidate for a unified theory of all fundamental physical forces and elementary particles. This claim, and the wish to reconcile general relativity and quantum mechanics, have provided the main impetus for the development of the theory over the past two decades. More recently the discovery of dualities, and of important new tools such as D-branes, has greatly reinforced this point of view. On the one hand there is now good reason to believe that the underlying theory is unique. On the other hand, we have for the first time working (though unrealistic) microscopic models of black hole mechanics. Furthermore, these recent developments have lead to new ideas about compactification and the emergence of low-energy physics.

While pursuing the goal of unification we have also witnessed a dramatic return to the "historic origins" of string theory as a dual model for meson physics. Indeed, the study of stringy black branes has uncovered a surprising relation between string theory and large-N gauge dynamics. This was cristallized in the AdS/CFT correspondence, which has revived the old hope for a string description of the strong interaction. The AdS/CFT correspondence is moreover a prime illustration of the central role of string theory in modern theoretical physics. Much like quantum field theory in the past, it provides a fertile springboard for new tools, concepts and insights, which should have ramifications in wider areas of physics and mathematics.

The main lectures of the Les Houches school covered most of the recent developments, in a distilled and pedagogical fashion. Students were expected to have a good knowledge of quantum field theory, and of basic string theory at the level, for instance, of the first ten chapters of Green, Schwarz and Witten. The emphasis was on acquiring a working knowledge of advanced string theory in its present form, and on critically assessing open problems and future directions.

The lectures by Bernard de Wit were a comprehensive introduction to supergravities in different dimensions and with various numbers of supersymmetries. Topics covered include the allowed low-energy couplings, duality symmetries, compactifications and supersymmetry in curved backgrounds.

Part of this is older material not easily accessible in the literature, and presented here from a modern perspective.

Eliezer Rabinovici lectured on supersymmetric gauge theories, reviewing earlier and more recent results for $N = 1$, 2 and 4 supersymmetries in four dimensions. These results include the structure of the effective lagrangians, non-renormalization theorems, dualities, the celebrated Seiberg-Witten solution and brane engineering of effective gauge theories.

M-theory and string dualities were introduced in the lectures by Ashoke Sen. He reviewed the conjectured relations between the five perturbative string theories, the maximal $N = 1$ supergravity in eleven dimensions and their compactifications. He summarized our present-day knowledge of the still elusive fundamental or "M theory", from which the above theories derive as special limits. More recent topics include non-BPS branes, where duality is of limited (but not zero) use.

Philip Candelas gave a pedagogical introduction to the important subject of Calabi Yau compactifications. He first reviewed the older material, and then discussed more recent aspects, including second quantized mirror symmetry, conifold transitions and some intriguing relations to number theory. Unfortunately a written version of his lectures could not be included in this volume.

The holographic gauge/string theory correspondence was the subject of the lectures by Juan Maldacena and by Igor Klebanov. Maldacena introduced the conjectured equivalence between string theory in the near-horizon geometries of various black branes and gauge theories in the large $N_{\rm color}$ limit. He focused on the celebrated example of $N = 4$ four-dimensional super Yang Mills dual to string theory in $AdS_5 \times S_5$, and gave a critical review of the existing evidence for this correspondence. He also discussed analogous conjectures in other spacetime dimensions, in particular those relevant to the study of stringy black holes, and of the still elusive little string theory.

Igor Klebanov then concentrated on this duality in the phenomenologically more interesting contexts of certain $N = 1$ and 2 supersymmetric gauge theories in four dimensions. He reviewed the relevant geometries on the supergravity side, which include non-trivial fluxes and fractional branes, and discussed the gravity duals of renormalization group flow, confinement and chiral symmetry breaking. These results have revived and made sharper the old ideas about the "master field" of large N gauge theory.

The lectures of Michael Green dealt with some finer aspects of string dualities and of the gauge theory/string theory correspondence. He discussed higher derivative couplings in effective supergravity actions, focusing in particular on the contributions of instantons both in string theory and on the

Yang Mills side. His lectures also included some introductory material on D-branes. Unfortunately a written version of his lectures could not be included in this volume.

Andrew Strominger gave a detailed introduction to quantum gravity in a de Sitter spacetime. He discussed in particular whether ideas of holography, that have worked well in anti de Sitter, could also be applied in this case. This was one of the more speculative subjects in the school, but a fascinating one not the least because astrophysical observations seem to indicate that we actually live in an accelerating universe.

Finally Michael Douglas gave three lectures on D-brane geometry, and in particular on the problem of classifying all $N = 1$ string-theory vacua, while Alexander Gorsky discussed $N = 1$ and $N = 2$ supersymmetric gauge theories and their relation to integrable models. Nikita Nekrasov lectured on open strings and non-commutative gauge theories.

Some more advanced and/or topical subjects were covered in the accompanying series of seminars. Seminar speakers included Laurent Baulieu, Mirjam Cvetič, Frank Ferrari, Dan Freedman, Bernard Julia, Peter Mayr, Soo-Jong Rey, Augusto Sagnotti, Samson Shatashvili, and one of the organizers (C.B.). There was also a lively weekly student seminar and discussion sessions, which contributed greatly to the lively and stimulating atmosphere of the school. Some of the seminar speakers have kindly accepted to contribute to the present volume.

In the year that has elapsed since the end of the school there have been further developments in the subject. The pp-waves, which arise as Penrose limits of near-horizon geometries, offer for instance a new line of attack on the important problem of solving string theory in Ramond-Ramond backgrounds. Such developments and others will no doubt make, one day, the present volume obsolete. This is of course no reason for regret – to the contrary we hope that this may happen sooner rather than later, and that the participants of this school will help shape the (non-recognisable?) future form of M theory.

Acknowledgements

This Les Houches school and the present proceedings were made possible due to the contributions of many individuals and several different funding sources.

Funding sources included the NATO Scientific Affairs Division through its Advanced Study Institute program, the European Union through its High-Level Scientific Conferences program, the Commissariat à l'Énergie Atomique, the Université Joseph Fourier de Grenoble, and two European

IHP Networks: "Superstring Theory" (HPRN-CT-2000-00122) and "The Quantum Structure of Spacetime" (HPRN-CT-2000-00131).

Among the many people contributing to the success of the school, we should mention

- the board of the School and in particular François David, who has worked tirelessly at all different stages (funding applications, admissions, running the session, preparation of proceedings) exceeding often the organizers in zeal and energy;

- the secretaries Mmes G. D'Henry, I. Lelièvre and B. Rousset (and the other personnel of the school), who helped solve administrative and everyday problems;
  and last but not least

- the lecturers, for their efforts in presenting hard material in a clear and pedagogical fashion, and also for writing up their lecture notes.

Costas Bachas
Adel Bilal
Michael Douglas
Nikita Nekrasov
François David

CONTENTS

Seminar 2. Special Holonomy Spaces and M-Theory by M. Cvetič, G.W. Gibbons, H. Lü and C.N. Pope 523

Seminar 3. Four Dimensional Non-Critical Strings by F. Ferrari 547

ORGANIZERS

C. BACHAS, Laboratoire de Physique Théorique, ENS, 24 rue Lhomond, 75231 Paris, France

A. BILAL, Institut de Physique, Université de Neuchâtel, 1 rue Breguet, 2000 Neuchâtel, Switzerland

M. DOUGLAS, Department of Physics & Astronomy, Rutgers, The State University of New Jersey, Piscataway, NJ 08854-8019, U.S.A.

N. NEKRASOV, I.H.E.S, 35 route de Chartres, 91440 Bures-sur-Yvette, France

F. DAVID, SPhT, CEA Saclay, 91191 Gif-sur-Yvette, France

LECTURERS

P. CANDELAS, Mathematical Institute, Oxford University, 24-29 St. Giles, Oxford OX1 3LB , U.K.

M. GREEN, DAMPT, Wilberforce Road, Cambridge CB3 OWA, U.K.

I. KLEBANOV, Joseph Henry Laboratories, Princeton University, Princeton, NJ 08544, U.S.A.

J. MALDACENA, Jefferson Physical Laboratory, Harvard University, Cambridge, MA 02138, U.S.A.

E. RABINOVICI, Racah Institute of Physics, Hebrew University, 91904 Jerusalem, Israel

A. SEN, Harish-Chandra Research Institute, Chhatnag Road, Jhusi, 211019 Allahabad, India

A. STROMINGER, Jefferson Physical Laboratory, Harvard University, Cambridge, MA 02138, U.S.A.

B. DE WIT, Institute for Theoretical Physics, Spinoza Institute, Utrecht University, 35P4 CE Utrecht, The Netherlands

SEMINAR SPEAKERS

L. BAULIEU, LPTHE, Université Pierre et Marie Curie, Tour 16, 4 place Jussieu, 75231 Paris Cedex 05, France

M. CVETIC, Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, PA 19104, U.S.A.

D. FREEDMAN, Center for Theoretical Physics, MIT, Cambridge, MA 02139, U.S.A.

A. GORSKY, ITEP, B. Cheremushkinskaya 25, 117259 Moscow, Russia

B. JULIA, LPT/ENS, 24 rue Lhomond, 75231 Paris, France

P. MAYR, CERN Theory Division, 1211 Genova 23, Switzerland

S. REY, Center for Theoretical Physics, Seoul National University, 151-747 Seoul, Corea

A. SAGNOTTI, Department of Physics, University Roma II, Tor Vergata, Via della Ricerca Scientifica 1, 00133 Roma, Italy

S. SHATASHVILI, Department of Physics, Yale University, New Haven, CT 06520, U.S.A.

PARTICIPANTS

D. ABANIN, ITEP, B. Cheremushkinskaya ul 25, 117259 Moscow, Russia

A. ALEXANDROV, ITEP, B. Cheremushkinskaya 25, 117259 Moscow, Russia

D. BELOV, Steklov Mathematical Institute, Gubkin St. 8, GSP-1, Moscow 117966, Russia

I. BENA, Department of Physics, University of California, Santa Barbara, CA 93106, U.S.A.

D. BERMAN, Racah Institute of Physics, Hebrew University, Jerusalem 91904, Israel

V. BRAUN, Humboldt University, Physik QFT, Invalidenstr. 110, 10115 Berlin, Germany

L. CARLEVARO, Université de Neuchâtel, Institut de Physique, Département de Physique Théorique, Rue Breguet 1, 2002 Neuchâtel, Switzerland

N. COUCHOUD, LPTHE, Université Pierre et Marie Curie, Tour 16, 4 place Jussieu, 75252 Paris Cedex 05, France

G. D'APPOLLONIO, Dipartimento di Fisica, Sezione INFN, L.go Fermi 2, 50125 Firenze, Italy

V. DOLGUSHEV, Tomsk State University, Physics Department, Lenin Av. 36, Tomsk 63050, Russia

R. DUIVENVOORDEN, Instituut voor Theoretische Fysica, University of Amsterdam, Valckenierstraat 65, 1018 XE Amsterdam, The Netherlands

A. DYMARSKY, Moscow State University, Physics Faculty, Theoretical Physics department, Vorobevy Gory, Moscow 119899, Russia

F. FERRARI, Joseph Henry Laboratories, Jadwin Hall, Princeton University, Princeton, NJ 08540, U.S.A.

B. FLOREA, Mathematical Institute, University of Oxford, 24-29 St. Giles, Oxford OX1 3LB, U.K.

U. GRAN, Department of Theoretical Physics, Chalmers University of Technology, Göteborg University, 41296 Gothenburg, Sweden

S. GUKOV, Jefferson Physical Laboratories, Harvard University, Cambridge, MA 02138, U.S.A.

S. GURRIERI, Centre de Physique Théorique Luminy, Case 907, 13288 Marseille Cedex 9, France

M. HAACK, Martin Luther Universität Wittenberg, Fachbereich Physik, Friedmann Bach Platz 6, 06108 Halle, Germany

P. HENRY, LPT, ENS, 24 rue Lhomond, 75005 Paris, France

C. HERZOG, Princeton University, Physics Department, Princeton, NJ 08544, U.S.A.

V. HUBENY, Department of Physics, 382 via Pueblo Mall, Stanford University, Stanford, CA 94305-4060, U.S.A.

P. KASTE, CPHT, École Polytechnique, 91128 Palaiseau, France

S. KLEVTSOV, Moscow State University, Physics Faculty, Theoretical Physics Department, Vorobevy Gory, Moscow 119899, Russia

A. KONECHNY, University of California Berkeley, Theoretical Physics Group, Mail Stop 50A-501 4BNL, Berkeley, CA 94720, U.S.A.

I. LOW, Department of Physics, Harvard University, Cambridge, MA 02138, U.S.A.

D. MALYSHEV, Moscow State University, Physics Faculty, Quantum Statistics and Quantum Field Theory Department, Vorobevy Gory, Moscow 119899, Russia

I. MASINA, SPhT, CEA Saclay, Orme des Merisiers, bâtiment 774, 91191 Gif-sur-Yvette, France

L. McALLISTER, Stanford University, Department of Physics, 382 Via Pueblo Mall, Stanford, CA 94305, U.S.A.

S. MORIYAMA, Department of Physics, Kyoto University, Kitashirakawa-Oiwakecho, Sakyo-ku, Kyoto 606-8502, Japan

V.S. NEMANI, Tata Institute of Fundamental Research, Department of Theoretical Physics, Homi Bhabha Road, Colaba, Mumbai 400005, India

F. NITTI, New York University, Department of Physics, 4 Washington Place, New York, NY 10003, U.S.A.

D. NOGRADI, University of Leiden, Institute Lorentz, P.O. Box 9506, 2300 RA Leiden, The Netherlands

L. PANDO ZAYAS, Michigan Center for Theoretical Physics, Randall Laboratory of Physics, The University of Michigan, Ann Arbor, MI 48109-1120, U.S.A.

G. PANOTOPOULOS, University of Crete, Physics Department, P.O. Box 2208, Heraklion, Greece

J. PARK, California Institute of Technology, 452-48 Cal Tech, 1200 East California Blvd., Pasadena, CA 91125, U.S.A.

V. PESTUN, ITEP, B. Cheremushkinskaya 25, 117259 Moscow, Russia

R. RABADAN, Universidad Autonoma de Madrid, Departamento De Fisica Teorica C-XI, 28049 Madrid, Spain

S. RIBAULT, Centre de Physique Théorique, École Polytechnique, 91128 Palaiseau, France

F. RICCIONI, Dipartimento di Fisica, Universita di Roma “Tor Vergata”, via della Ricerca Scientifica 1, 00133 Roma, Italy

D. ROEST, Institute for Theoretical Physics, Nijenborgh 4, 9747 AG Groningen, The Netherlands

C. ROMELSBERGER, Department of Physics & Astronomy, Rutgers, The State University of New Jersey, 136 Frelinghuisen Road, Piscataway, NJ 08854-8019, U.S.A.

V. RYCHKOV, Princeton University, Mathematics Department, Fine Hall, Washington Road, Princeton, NJ 08544, U.S.A.

K. SARAIKIN, ITEP, B. Cheremushkinskaya 25, 117259 Moscow, Russia

S. SCHÄFER-NAMEKI, DAMTP, Centre for Mathematical Sciences, University of Cambridge, Wilberforce Road, Cambridge CB3 0WA, U.K.

M. SCHNABL, SISSA, Via Beirut 4, 34014 Trieste, Italy

G. SERVANT, SPhT, CEA Saclay, 91191 Gif-sur-Yvette, France

A. SHCHERBAKOV, Theoretical Physics Department, Dnepropetrovsk National University, Naukova St. 13, Dnepropetrovsk 49050, Ukraine

M. SMEDBÄCK, Department of Theoretical Physics, Uppsala University, Box 803, 751 08 Uppsala, Sweden

A. SOLOVYOV, Institute for Theoretical Physics, 14 B Metrologicheskaya St., Kiev, Ukraine

M. SPRADLIN, Department of Physics, Harvard University, 17A Oxford Street, Cambridge, MA 02138, U.S.A.

C. STAHN, Department of Applied Mathematics and Theoretical Physics, Cambridge CB3 OWA, U.K.

S. ULANOV, Theoretical Physics Department, Dnepropetrovsk National University, Naukova St. 13, Dnepropetrovsk 49050, Ukraine

D. VASSILIEV, MIPT, ITEP, St. B. Cheremushkinskaya 25, 117259 Moscow, Russia

A. VOLOVICH, Physics Department, Harvard University, Cambridge, MA 02138, U.S.A.

J. WALCHER, ETH, Zurich, Institute for Theoretical Physics, 8093 Zurich, Switzerland

J.-T. YEE, School of Physics, Seoul National University, Seoul 151-742, South Korea

A. ZOTOV, ITEP, Ul B. Cheremushkinskaya 25, 117259 Moscow, Russia

LECTURE 1

SUPERGRAVITY

B. DE WIT

Institute for Theoretical Physics &
Spinoza Institute, Utrecht University,
The Netherlands

Contents

SUPERGRAVITY

B. de Wit

1 Introduction

Supergravity plays a prominent role in our ideas about the unification of fundamental forces beyond the standard model, in our understanding of many central features of superstring theory, and in recent developments of the conceptual basis of quantum field theory and quantum gravity. The advances made have found their place in many reviews and textbooks (see, [1]), but the subject has grown so much and has so many different facets that no comprehensive treatment is available as of today. Also in these lectures, which will cover a number of basic aspects of supergravity, many topics will be left untouched.

During its historical development the perspective of supergravity has changed. Originally it was envisaged as an elementary field theory which should be free of ultraviolet divergencies and thus bring about the long awaited unification of gravity with the other fundamental forces in nature. But nowadays supergravity is primarily viewed as an effective field theory describing the low-mass degrees of freedom of a more fundamental underlying theory. The only candidate for such a theory is superstring theory (for some reviews and textbooks, see, [2]), or rather, yet another, somewhat hypothetical, theory, called M-theory. Although we know a lot about M-theory, its underlying principles have only partly been established. String theory and supergravity in their modern incarnations now represent some of the many faces of M-theory. String theory is no longer a theory exclusively of strings but includes other extended objects that emerge in the supergravity context as solitonic objects. Looking backwards it becomes clear that there are many reasons why neither superstrings nor supergravity could account

I am grateful to Sergio Ferrara, Murat Günaydin, Francisco Morales, Jan Louis, Hermann Nicolai, Peter van Nieuwenhuizen, Soo-Jong Rey, Henning Samtleben, Ergin Sezgin, Kostas Skenderis, Mario Trigiante, Stefan Vandoren and Toine Van Proeyen for many helpful and stimulating discussions.

for all the relevant degrees of freedom and we have learned to appreciate that M-theory has many different realizations.

Because supersymmetry is such a powerful symmetry it plays a central role in almost all these developments. It controls the dynamics and, because of nonrenormalization theorems, precise predictions can be made in many instances, often relating strong- to weak-coupling regimes. To appreciate the implications of supersymmetry, Section 2 starts with a detailed discussion of supersymmetry and its representations. Subsequently supergravity theories are introduced in Section 3, mostly concentrating on the maximally supersymmetric cases. In Section 4 gauged nonlinear sigma models with homogeneous target spaces are introduced, paving the way for the construction of gauged supergravity. This construction is explained in Section 5, where the emphasis is on gauged supergravity with 32 supercharges in 4 and 5 spacetime dimensions. These theories can describe anti-de Sitter ground states which are fully supersymmetric. This is one of the motivations for considering anti-de Sitter supersymmetry and the representations of the anti-de Sitter group in Section 6. Section 7 contains a short introduction to superconformal transformations and superconformally invariant theories. This section is self-contained, but it is of course related to the discussion in Section 6 on anti-de Sitter representations as well as to the adS/CFT correspondence.

This school offers a large number of lectures dealing with gravity, gauge theories and string theory from various perspectives. We intend to stay within the supergravity perspective and to try and indicate what the possible implications of supersymmetry and supergravity are for these subjects. Our hope is that the material presented below will offer a helpful introduction to and will blend in naturally with the material presented in other lectures.

2 Supersymmetry in various dimensions

An enormous amount of information about supersymmetric theories is contained in the structure of the underlying representations of the supersymmetry algebra (for some references, see [1, 3–6]). Here we should make a distinction between a supermultiplet of fields which transform irreducibly under the supersymmetry transformations, and a supermultiplet of states described by a supersymmetric theory. In this section[1] we concentrate on supermultiplets of states, primarily restricting ourselves to flat Minkowski spacetimes of dimension D. The relevant symmetries in this case form an

[1]The material presented in this and the following section is an extension of the second section of [6].

extension of the Poincaré transformations, which consist of translations and Lorentz transformations. However, many of the concepts that we introduce will also play a role in the discussion of other superalgebras, such as the anti-de Sitter (or conformal) superalgebras. For a recent practical introduction to superalgebras, see [7].

2.1 The Poincaré supersymmetry algebra

The generators of the super-Poincaré algebra comprise the supercharges, transforming as spinors under the Lorentz group, the energy and momentum operators, the generators of the Lorentz group, and possibly additional generators that commute with the supercharges. For the moment we ignore these additional charges, often called [illegible] charges[2]. There are other relevant superalgebras, such as the supersymmetric extensions of the anti-de Sitter (or the conformal) algebras. These will be encountered in due course.

The most important anti-commutation relation of the super-Poincaré algebra is the one of two supercharges,

$$\{Q_\alpha, \bar{Q}_\beta\} = -2iP_\mu \, (\Gamma^\mu)_{\alpha\beta} \,, \tag{2.1}$$

where we suppressed the central charges. Here Γ^μ are the gamma matrices that generate the Clifford algebra $\mathcal{C}(D-1,1)$ with Minkowskian metric $\eta_{\mu\nu} = \mathrm{diag}(-,+,\cdots,+)$.

The size of a supermultiplet depends exponentially on the number of independent supercharge components Q. The first step is therefore to determine Q for any given number of spacetime dimensions D. The result is summarized in Table 1. As shown, there exist five different sequences of spinors, corresponding to spacetimes of particular dimensions. When this dimension is odd, it is possible in certain cases to have Majorana spinors. These cases constitute the first sequence. The second one corresponds to those odd dimensions where Majorana spinors do not exist. The spinors are then Dirac spinors. In even dimension one may distinguish three sequences. In the first one, where the number of dimensions is a multiple of 4, charge conjugation relates positive- with negative-chirality spinors. All spinors in this sequence can be restricted to Majorana spinors. For the remaining two sequences, charge conjugation preserves the chirality of the spinor. Now there are again two possibilities, depending on whether Majorana spinors

[2] The terminology adopted in the literature is not always very precise. Usually, all charges that commute with the supercharges, but not necessarily with all the generators of the Poincaré algebra, are called “central charges”. We adhere to this nomenclature. Observe that the issue of central charges is different when not in flat space, as can be seen, for example, in the context of the anti-de Sitter superalgebra (discussed in Sect. 6).

Table 1. The supercharges in flat Minkowski spacetimes of dimension D. In the second column, $Q_{\rm irr}$ specifies the real dimension of an irreducible spinor in a D-dimensional Minkowski spacetime. The third column specifies the group $H_{\rm R}$ for N-extended supersymmetry, defined in the text, acting on N-fold reducible spinor charges. The fourth column denotes the type of spinors: Majorana (M), Dirac (D), Weyl (W) and Majorana-Weyl (MW).

D	$Q_{\rm irr}$	$H_{\rm R}$	type
3, 9, 11, mod 8	$2^{(D-1)/2}$	$SO(N)$	M
5, 7, mod 8	$2^{(D+1)/2}$	$USp(2N)$	D
4, 8, mod 8	$2^{D/2}$	$U(N)$	M
6, mod 8	$2^{D/2}$	$USp(2N_+) \times USp(2N_-)$	W
2, 10, mod 8	$2^{D/2-1}$	$SO(N_+) \times SO(N_-)$	MW

can exist or not. The cases where we cannot have Majorana spinors, corresponding to $D = 6$ mod 8, comprise the fourth sequence. For the last sequence with $D = 2$ mod 8, Majorana spinors exist and the charges can be restricted to so-called Majorana-Weyl spinors.

One can consider [illegible], where the spinor charges transform reducibly under the Lorentz group and comprise N irreducible spinors. For Weyl charges, one can consider combinations of N_+ positive- and N_- negative-chirality spinors. In all these cases there exists a group $H_{\rm R}$ of rotations of the spinors which commute with the Lorentz group and leave the supersymmetry algebra invariant. This group, often referred to as the "R-symmetry" group, is thus defined as the largest subgroup of the automorphism group of the supersymmetry algebra that commutes with the Lorentz group. It is often realized as a manifest invariance group of a supersymmetric field theory, but this is by no means necessary. There are other versions of the R-symmetry group $H_{\rm R}$ which play a role, for instance, in the context of the Euclidean rest-frame superalgebra for massive representations or for the anti-de Sitter superalgebra. Those will be discussed later in this section. In Table 1 we have listed the corresponding $H_{\rm R}$ groups for N irreducible spinor charges. Here we have assumed that $H_{\rm R}$ is compact so that it preserves a positive-definite metric. In the latter two sequences of spinor charges shown in Table 1, we allow $N_\pm$ charges of opposite chirality, so that $H_{\rm R}$ decomposes into the product of two such groups, one for each chiral sector.

Another way to present some of the results above, is shown in Table 2. Here we list the real dimension of an irreducible spinor charge and the corresponding spacetime dimension. In addition we include the number of

Table 2. Simple supersymmetry in various dimensions. We present the dimension of the irreducible spinor charge with $2 \leq Q_{\rm irr} \leq 32$ and the corresponding spacetime dimensions D. The third column represents the number of bosonic + fermionic *massless* states for the shortest supermultiplet.

$Q_{\rm irr}$	D	shortest supermultiplet
32	$D = 11$	$128 + 128$
16	$D = 10, 9, 8, 7$	$8 + 8$
8	$D = 6, 5$	$4 + 4$
4	$D = 4$	$2 + 2$
2	$D = 3$	$1 + 1$

states of the shortest[3] supermultiplet of massless states, written as a sum of bosonic and fermionic states. We return to a more general discussion of the R-symmetry groups and their consequences in Section 2.5.

2.2 Massless supermultiplets

Because the momentum operators P_μ commute with the supercharges, we may consider the states at arbitrary but fixed momentum P_μ, which, for massless representations, satisfies $P^2 = 0$. The matrix $P_\mu \Gamma^\mu$ on the right-hand side of (2.1) has therefore zero eigenvalues. In a positive-definite Hilbert space some (linear combinations) of the supercharges must therefore vanish. To exhibit this more explicitly, let us rewrite (2.1) as (using $\bar{Q} = iQ^\dagger \Gamma^0$),

$$\{Q_\alpha, Q^\dagger_\beta\} = 2\, (\not{P}\Gamma^0)_{\alpha\beta}\,. \tag{2.2}$$

For light-like $P^\mu = (P^0, \vec{P})$ the right-hand side is proportional to a projection operator $(\mathbf{1} + \Gamma_{||}\Gamma^0)/2$. Here $\Gamma_{||}$ is the gamma matrix along the spatial momentum $\vec{P}$ of the states. The supersymmetry anti-commutator can then be written as

$$\{Q_\alpha, Q^\dagger_\beta\} = 2\, P^0 \Big(\mathbf{1} + \tilde{\Gamma}_D \tilde{\Gamma}_\perp\Big)_{\alpha\beta}\,. \tag{2.3}$$

Here $\tilde{\Gamma}_D$ consists of the product of all D independent gamma matrices, and $\tilde{\Gamma}_\perp$ of the product of all $D-2$ gamma matrices in the transverse directions (, perpendicular to $\vec{P}$), with phase factors such that

$$(\tilde{\Gamma}_D)^2 = (\tilde{\Gamma}_\perp)^2 = \mathbf{1}\,, \qquad [\tilde{\Gamma}_D, \tilde{\Gamma}_\perp] = 0\,. \tag{2.4}$$

[3]By the *shortest* multiplet, we mean the multiplet with the helicities of the states as low as possible. This is usually (one of) the smallest possible supermultiplet(s).

This shows that the right-hand side of (2.3) is proportional to a projection operator, which projects out half of the spinor space. Consequently, half the spinors must vanish on physical states, whereas the other ones generate a Clifford algebra.

Denoting the real dimension of the supercharges by Q, the representation space of the charges decomposes into the two chiral spinor representations of $SO(Q/2)$. When confronting these results with the last column in Table 2, it turns out that the dimension of the shortest supermultiplet is not just equal to $2^{Q_{\rm irr}/4}$, as one might naively expect. For $D = 6$, this is so because the representation is complex. For $D = 3, 4$ the representation is twice as big because it must also accommodate fermion number (or, alternatively, because it must be CPT self-conjugate). The derivation for $D = 4$ is presented in many places (see, for instance [1,4,5]). For $D = 3$ we refer to [8].

The two chiral spinor subspaces correspond to the bosonic and fermionic states, respectively. For the massless multiplets, the dimensions are shown in Table 2. Bigger supermultiplets can be obtained by combining irreducible multiplets by requiring them to transform nontrivially under the Lorentz group. We shall demonstrate this below in three relevant cases, corresponding to $D = 11, 10$ and 6 spacetime dimensions. Depending on the number of spacetime dimensions, many supergravity theories exist. Pure supergravity theories with spacetime dimension $4 \leq D \leq 11$ can exist with $Q = 32, 24, 20, 16, 12, 8$ and 4 supersymmetries[4]. Some of these theories will be discussed later in more detail (in particular supergravity in $D = 11$ and 10 spacetime dimensions).

2.2.1 $D = 11$ supermultipets

In 11 dimensions we are dealing with 32 independent real supercharges. In odd-dimensional spacetimes irreducible spinors are subject to the eigenvalue condition $\tilde{\Gamma}_D = \pm\mathbf{1}$. Therefore (2.3) simplifies and shows that the 16 nonvanishing spinor charges transform according to a single spinor representation of the helicity group $SO(9)$.

On the other hand, when regarding the 16 spinor charges as gamma matrices, it follows that the representation space constitutes the spinor representation of $SO(16)$, which decomposes into two chiral subspaces, one corresponding to the bosons and the other one to the fermions. To determine the helicity content of the bosonic and fermionic states, one considers

[4] In $D = 4$ there exist theories with $Q = 12, 20$ and 24; in $D = 5$ there exists a theory with $Q = 24$ [9]. In $D = 6$ there are three theories with $Q = 32$ and one with $Q = 24$. So far these supergravities have played no role in string theory. For a more recent discussion, see [10].

the embedding of the $SO(9)$ spinor representation in the $SO(16)$ vector representation. It then turns out that one of the **128** representations branches into helicity representations according to $\mathbf{128} \to \mathbf{44} + \mathbf{84}$, while the second one transforms irreducibly according to the **128** representation of the helicity group.

The above states comprise precisely the massless states corresponding to $D = 11$ supergravity [11]. The graviton states transform in the **44**, the antisymmetric tensor states in the **84** and the gravitini states in the **128** representation of $SO(9)$. Bigger supermultiplets consist of multiples of 256 states. For instance, without central charges, the smallest massive supermultiplet comprises $32\,768 + 32\,768$ states. These multiplets will not be considered here.

2.2.2 $D = 10$ supermultiplets

In 10 dimensions the supercharges are both Majorana and Weyl spinors. The latter means that they are eigenspinors of $\tilde{\Gamma}_D$. According to (2.3), when we have simple ([illegible], nonextended) supersymmetry with 16 charges, the nonvanishing charges transform in a chiral spinor representation of the $SO(8)$ helicity group. With 8 nonvanishing supercharges we are dealing with an 8-dimensional Clifford algebra, whose irreducible representation space corresponds to the bosonic and fermionic states, each transforming according to a chiral spinor representation. Hence we are dealing with three 8-dimensional representations of $SO(8)$, which are inequivalent. One is the representation to which we assign the supercharges, which we will denote by $\mathbf{8}_s$; to the other two, denoted as the $\mathbf{8}_v$ and $\mathbf{8}_c$ representations, we assign the bosonic and fermionic states, respectively. The fact that $SO(8)$ representations appear in a three-fold variety is known as [illegible], which is a characteristic property of the group $SO(8)$. With the exception of certain representations, such as the adjoint and the singlet representation, the three types of representation are inequivalent. They are traditionally distinguished by labels s, v and c (see, for instance [12])[5].

The smallest massless supermultiplet has now been constructed with 8 bosonic and 8 fermionic states and corresponds to the vector multiplet of supersymmetric Yang-Mills theory in 10 dimensions [13]. Before constructing the supermultiplets that are relevant for $D = 10$ supergravity, let us first discuss some other properties of $SO(8)$ representations. One way to distinguish the inequivalent representations, is to investigate how they decompose

[5]The representations can be characterized according to the four different conjugacy classes of the $SO(8)$ weight vectors, denoted by 0, v, s and c. In this context one uses the notation $\mathbf{1}_0$, $\mathbf{28}_0$, and $\mathbf{35}_0$, $\mathbf{35}'_0$, $\mathbf{35}''_0$ for $\mathbf{35}_v$, $\mathbf{35}_s$, $\mathbf{35}_c$, respectively.

Table 3. Massless $N = 1$ supermultiplets in $D = 10$ spacetime dimensions containing $8 + 8$ or $64 + 64$ bosonic and fermionic degrees of freedom.

supermultiplet	bosons	fermions
vector multiplet	$\mathbf{8}_v$	$\mathbf{8}_c$
graviton multiplet	$\mathbf{1} + \mathbf{28} + \mathbf{35}_v$	$\mathbf{8}_s + \mathbf{56}_s$
gravitino multiplet	$\mathbf{1} + \mathbf{28} + \mathbf{35}_c$	$\mathbf{8}_s + \mathbf{56}_s$
gravitino multiplet	$\mathbf{8}_v + \mathbf{56}_v$	$\mathbf{8}_c + \mathbf{56}_c$

into representations of an $SO(7)$ subgroup. Each of the 8-dimensional representations leaves a different $SO(7)$ subgroup of $SO(8)$ invariant. Therefore there is an $SO(7)$ subgroup under which the $\mathbf{8}_v$ representation branches into

$$\mathbf{8}_v \longrightarrow \mathbf{7} + \mathbf{1}.$$

Under this $SO(7)$ the other two 8-dimensional representations branch into

$$\mathbf{8}_s \longrightarrow \mathbf{8}\,, \qquad \mathbf{8}_c \longrightarrow \mathbf{8}\,,$$

where $\mathbf{8}$ is the spinor representation of $SO(7)$. Corresponding branching rules for the 28-, 35- and 56-dimensional representations are

$$\begin{array}{rcl} \mathbf{28} & \longrightarrow & \mathbf{7} + \mathbf{21}\,, \\ \mathbf{35}_v & \longrightarrow & \mathbf{1} + \mathbf{7} + \mathbf{27}\,, \\ \mathbf{35}_{c,s} & \longrightarrow & \mathbf{35}\,, \end{array} \qquad \begin{array}{rcl} & & \\ \mathbf{56}_v & \longrightarrow & \mathbf{21} + \mathbf{35}\,, \\ \mathbf{56}_{c,s} & \longrightarrow & \mathbf{8} + \mathbf{48}\,. \end{array} \tag{2.5}$$

In order to obtain the supersymmetry representations relevant for supergravity we consider tensor products of the smallest supermultiplet consisting of $\mathbf{8}_v + \mathbf{8}_c$, with one of the 8-dimensional representations. There are thus three different possibilities, each leading to a 128-dimensional supermultiplet. Using the multiplication rules for $SO(8)$ representations,

$$\begin{array}{rcl} \mathbf{8}_v \times \mathbf{8}_v & = & \mathbf{1} + \mathbf{28} + \mathbf{35}_v\,, \\ \mathbf{8}_s \times \mathbf{8}_s & = & \mathbf{1} + \mathbf{28} + \mathbf{35}_s\,, \\ \mathbf{8}_c \times \mathbf{8}_c & = & \mathbf{1} + \mathbf{28} + \mathbf{35}_c\,, \end{array} \qquad \begin{array}{rcl} \mathbf{8}_v \times \mathbf{8}_s & = & \mathbf{8}_c + \mathbf{56}_c\,, \\ \mathbf{8}_s \times \mathbf{8}_c & = & \mathbf{8}_v + \mathbf{56}_v\,, \\ \mathbf{8}_c \times \mathbf{8}_v & = & \mathbf{8}_s + \mathbf{56}_s\,, \end{array} \tag{2.6}$$

it is straightforward to obtain these new multiplets. Multiplying $\mathbf{8}_v$ with $\mathbf{8}_v + \mathbf{8}_c$ yields $\mathbf{8}_v \times \mathbf{8}_v$ bosonic and $\mathbf{8}_v \times \mathbf{8}_c$ fermionic states, and leads to the second supermultiplet shown in Table 3. This supermultiplet contains the representation $\mathbf{35}_v$, which can be associated with the states of the graviton in $D = 10$ dimensions (the field-theoretic identification of the various states has been clarified in many places; see the Appendix in [6]). Therefore this supermultiplet will be called the Multiplication with $\mathbf{8}_c$ or $\mathbf{8}_s$ goes in the same fashion, except that we will associate the $\mathbf{8}_c$ and $\mathbf{8}_s$ representations with fermionic quantities (note that these are the representations to which the fermion states of the Yang-Mills multiplet and the supersymmetry charges are assigned). Consequently, we interchange the boson and fermion assignments in these products. Multiplication with $\mathbf{8}_c$ then leads to $\mathbf{8}_c \times \mathbf{8}_c$ bosonic and $\mathbf{8}_c \times \mathbf{8}_v$ fermionic states, whereas multiplication with $\mathbf{8}_s$ gives $\mathbf{8}_s \times \mathbf{8}_c$ bosonic and $\mathbf{8}_s \times \mathbf{8}_v$ fermionic states. These supermultiplets contain fermions transforming according to the $\mathbf{56}_s$ and $\mathbf{56}_c$ representations, respectively, which can be associated with gravitino states, but no graviton states as those transform in the $\mathbf{35}_v$ representation. Therefore these two supermultiplets are called . We have thus established the existence of two inequivalent gravitino multiplets. The explicit $SO(8)$ decompositions of the vector, graviton and gravitino supermultiplets are shown in Table 3.

By combining a graviton and a gravitino multiplet it is possible to construct an $N = 2$ supermultiplet of 128 + 128 bosonic and fermionic states. However, since there are two inequivalent gravitino multiplets, there will also be two inequivalent $N = 2$ supermultiplets containing the states corresponding to a graviton and two gravitini. According to the construction presented above, one $N = 2$ supermultiplet may be be viewed as the tensor product of two identical supermultiplets (namely $\mathbf{8}_v + \mathbf{8}_c$). Such a multiplet follows if one starts from a supersymmetry algebra based on Majorana-Weyl spinor charges Q with the chirality. The states of this multiplet decompose as follows:

$$\begin{array}{c} N = 2 \quad \text{(IIB)} \\[1ex] (\mathbf{8}_v + \mathbf{8}_c) \times (\mathbf{8}_v + \mathbf{8}_c) \end{array} \Longrightarrow \begin{cases} \textbf{bosons:} \\ \mathbf{1} + \mathbf{1} + \mathbf{28} + \mathbf{28} + \mathbf{35}_v + \mathbf{35}_c \\ \\ \textbf{fermions:} \\ \mathbf{8}_s + \mathbf{8}_s + \mathbf{56}_s + \mathbf{56}_s . \end{cases} \tag{2.7}$$

This is the multiplet corresponding to IIB supergravity [14]. Because the supercharges have the same chirality, one can perform rotations between these spinor charges which leave the supersymmetry algebra unaffected. Hence the automorphism group H_{R} is equal to $SO(2)$. This feature reflects itself in the multiplet decomposition, where the $\mathbf{1}$, $\mathbf{8}_s$, $\mathbf{28}$ and $\mathbf{56}_s$ representations are degenerate and constitute doublets under this $SO(2)$ group.

A second supermultiplet may be viewed as the tensor product of a $(\mathbf{8}_v + \mathbf{8}_s)$ supermultiplet with a second supermultiplet $(\mathbf{8}_v + \mathbf{8}_c)$. In this case the supercharges constitute two Majorana-Weyl spinors of opposite chirality.

Now the supermultiplet decomposes as follows:

$$\begin{array}{c} N=2 \qquad\qquad \text{(IIA)} \\[1em] (\mathbf{8}_v+\mathbf{8}_s)\times(\mathbf{8}_v+\mathbf{8}_c) \end{array} \quad \Longrightarrow \quad \begin{cases} \textbf{bosons:} \\ \mathbf{1}+\mathbf{8}_v+\mathbf{28}+\mathbf{35}_v+\mathbf{56}_v \\ \\ \textbf{fermions:} \\ \mathbf{8}_s+\mathbf{8}_c+\mathbf{56}_s+\mathbf{56}_c . \end{cases} \tag{2.8}$$

This is the multiplet corresponding to IIA supergravity [15]. It can be obtained by a straightforward reduction of $D = 11$ supergravity. The latter follows from the fact that two $D = 10$ Majorana-Weyl spinors with opposite chirality can be combined into a single $D = 11$ Majorana spinor. The formula below summarizes the massless states of IIA supergravity from an 11-dimensional perspective. The massless states of 11-dimensional supergravity transform according to the **44**, **84** and **128** representation of the helicity group $SO(9)$. They correspond to the degrees of freedom described by the metric, a 3-rank antisymmetric gauge field and the gravitino field, respectively. We also show how the 10-dimensional states can subsequently be branched into 9-dimensional states, characterized in terms of representations of the helicity group $SO(7)$:

$$\begin{array}{ccl} \mathbf{44} & \Longrightarrow & \begin{cases} \mathbf{1} & \longrightarrow & \mathbf{1} \\ \mathbf{8}_v & \longrightarrow & \mathbf{1}+\mathbf{7} \\ \mathbf{35}_v & \longrightarrow & \mathbf{1}+\mathbf{7}+\mathbf{27} \end{cases} \\ \mathbf{84} & \Longrightarrow & \begin{cases} \mathbf{28} & \longrightarrow & \mathbf{7}+\mathbf{21} \\ \mathbf{56}_v & \longrightarrow & \mathbf{21}+\mathbf{35} \end{cases} \\ \mathbf{128} & \Longrightarrow & \begin{cases} \mathbf{8}_s & \longrightarrow & \mathbf{8} \\ \mathbf{8}_c & \longrightarrow & \mathbf{8} \\ \mathbf{56}_s & \longrightarrow & \mathbf{8}+\mathbf{48} \\ \mathbf{56}_c & \longrightarrow & \mathbf{8}+\mathbf{48}. \end{cases} \end{array} \tag{2.9}$$

Clearly, in $D = 9$ we have a degeneracy of states, associated with the group $H_{\mathrm{R}} = SO(2)$. We note the presence of graviton and gravitino states, transforming in the **27** and **48** representations of the $SO(7)$ helicity group.

One could also take the states of the IIB supergravity and decompose them into $D = 9$ massless states. This leads to precisely the same supermultiplet as the reduction of the states of IIA supergravity. Indeed, the reductions of IIA and IIB supergravity to 9 dimensions, yield the same theory [16–18]. However, the massive states are still characterized in terms of the group $SO(8)$, which in $D = 9$ dimensions comprises the rest-frame rotations. Therefore the Kaluza-Klein states that one obtains when compactifying the ten-dimensional theory on a circle remain for the IIA and IIB theories (see [19] for a discussion of this phenomenon and its

Table 4. Shortest massless supermultiplets of $D = 6$ N_+-extended chiral supersymmetry. The states transform both in the $SU_+(2)$ helicity group and under a $USp(2N_+)$ group. For odd values of N_+ the representations are complex, for even N_+ they can be chosen real. Of course, an identical table can be given for negative-chirality spinors.

$SU_+(2)$	$N_+ = 1$	$N_+ = 2$	$N_+ = 3$	$N_+ = 4$
5				1
4			1	8
3		1	6	27
2	1	4	14	48
1	2	5	14	42
	$(2+2)_{\mathbf{C}}$	$(8+8)_{\mathbf{R}}$	$(32+32)_{\mathbf{C}}$	$(128+128)_{\mathbf{R}}$

consequences). It turns out that the $Q = 32$ supergravity multiplets are unique in all spacetime dimensions $D > 2$, except for $D = 10$. Maximal supergravity will be introduced in Section 3. The field content of the maximal $Q = 32$ supergravity theories for dimensions $3 \leq D \leq 11$ will be presented in two tables (Tables 10 and 11).

2.2.3 $D = 6$ supermultiplets

In 6 dimensions we have chiral spinors, which are not Majorana. Because the charge conjugated spinor has the same chirality, the chiral rotations of the spinors can be extended to the group $USp(2N_+)$, for N_+ chiral spinors. Likewise N_- negative-chirality spinors transform under $USp(2N_-)$. This feature is already incorporated in Table 1. In principle we have N_+ positive- and N_- negative-chirality charges, but almost all information follows from first considering the purely chiral case. In Table 4 we present the decomposition of the various helicity representations of the smallest supermultiplets based on $N_+ = 1, 2, 3$ or 4 supercharges. In $D = 6$ dimensions the helicity group $SO(4)$ decomposes into the product of two $SU(2)$ groups: $SO(4) \cong (SU_+(2) \times SU_-(2))/\mathbf{Z}_2$. When we have supercharges of only one chirality, the smallest supermultiplet will only transform under one $SU(2)$ factor of the helicity group, as is shown in Table 4[6].

Let us now turn to specific supermultiplets. Let us recall that the helicity assignments of the states describing gravitons, gravitini, vector and

[6]The content of this table also specifies the shortest *massive* supermultiplets in four dimensions as well as with the shortest *massless* multiplets in five dimensions. The $SU(2)$ group is then associated with spin or with helicity, respectively.

(anti)selfdual tensor gauge fields, and spinor fields are $(3,3)$, $(2,3)$ or $(3,2)$, $(2,2)$, $(3,1)$ or $(1,3)$, and $(2,1)$ or $(1,2)$. Here (m,n) denotes that the dimensionality of the reducible representations of the two $SU(2)$ factors of the helicity group are of dimension m and n. For the derivation of these assignments, see for instance one of the Appendices in [6].

In the following we will first restrict ourselves to helicities that correspond to at most the three-dimensional representation of either one of the $SU(2)$ factors. Hence we have only $(3,3)$, $(3,2)$, $(2,3)$, $(3,1)$ or $(1,3)$ representations, as well as the lower-dimensional ones. When a supermultiplet contains $(3,2)$ or $(2,3)$ representations, we insist that it will also contain a single $(3,3)$ representation, because gravitini without a graviton are not expected to give rise to a consistent interacting field theory. The multiplets of this type are shown in Table 5. There are no such multiplets for more than $Q = 32$ supercharges.

There are supermultiplets with higher $SU(2)$ helicity representations, which contain neither gravitons nor gravitini. Some of these multiplets are shown in Table 6 and we will discuss them in due course.

We now elucidate the construction of the supermultiplets listed in Table 5. The simplest case is $(N_+, N_-) = (1,\ 0)$, where the smallest supermultiplet is the $(1,\ 0)$ [illegible], consisting of a complex doublet of spinless states and a chiral spinor. Taking the tensor product of the smallest supermultiplet with the $(2,1)$ helicity representation gives the $(1,\ 0)$ [illegible], with a selfdual tensor, a spinless state and a doublet of chiral spinors. The tensor product with the $(1,2)$ helicity representation yields the $(1,\ 0)$ [illegible], with a vector state, a doublet of chiral spinors and a scalar. Multiplying the hypermultiplet with the $(2,\ 3)$ helicity representation, one obtains the states of $(1,\ 0)$ [illegible]. Observe that the selfdual tensor fields in the tensor and supergravity supermultiplet are of opposite selfduality phase.

Next consider $(N_+, N_-) = (2,0)$ supersymmetry. The smallest multiplet, shown in Table 4, then corresponds to the $(2,0)$ [illegible], with the bosonic states decomposing into a selfdual tensor, and a five-plet of spinless states, and a four-plet of chiral fermions. Multiplication with the $(1,3)$ helicity representation yields the $(2,0)$ supergravity multiplet, consisting of the graviton, four chiral gravitini and five selfdual tensors [20]. Again, the selfdual tensors of the tensor and of the supergravity supermultiplet are of opposite selfduality phase.

Of course, there exists also a nonchiral version with 16 supercharges, namely the one corresponding to $(N_+, N_-) = (1,1)$. The smallest multiplet is now given by the tensor product of the supermultiplets with $(1,\ 0)$ and $(0,\ 1)$ supersymmetry. This yields the vector multiplet, with the vector state and four scalars, the latter transforming with respect to the $(2,\ 2)$

Table 5. Some relevant $D = 6$ supermultiplets with (N_+, N_-) supersymmetry. The states $(m, n; \tilde{m}, \tilde{n})$ are assigned to (m, n) representations of the helicity group $SU_+(2) \times SU_-(2)$ and $(\tilde{m}, \tilde{n})$ representations of $USp(2N_+) \times USp(2N_-)$. The second column lists the number of bosonic + fermionic states for each multiplet.

multiplet	#	bosons	fermions
(1, 0) hyper	$4+4$	$(1,1;2,1) + \text{h.c.}$	$(2,1;1,1)$
(1, 0) tensor	$4+4$	$(3,1;1,1) + (1,1;1,1)$	$(2,1;2,1)$
(1, 0) vector	$4+4$	$(2,2;1,1)$	$(1,2;2,1)$
(1, 0) graviton	$12+12$	$(3,3;1,1) + (1,3;1,1)$	$(2,3;2,1)$
(2, 0) tensor	$8+8$	$(3,1;1,1) + (1,1;5,1)$	$(2,1;4,1)$
(2, 0) graviton	$24+24$	$(3,3;1,1) + (1,3;5,1)$	$(2,3;4,1)$
(1, 1) vector	$8+8$	$(2,2;1,1) + (1,1;2,2)$	$(2,1;1,2)$ $+(1,2;2,1)$
(1, 1) graviton	$32+32$	$(3,3;1,1)$ $+(1,3;1,1) + (3,1;1,1)$ $+(1,1;1,1) + (2,2;2,2)$	$(3,2;1,2)$ $+(2,3;2,1)$ $+(1,2;1,2)$ $+(2,1;2,1)$
(2, 1) graviton	$64+64$	$(3,3;1,1)$ $+(1,3;5,1) + (3,1;1,1)$ $+(2,2;4,2) + (1,1;5,1)$	$(3,2;1,2)$ $+(2,3;4,1)$ $+(1,2;5,2)$ $+(2,1;4,1)$
(2, 2) graviton	$128+128$	$(3,3;1,1)$ $+(3,1;1,5) + (1,3;5,1)$ $+(2,2;4,4) + (1,1;5,5)$	$(3,2;4,1)$ $+(2,3;1,4)$ $+(2,1;4,5)$ $+(1,2;5,4)$

representation of $USp(2) \times USp(2)$. There are two doublets of chiral fermions with opposite chirality, each transforming as a doublet under the corresponding $USp(2)$ group. Taking the tensor product of the vector multiplet with the (2, 2) representation of the helicity group yields the states of the (1, 1) [illegible] multiplet. It consists of 32 bosonic states, corresponding to a graviton, a tensor, a scalar and four vector states, where the latter transform under the (2, 2) representation of $USp(2) \times USp(2)$. The 32 fermionic states comprise two doublets of chiral gravitini and two chiral spinor doublets, transforming as doublets under the appropriate $USp(2)$ group.

Table 6. $D = 6$ supermultiplets without gravitons and gravitini with $(N, 0)$ supersymmetry, a single $(5; 1)$ highest-helicity state and at most 32 supercharges. The theories based on these multiplets have only rigid supersymmetry. The multiplets are identical to those that underly the five-dimensional N-extended supergravities. They are all chiral, so that the helicity group in six dimensions is restricted to $SU(2) \times \mathbf{1}$ and the states are characterized as representations of $USp(2N)$. The states $(n; \tilde{n})$ are assigned to the n-dimensional representation of $SU(2)$ and the $\tilde{n}$-dimensional representation of $USp(2N)$. The second column lists the number of bosonic + fermionic states for each multiplet.

supersymmetry	#	bosons	fermions
(1, 0)	$8+8$	$(5;1)+(3;1)$	$(4;2)$
(2, 0)	$24+24$	$(5;1)+(1,1)$ $+(3;1)+(3;5)$	$(4;4)+(2;4)$
(3, 0)	$64+64$	$(5;1)+(1;14)$ $+(3;1)+(3;14)$	$(4;6)$ $+(2;6)+(2;14)$
(4, 0)	$128+128$	$(5;1)+(3;27)$ $+(1;42)$	$(4;8)+(2;48)$

Subsequently we discuss the case $(N_+, N_-) = (2, 1)$. Here a supergravity multiplet exists [21] and can be obtained from the product of the states of the $(2, 0)$ tensor multiplet with the $(0, 1)$ tensor multiplet. There is in fact a smaller supermultiplet, which we discard because it contains gravitini but no graviton states.

Finally, we turn to the case of $(N_+, N_-) = (2, 2)$. The smallest supermultiplet is given by the tensor product of the smallest (2, 0) and (0, 2) supermultiplets. This yields the 128 + 128 states of the (2, 2) [illegible]-[illegible] multiplet. These states transform according to representations of $USp(4) \times USp(4)$.

In principle, one can continue and classify representations for other values of (N_+, N_-). As is obvious from the construction that we have presented, this will inevitably lead to states transforming in higher-helicity representations. Some of these multiplets will suffer from the fact that they have more than one graviton state, so that we expect them to be inconsistent at the nonlinear level. However, there are the chiral theories which contain neither graviton nor gravitino states. Restricting ourselves to 32 supercharges and requiring the highest helicity to be a five-dimensional representation of one of the $SU(2)$ factors, there are just four theories, summarized in Table 6. For a recent discussion of one of these theories, see [10].

2.3 Massive supermultiplets

Generically massive supermultiplets are bigger than massless ones because the number of supercharges that generate the multiplet is not reduced, unlike for massless supermultiplets where one-half of the supercharges vanishes. However, in the presence of mass parameters the superalgebra may also contain central charges, which could give rise to a shortening of the representation in a way similar to what happens for the massless supermultiplets. This only happens for special values of these charges. The shortened supermultiplets are known as BPS multiplets. Central charges and multiplet shortening are discussed in Section 2.4. In this section we assume that the central charges are absent.

The analysis of massive supermultiplets takes place in the restframe. The states then organize themselves into representations of the rest-frame rotation group, $SO(D-1)$, associated with spin. The supercharges transform as spinors under this group, so that one obtains a Euclidean supersymmetry algebra,

$$\{Q_\alpha, Q^\dagger_\beta\} = 2M\,\delta_{\alpha\beta}\,. \tag{2.10}$$

Just as before, the spinor charges transform under the automorphism group of the supersymmetry algebra that commutes with the spin rotation group. This group will also be denoted by $H_{\rm R}$; it is the nonrelativistic variant of the R-symmetry group that was introduced previously. Obviously the nonrelativistic group can be bigger than its relativistic counterpart, as it is required to commute with a smaller group. For instance, in $D=4$ spacetime dimensions, the relativistic R-symmetry group is equal to $U(N)$, while the nonrelativistic one is the group $USp(2N)$, which contains $U(N)$ as a subgroup according to $\mathbf{2N} = \mathbf{N} + \overline{\mathbf{N}}$. Table 7 shows the smallest massive representations for $N \leq 4$ in $D=4$ dimensions as an illustration. Clearly the states of given spin can be assigned to representations of the nonrelativistic group $H_{\rm R} = USp(2N)$ and decomposed in terms of irreducible representations of the relativistic R-symmetry group $U(N)$. More explicit derivations can be found in [4].

Knowledge of the relevant groups $H_{\rm R}$ is important and convenient in writing down the supermultiplets. It can also reveal certain relations between supermultiplets, even between supermultiplets living in spacetimes of different dimension. Obviously, supermultiplets living in higher dimensions can always be decomposed into supermultiplets living in lower dimensions, and massive supermultiplets can be decomposed in terms of massless ones, but sometimes there exists a relationship that is less trivial. For instance, the $D=4$ multiplets shown in Table 7 coincide with the supermultiplets of chirally extended supersymmetry in $D=6$ dimensions shown in Table 4. In particular the $N=4$ supermultiplet of Table 7 appears in many places and coincides with the massless $N=8$ supermultiplet in $D = 5$ dimensions, which is shown in Tables 10 and 11. The reasons for this are clear. The $D = 5$ and the chiral $D = 6$ massless supermultiplets are subject to the same helicity group $SU(2)$, which in turn coincides with the spin rotation group for $D = 4$. Not surprisingly, also the relevant automorphism groups $H_{\rm R}$ coincide, as the reader can easily verify. Since the number of *ff* supercharges is equal in these cases and given by $Q_{\rm eff} = 16$ (remember that only half of the charges play a role in building up massless supermultiplets), the multiplets must indeed be identical.

Here we also want to briefly draw the attention to the relation between off-shell multiplets and massive representations. So far we discussed supermultiplets consisting of states on which the supercharges act. These states can be described by a field theory in which the supercharges generate corresponding supersymmetry variations on the fields. Very often the transformations on the fields do close according to the supersymmetry algebra unless one imposes the equations of motion for the fields. Such representations are called - representations. The lack of closure has many consequences, for instance, when determining quantum corrections. In certain cases one can improve the situation by introducing extra fields which do not directly correspond to physical fields. These fields are known as - *fi* . By employing such fields one may be able to define an *ff-* representation, where the transformations close upon (anti)commutation without the need for imposing field equations. Unfortunately, many theories do not possess (finite-dimensional) off-shell representations. Notorious examples are gauge theories and supergravity theories with 16 or more supercharges. This fact makes is much more difficult to construct an extended variety of actions for these theories, because the transformation rules are implicitly dependent on the action. There is an off-shell counting argument. according to which the field degrees of freedom should comprise a supermultiplet (while the states that are described could be massless). For instance, the off-shell description of the $N = 2$ vector multiplet in $D = 4$ dimensions can be formulated in terms of a gauge field (with three degrees of freedom), a fermion doublet (with eight degrees of freedom) and a triplet of auxiliary scalar fields (with three degrees of freedom), precisely in accord with the $N = 2$ entry in Table 7. In fact, this multiplet coincides with the multiplet of the currents that couple to an $N = 2$ supersymmetric gauge theory.

The $N = 4$ multiplet in Table 7 corresponds to the gravitational supermultiplet of currents [22]. These are the currents that couple to the fields of $N = 4$ conformal supergravity. Extending the number of supercharges beyond 16 will increase the minimal spin of a massive multiplet beyond spin-2. Since higher-spin fields can usually not be coupled, one may conclude that conformal supergravity does not exist for more than 16 charges. For that

Table 7. Minimal $D = 4$ massive supermultiplets without central charges for $N \leq 4$. The states are listed as $USp(2N)$ representations which are subsequently decomposed into representations of $U(N)$.

spin	$N = 1$	$N = 2$
1		$1 = 1$
1/2	$1 = 1$	$4 = 2 + \bar{2}$
0	$2 = 1 + \bar{1}$	$5 = 3 + 1 + \bar{1}$
	$N = 3$	$N = 4$
2		$1 = 1$
3/2	$1 = 1$	$8 = 4 + \bar{4}$
1	$6 = 3 + \bar{3}$	$27 = 15 + 6 + \bar{6}$
1/2	$14 = 8 + 3 + \bar{3}$	$48 = 20 + \overline{20} + 4 + \bar{4}$
0	$14 = 6 + \bar{6} + 1 + \bar{1}$	$42 = 20' + 10 + \overline{10} + 1 + \bar{1}$

reason there can be no off-shell formulations for supergravity with more than 16 charges. Conformal supergravity will be discussed in Section 7.

In Section 2.5 we will present a table listing the various groups $H_{\rm R}$ for spinors associated with certain Clifford algebras $\mathcal{C}(p, q)$ with corresponding rotation groups $SO(p, q)$. Subsequently we then discuss some further implications of these results.

2.4 Central charges and multiplet shortening

The supersymmetry algebra of the maximal supergravities comprises general coordinate transformations, local supersymmetry transformations and the gauge transformations associated with the antisymmetric gauge fields[7]. These gauge transformations usually appear in the anticommutator of two supercharges, and may be regarded as central charges. In perturbation theory, the theory does not contain charged fields, so these central charges simply vanish on physical states. However, at the nonperturbative level, there may be solitonic or other states that carry charges. An example are magnetic monopoles, dyons, or black holes. At the M-theory level, these charges are associated with certain brane configurations. On such states, some of the central charges may take finite values. Without further knowledge about the kind of states that may emerge at the nonperturbative level,

[7]There may be additional gauge transformations that are of interest to us. As we discuss in due course, it is possible to have (part of) the automorphism group $H_{\rm R}$ realized as a local invariance. However, the corresponding gauge fields are composite and do not give rise to physical states (at least, not in perturbation theory).

Table 8. Decomposition of the central extension in the supersymmetry algebra with $Q = 32$ supercharge components in terms of p-rank Lorentz tensors. The second row specifies the number of independent components for each p-rank tensor charge. The total number of central charges is equal to $528 - D$, because we have not listed the D independent momentum operators.

D	$H_{\rm R}$	$p=0$	$p=1$	$p=2$	$p=3$	$p=4$	$p=5$
11	1			1 [55]			1 [462]
10A	1	1 [1]	1 [10]	1 [45]		1 [210]	$1+1$ [126]
10B	$SO(2)$		2 [10]		1 [120]		$1+2$ [126]
9	$SO(2)$	$1+2$ [1]	2 [9]	1 [36]	1 [84]	$1+2$ [126]	
8	$U(2)$	$3+\bar{3}$ [1]	3 [8]	$1+\bar{1}$ [28]	$1+3$ [56]	$3+\bar{3}$ [35]	
7	$USp(4)$	10 [1]	5 [7]	$1+5$ [21]	10 [35]		
6	$USp(4)$ $\times USp(4)$	$(4,4)$ [1]	$(1,1)$ $+(5,1)$ $+(1,5)$ [6]	$(4,\ 4)$ [15]	$(10,1)$ $+(1,10)$ [10]		
5	$USp(8)$	$1+27$ [1]	27 [5]	36 [10]			
4	$U(8)$	$28+\overline{28}$ [1]	63 [4]	$36+\overline{36}$ [3]			
3	$SO(16)$	120 [1]	135 [3]				

we can generally classify the possible central charges, by considering a decomposition of the anticommutator. This anticommutator carries at least two spinor indices and two indices associated with the group $H_{\rm R}$. Hence we may write

$$\{Q_\alpha, Q_\beta\} \propto \sum_p (\Gamma^{\mu_1\cdots\mu_p} C)_{\alpha\beta}\, Z_{\mu_1\cdots\mu_p}\,, \tag{2.11}$$

where $\Gamma^{\mu_1\cdots\mu_p}$ is the antisymmetrized product of p gamma matrices, C is the charge-conjugation matrix and $Z_{\mu_1\cdots\mu_p}$ is the central charge, which transforms as an antisymmetric p-rank Lorentz tensor and depends on possible additional $H_{\rm R}$ indices attached to the supercharges. The central charge must be symmetric or antisymmetric in these indices, depending on whether the $(\Gamma^{\mu_1\cdots\mu_p}C)_{\alpha\beta}$ is antisymmetric or symmetric in α,β, so that the product with $Z_{\mu_1\cdots\mu_p}$ is always symmetric in the combined indices of the supercharges. For given spacetime dimension all possible central charges can be classified[8]. For the maximal supergravities in spacetime dimensions $3\leq D\leq 11$ this classification is given in Table 8, where we list all possible charges and their $H_{\rm R}$ representation assignments. Because we have 32 supercharge components, the sum of the independent momentum operators and the central charges must be equal to $(32\times 33)/2 = 528$. The results of the table are in direct correspondence with the eleven-dimensional superalgebra with the most general central charges,

$$\{Q_\alpha,\bar Q_\beta\} = -2iP_M\Gamma^M_{\alpha\beta} + Z_{MN}\Gamma^{MN}_{\alpha\beta} + Z_{MNPQR}\Gamma^{MNPQR}_{\alpha\beta}\,. \tag{2.12}$$

The two central charges, Z_{MN} and Z_{MNPQR} can be associated with the winding numbers of two- and five-branes.

In order to realize the supersymmetry algebra in a positive-definite Hilbert space, the right-hand side of the anticommutator is subject to a positivity condition, which generically implies that the mass of the multiplet is larger than or equal to the central charges. Especially in higher dimensions, the bound may can take a complicated form. This positivity bound is known as the Bogomol'nyi bound. When the bound is saturated one speaks of BPS states. For BPS multiplets some of the supercharges must vanish on the states, in the same way as half of the charges vanish for the massless supermultiplets. This vanishing of some of the supercharges leads to a shortening of the multiplet. Qualitatively, this phenomenon of multiplet shortening is the same as for massless supermultiplets, but here the fraction of the charges that vanishes is not necessarily equal to 1/2. Hence one speaks of 1/2-BPS, 1/4-BPS supermultiplets, etcetera, to indicate which fraction of the supercharges vanishes on the states. The fact that the BPS supermultiplets have a completely different field content than the generic massive supermultiplets makes that they exhibit a remarkable stability under "adiabatic" deformations. This means that perturbative results based on BPS supermultiplets can often be extrapolated to a nonperturbative regime.

For higher extended supersymmetry the difference in size of BPS supermultiplets and massive supermultiplets can be enormous in view of the fact that the number of states depend exponentionally on the number of nonvanishing central charges. For lower supersymmetry the multiplets can be

[8]For related discussions see, for example [23–25] and references therein.

comparable in size, but nevertheless they are quite different. For instance, consider $N = 2$ vector supermultiplets in four spacetime dimensions. Without central charges, such a multiplet comprises $8 + 8$ states, corresponding to the three states of spin-1, the eight states of four irreducible spin-$\frac{1}{2}$ representations, and five states with spin 0. On the other hand there is another massive vector supermultiplet, which is BPS and comprises the three states of spin-1, the four states of two spin-$\frac{1}{2}$ representations and two states of spin-0. These states are subject to a nonvanishing central charge which requires that the states are all doubly degenerate, so that we have again $8+8$ states, but with a completely different spin content. When decomposing these multiplets into massless $N = 2$ supermultiplets, the first multiplet decomposes into a massless vector multiplet and a hypermultiplet. Hence this is the multiplet one has in the Higgs phase, where the hypermultiplet provides the scalar degree of freedom that allows the conversion of the massless to massive spin-1 states. This multiplet carries no central charge. The second supermultiplet, which is BPS, appears as a massive charged vector multiplet when breaking a non-Abelian supersymmetric gauge theory to an Abelian subgroup. This realization is known as the Coulomb phase.

In view of the very large variety of BPS supermultiplets, we do not continue this general discussion of supermultiplets with central charges. In later sections we will discuss specific BPS supermultiplets as well as other mechanisms of multiplet shortening in anti-de Sitter space.

2.5 On spinors and the R-symmetry group H_{R}

In this section we return once more to the spinor representations and the corresponding automorphism group H_{R}, also known as the R-symmetry group. Table 9 summarizes information for spinors up to (real) dimension 32 associated with the groups $SO(p,q)$, where we restrict $q \leq 2$. From this table we can gain certain insights into the properties of spinors living in Euclidean, Minkowski and (anti-)de Sitter spaces as well as the supersymmetry algebras based on these spinors. Let us first elucidate the information presented in the table. Subsequently we shall discuss some correspondences between the various spinors in different dimensions.

We consider the Clifford algebras $\mathcal{C}(p,q)$ based on $p+q$ generators, denoted by $e_1, e_2, \ldots, e_{p+q}$, with a nondegenerate metric of signature (p,q). This means that p generators square to the identity and q to minus the identity. We list the real dimension of the irreducible Clifford algebra representation, denoted by $d_{\mathcal{C}}$, and the values r (equal to $0,\ldots,3$), where r is defined by $r \equiv p - q \bmod 4$. The value for r determines the square of the matrix built from forming the product $e_1 \cdot e_2 \cdots e_{p+q}$ of all the Clifford algebra generators. For $r = 0, 1$ this square equals the identity, while for $r = 2, 3$ the square equals minus the identity. Therefore, for $r = 0$, the subalgebra $\mathcal{C}_+(p,q)$ generated by products of numbers of generators is not simple and breaks into two simple ideals, while for $r = 1$, the Clifford algebra $\mathcal{C}(p,q)$ decomposes into two simple ideals.

We also present the centralizer of the irreducible representations of the Clifford algebra, which, according to Schur's lemma, must form a division algebra and is thus isomorphic to the real numbers ($\mathbf{R}$), the complex numbers ($\mathbf{C}$), or the quaternions ($\mathbf{H}$). This means that the irreducible representation commutes with the identity and none, one or three complex structures, respectively, which generate the corresponding division algebra. Table 9 reflects also the so-called periodicity theorem [26], according to which there exists an isomorphism between the Clifford algebras $\mathcal{C}(p{+}8,q)$ (or $\mathcal{C}(p,q{+}8)$) and $\mathcal{C}(p,q)$ times the 16×16 real matrices. Therefore, the dimension of the representations of $\mathcal{C}(p{+}8,q)$ (or $\mathcal{C}(p,q{+}8)$) and $\mathcal{C}(p,q)$ differs by a factor 16.

Finally the table lists the branching of the Clifford algebra representation into $SO(p,q)$ spinor representations. When $r = 0$ the Clifford algebra representation decomposes into two chiral spinors. Observe that for $r = 2$ we can also have chiral spinors, but they are complex so that their real dimension remains unaltered. For $r = 2, 3$ there are no chiral spinors, but nevertheless in certain cases the Clifford algebra representation can still decompose into two irreducible spinor representations. The last column gives the compact group $H_{\rm R}$, consisting of the linear transformations that commute with the group $SO(p,q)$ and act on N irreducible spinors, leaving a positive-definite metric invariant. For $r = 0$, a group $H_{\rm R}$ should be assigned to each of the chiral sectors separately. Again, according to Schur's lemma the centralizer of $SO(p,q)$ must form a division algebra for irreducible spinor representations. Correspondingly, the group $SO(p,q)$ commutes with the identity and none, one or three complex structures, which leads to $H_{\rm R} = SO(N)$, $U(N)$, or $USp(2N)$, respectively. We note that the results of Table 9 are in accord with the results presented earlier in Tables 1 and 2.

We now discuss and clarify a number of correspondences between spinors living in different dimensions. The first correspondence is between spinors of $SO(p,1)$ and $SO(p-1,0)$. According to the table, for any $p > 1$, the dimensions of the corresponding spinors differ by a factor two, while their respective groups $H_{\rm R}$ always coincide. From a physical perspective, this correspondence can be understood from the fact that $SO(p-1,0)$ is the group of massless spinor states in flat Minkowski space of dimension $D = p+1$. In a field-theoretic context the reduction of the spinor degrees of freedom is effected by the massless Dirac equation and the automorphism groups $H_{\rm R}$ that commute with the Lorentz transformations and the transverse helicity rotations, respectively, simply coincide. The two algebras (2.1) and (2.3) thus share the same automorphism group. From a

Table 9. Representations of the Clifford algebras $\mathcal{C}(p,q)$ with $q \leq 2$ and their centralizers, and the $SO(p,q)$ spinors of maximal real dimension 32 and their R-symmetry group. We also list the dimensions of the Clifford algebra and spinor representation, as well as $r = p - q$ mod 4.

$d_{\mathcal{C}}$	$\mathcal{C}(p,q)$	r	centralizer	$d_{SO(p,q)}$	$H_{\rm R}$
1	$\mathcal{C}(1,0)$	1	**R**	1	$SO(N)$
2	$\mathcal{C}(0,1)$	3	**C**	$1+1$	$SO(N)$
2	$\mathcal{C}(1,1)$	0	**R**	$1+1$	$SO(N)$
2	$\mathcal{C}(2,0)$	2	**R**	2	$U(N)$
2	$\mathcal{C}(2,1)$	1	**R**	2	$SO(N)$
4	$\mathcal{C}(0,2)$	2	**H**	$2+2$	$U(N)$
4	$\mathcal{C}(1,2)$	3	**C**	$2+2$	$SO(N)$
4	$\mathcal{C}(2,2)$	0	**R**	$2+2$	$SO(N)$
4	$\mathcal{C}(3,0)$	3	**H**	4	$USp(2N)$
4	$\mathcal{C}(3,1)$	2	**C**	4	$U(N)$
4	$\mathcal{C}(3,2)$	1	**R**	4	$SO(N)$
8	$\mathcal{C}(4,0)$	0	**H**	$4+4$	$USp(2N)$
8	$\mathcal{C}(4,1)$	3	**H**	8	$USp(2N)$
8	$\mathcal{C}(4,2)$	2	**C**	8	$U(N)$
8	$\mathcal{C}(5,0)$	1	**H**	8	$USp(2N)$
16	$\mathcal{C}(5,1)$	0	**H**	$8+8$	$USp(2N)$
16	$\mathcal{C}(5,2)$	3	**C**	16	$USp(2N)$
16	$\mathcal{C}(6,0)$	2	**H**	$8+8$	$U(N)$
16	$\mathcal{C}(6,1)$	1	**H**	16	$USp(2N)$
16	$\mathcal{C}(7,0)$	3	**C**	$8+8$	$SO(N)$
16	$\mathcal{C}(8,0)$	0	**R**	$8+8$	$SO(N)$
16	$\mathcal{C}(9,0)$	1	**R**	16	$SO(N)$
32	$\mathcal{C}(6,2)$	0	**H**	$16+16$	$USp(2N)$
32	$\mathcal{C}(7,1)$	2	**C**	$16+16$	$U(N)$
32	$\mathcal{C}(7,2)$	3	**H**	32	$USp(2N)$
32	$\mathcal{C}(8,1)$	3	**C**	$16+16$	$SO(N)$
32	$\mathcal{C}(9,1)$	0	**R**	$16+16$	$SO(N)$
32	$\mathcal{C}(10,0)$	2	**R**	32	$U(N)$
32	$\mathcal{C}(10,1)$	1	**R**	32	$SO(N)$
64	$\mathcal{C}(8,2)$	2	**H**	$32+32$	$U(N)$
64	$\mathcal{C}(9,2)$	3	**C**	$32+32$	$SO(N)$
64	$\mathcal{C}(10,2)$	0	**R**	$32+32$	$SO(N)$

mathematical viewpoint, this correspondence is related to the isomorphism

$$\mathcal{C}(p,q) \cong \mathcal{C}(p-1,q-1) \otimes \mathcal{C}(1,1)\,, \tag{2.13}$$

where we note that $\mathcal{C}(1,1)$ is isomorphic with the real 2×2 matrices.

Inspired by the first correspondence one may investigate a second one between spinors of $SO(p,1)$ and $SO(p,0)$ with $p>1$. Physically this correspondence is relevant when considering relativistic massive spinors in flat Minkowski spacetime of dimension $D=p+1$, which transform in the rest-frame under p-dimensional spin rotations. As the table shows, this correspondence is less systematic and, indeed, an underlying isomorphism for the corresponding Clifford algebras is lacking. The results of the table should therefore be applied with care. In a number of cases the relativistic spinor transforms irreducibly under the nonrelativistic rotation group. In that case the dimension of the automorphism group $H_{\rm R}$ can increase, as it does for $p=3$ and 10, but not for $p=5$ and 9. For $p=3$ (or, equivalently, $D=4$) the implications of the fact that the nonrelativistic automorphism group $USp(2N)$ is bigger than the relativistic one, have already been discussed in Section 2.3. In the remaining cases, $p=4,6,7,8$ (always modulo 8), the relativistic spinor decomposes into two nonrelativistic spinors. Because the number of irreducible spinors is then doubled, the nonrelativistic automorphism group has a tendency to increase, but one should consult the table for specific cases.

The third correspondence relates spinors of $SO(p,2)$ and $SO(p,1)$ with $p>1$. Again the situation depends sensitively on the value for p. In a number of cases (, $p=2,3,4,6$) the spinor dimension is the same for both groups. This can be understood from the fact that the Clifford algebra representations are irreducible with respect to $SO(p,1)$, so that one can always extend the generators of $SO(p,1)$, which are proportional to $\Gamma^{[a}\Gamma^{b]}$, to those of $SO(p,2)$ by including the gamma matrices Γ^a. However, the R-symmetry group is not necessarily the same. For $p=4$ the $SO(4,2)$ spinors allow the R-symmetry group $U(N)$, while for $SO(4,1)$ the R-symmetry group is larger and equal to $USp(2N)$. Therefore theories formulated in flat Minkowski spacetime of dimension $D=3,4,5,6$ can in principle be elevated to anti-de Sitter space. For $D=5$ the R-symmetry reduces to $U(N)$, while for $D=3,4,6$ the R-symmetry remains the same. In the remaining dimensions, $D=2,7,8,9,10$, a single Minkowski spinor can not be elevated to anti-de Sitter space, and one must at least start from an even number of flat Minkowski spinors (so that N is even). For these cases, it is hard to make general statements about the fate of the R-symmetry when moving to anti-de Sitter space and one has to consult Table 9.

The fourth correspondence is again more systematic, as it is based on the isomorphism (2.13). The correspondence relates spinors of $SO(p,2)$

and of $SO(p-1,1)$. For all $p>1$ the spinor dimension differs by one-half while the R-symmetry group remains the same. Observe that $SO(p,2)$ can be regarded as the group of conformal symmetries in a Minkowski space of p dimensions, or as the isometry group of an anti-de Sitter space of dimension $p+1$. This correspondence extends this statement to the level of spinors. It implies that the extension of the Poincaré superalgebra in $D=p$ spacetime dimensions to a superconformal algebra requires a doubling of the number of supercharges. This feature is well known [27] and the two supersymmetries are called Q- and S-supersymmetry. The anticommutator of two S-supersymmetry charges yields the conformal boosts. Both set of charges transform under the R-symmetry group of the Poincaré algebra, which plays a more basic role in the superconformal algebra as its generators appear in the anticommutator of a Q-supersymmetry and an S-supersymmetry charge. In the anti-de Sitter context, the spinor charge is irreducible but has simply twice as many components. We return to the superconformal invariance and related aspects in Section 7.

It is illuminating to exploit some of the previous correspondences and the relations between various supersymmetry representations in the context of the so-called adS/CFT correspondence [28]. We close this section by exhibiting a chain of relationships between various supermultiplets. We start with an $N=4$ supersymmetric gauge theory in $D=4$ spacetime dimensions, whose massless states are characterized as representations of the $SO(2)$ helicity group and the R-symmetry group $SU(4)$[9]. Hence we have a field theory with $Q=16$ supersymmetries, of which only 8 are realized on the massless supermultiplet. This supermultiplet decomposes as follows,

$$(\pm1,\mathbf{1})_{\mathbf{p}}+(0,\mathbf{6})_{\mathbf{p}}+(\tfrac{1}{2},\mathbf{4})_{\mathbf{p}}+(-\tfrac{1}{2},\overline{\mathbf{4}})_{\mathbf{p}}\,, \tag{2.14}$$

where $\mathbf{p}$ indicates the three-momentum, $|\mathbf{p}|$ the energy, and the entries in the parentheses denote the helicity and the $SU(4)$ representation of the states.

Multiplying this multiplet with a similar one, but now with opposite three-momentum $-\mathbf{p}$, yields a multiplet with zero momentum and with mass $M=2|\mathbf{p}|$. As it turns out the helicity states can now be assembled into states that transform under the 3-dimensional rotation group, so that they can be characterized by their spin. The resulting supermultiplet consists of $128+128$ degrees of freedom. While the states of the original multiplet (2.14) were only subject to the helicity group and 8 supersymmetries, the composite multiplet is now a full supermultiplet subject to 16 supersymmetries and the rotation (rather than the helicity) group.

[9] Because this multiplet is CPT self-conjugate, the $U(1)$ subgroup of $U(4)$ coincides with the helicity group and plays no independent role here.

Indeed, inspection shows that this composite multiplet is precisely the $N = 4$ massive multiplet shown in Table 7. In this form the relevant R-symmetry group is extended to $USp(8)$.

It is possible to cast the above product of states into a product of fields of the 4-dimensional gauge theory. One then finds that the spin-2 operators correspond to the energy-momentum tensor, which is conserved (divergence-free) and traceless, so that it has precisely the 5 independent components appropriate for spin-2. The spin-1 operators decompose into 15 conserved vectors, associated with the currents of $SU(4)$, and 6 selfdual antisymmetric tensors. The spin-0 operators are scalar composite operators. Furthermore there are 4 chiral and 4 antichiral vector-spinor operators, which are conserved and traceless (with respect to a contraction with gamma matrices) such that each of them correspond precisely to the 4 components appropriate for spin-$\frac{3}{2}$. These are the supersymmetry currents. Finally there are 20 and 4 chiral and antichiral spin-$\frac{1}{2}$ operators. This is precisely the supermultiplet of currents [22], which couples to the fields of conformal supergravity. Because neither the currents nor the conformal supergravity fields are subject to any field equations (unlike the supersymmetric gauge multiplet from which we started, which constitutes only an on-shell supermultiplet), it forms the basis for a proper off-shell theory of $N = 4$ conformal supergravity [22]. The presence of the traceless and conserved energy-momentum tensor and supersymmetry currents, and of the $SU(4)$ conserved currents, is a consequence of the superconformal invariance of the underlying 4-dimensional gauge theory. The $N = 4$ conformal supergravity theory couples consistently to the $N = 4$ supersymmetric gauge theory. Section 7 will further explain the general setting of superconformal theories that is relevant in this context.

The off-shell $N = 4$ conformal supergravity multiplet in 4 dimensions can also be interpreted as an on-shell supermultiplet in 5 dimensions with 32 supersymmetries. Because of the masslessness, the states are annihilated by half the supercharges and are still classified according to $SO(3)$, which now acts as the helicity group; the R-symmetry group coincides with the $USp(8)$ R-symmetry of the relativistic 5-dimensional supersymmetry algebra. Hence this is the same multiplet that describes $D = 5$ maximal supergravity. This theory has a nonlinearly realized $E_{6(6)}$ invariance whose linearly realized subgroup (which is relevant for the spectrum) equals $USp(8)$.

The latter theory can be gauged (we refer to Sect. 5 for a discussion of this) in which case it can possess an anti-de Sitter ground state. According to Table 9, a fully supersymmetric ground state leads to a $U(4)$ R-symmetry group. As we will discuss in Section 6, anti-de Sitter space leads to "remarkable representations". These are the singletons, which do not have a smooth Poincaré limit because they are associated with possible degrees of freedom living on a 4-dimensional boundary. Because the 5-dimensional anti-de Sitter superalgebra coincides with the 4-dimensional superconformal algebra, the 4-dimensional boundary theory must be consistent with superconformal invariance. Hence it does not come as a surprise that these singleton representations coincide with the supermultiplet of 4-dimensional $N = 4$ gauge theory. This set-up requires the gauge group of 5-dimensional supergravity to be chosen such as to preserve the relevant automorphism group. Therefore the gauge group must be equal to $SO(6) \cong SU(4)/\mathbf{Z}_2$. Indeed, this gauging allows for an anti-de Sitter maximally supersymmetric ground state [29]. Thus, the circle closes.

We stress that the above excursion, linking the various supermultiplets in different dimensions by a series of arguments, is purely based on symmetries. It does not capture the dynamical aspects of the adS/CFT correspondence and has no bearing on the nature of the gauge group in 4 dimensions. At this stage we thus have to content ourselves with the existence of this remarkable chain of correspondences. Many aspects of these correspondences will reappear in later sections.

3 Supergravity

In this section we discuss field theories that are invariant under local supersymmetry. Because of the underlying supersymmetry algebra, the invariance under local supersymmetry implies the invariance under spacetime diffeomorphisms. Therefore these theories are necessarily theories of gravity. We exhibit the initial steps in the construction of a supergravity theory, with and without a cosmological term. Then we concentrate on maximal supergravity theories in various dimensions, their symmetries, and dimensional compactifications on tori. At the end we briefly discuss some of the nonmaximal theories

3.1 Simple supergravity

The first steps in the construction of any supergravity theory are usually based on the observation that local supersymmetry implies the invariance under general coordinate transformation. Therefore one must introduce the fields needed to describe general relativity, namely a vielbein field $e_\mu^{\;a}$ and a spin-connection field $\omega_\mu^{\;ab}$. The vielbein field is nonsingular and its inverse is denoted by $e_a^{\;\mu}$. The vielbein defines a local set of tangent frames of the spacetime manifold, while the spin-connection field is associated with (local) Lorentz transformations of these frames. The world indices, $\mu, \nu, \ldots$, and the tangent space indices, $a, b, \ldots$, both run from 0 to $D-1$. For an introduction to the vielbein formalism we refer to [30]. Furthermore one needs one or several gravitino fields, which carry both a world index and a spinor index and which act as the gauge fields associated with local supersymmetry. For simplicity we only consider a single Majorana gravitino field, denoted by ψ_μ, but this restriction is not essential[10]. Hence any supergravity Lagrangian is expected to contain the Einstein-Hilbert Lagrangian of general relativity and the Rarita-Schwinger Lagrangian for the gravitino field,

$$\kappa^2\,\mathcal{L} = -\tfrac{1}{2}e\,R(\omega) - \tfrac{1}{2}e\bar\psi_\mu\,\Gamma^{\mu\nu\rho}D_\nu(\omega)\psi_\rho + \cdots\,, \tag{3.1}$$

where the covariant derivative on a spinor ψ reads

$$D_\mu(\omega)\psi = \left(\partial_\mu - \tfrac{1}{4}\omega_\mu{}^{ab}\,\Gamma_{ab}\right)\psi\,, \tag{3.2}$$

and $\omega_\mu{}^{ab}$ is the spin-connection field whose definition will be discussed in a sequel. The matrices $\frac{1}{2}\Gamma_{ab} = \frac{1}{4}[\Gamma_a,\Gamma_b]$ are the generators of the Lorentz transformations in spinor space, κ^2 is related to Newton's constant and $e = \det(e_\mu{}^a)$. Observe that the spinor covariant derivative on ψ_μ contains no affine connection, as it should [30].

We note the existence of two covariant tensors, namely the curvature associated with the spin connection $R_{\mu\nu}^{ab}(\omega)$ and the torsion tensor $R_{\mu\nu}^{a}(P)$, which carries this name because it is proportional to the antisymmetric part of the affine connection, $\Gamma^\rho_{[\mu\nu]}$, upon using the vielbein postulate,

$$\begin{aligned} R_{\mu\nu}^{ab}(\omega) &= \partial_\mu\omega_\nu^{\ ab} - \partial_\nu\omega_\mu^{\ ab} + \omega_\mu^{\ ac}\,\omega_{\nu\,c}^{\ \ b} - \omega_\nu^{\ ac}\,\omega_{\mu\,c}^{\ \ b}\,, \\ R_{\mu\nu}^{a}(P) &= D_\mu(\omega)e_\nu^{\ a} - D_\nu(\omega)e_\mu^{\ a}\,. \end{aligned} \tag{3.3}$$

We note that these tensors satisfy the Bianchi identities,

$$D_{[\mu}(\omega)R_{\nu\rho]}^{ab}(\omega) = 0\,, \qquad D_{[\mu}(\omega)R_{\nu\rho]}^{a}(P) + R_{[\mu\nu}^{ab}(\omega)\,e_{\rho]\,b} = 0\,. \tag{3.4}$$

It is suggestive to regard $e_\mu^{\ a}$ and $\omega_\mu^{\ ab}$ as the gauge fields of the Poincaré group. In that context $R(\omega)$ is written as $R(M)$, so that P and M denote the translation and the Lorentz generators of the Poincaré algebra. We will use this notation in later sections when discussing the anti-de Sitter and the conformal algebras. Here we will just use the notation $R(\omega)$ and define its contractions with the inverse vielbeine (related to the Ricci tensor and Ricci scalar) by

$$R_\mu^{a}(e,\omega) = e_b^{\ \nu}\,R_{\mu\nu}^{ab}(\omega)\,, \qquad R(e,\omega) = e_a^{\ \mu}e_b^{\ \nu}\,R_{\mu\nu}^{ab}(\omega)\,. \tag{3.5}$$

[10] For definiteness we consider a generic supergravity theory with one Majorana gravitino with an antisymmetric charge-conjugation matrix C and gamma matrices Γ_a satisfying $C\Gamma_aC^{-1} = -\Gamma_a^{\mathrm{T}}$. Furthermore $\Gamma_\mu = e_\mu{}^a\Gamma_a$.

The spin connection can be treated as an independent field (first-order formalism), which is then solved in terms of its field equations, or it can be fixed from the beginning (second-order formalism), for instance, by imposing the constraint,

$$R^{a}_{\mu\nu}(P) = 0\,. \tag{3.6}$$

Such a constraint is called "conventional" because it expresses one field in terms of other fields in an algebraic fashion. For pure gravity the first- and the second-order formulism lead to the same result. The constraint (3.6) can be solved algebraically and leads to,

$$\omega_\mu^{\ ab}(e) = \tfrac{1}{2} e_\mu^{\ c} (\Omega^{ab}{}_c - \Omega^b{}_c{}^a - \Omega_c{}^{ab})\,, \tag{3.7}$$

where the $\Omega_{ab}{}^c$ are the [illegible],

$$\Omega_{ab}{}^c = e_a^{\ \mu} e_b^{\ \nu}\, (\partial_\mu e_\nu^{\ c} - \partial_\mu e_\nu^{\ c})\,. \tag{3.8}$$

From the spin connection one defines the affine connection by $\Gamma_{\mu\nu}{}^\rho = e_a^{\ \rho}\, D_\mu(\omega) e_\nu^{\ a}$, which ensures the validity of the vielbein postulate. With the zero-torsion value (3.7) for the spin connection, the affine connection becomes equal to the Christoffel symbols and $R_{\mu\nu\rho}{}^\sigma = R^{ab}_{\mu\nu}(\omega)\, e_{\rho\, a}\, e_b^{\ \sigma}$ coincides with the standard Riemann tensor.

The action corresponding to the above Lagrangian is locally supersymmetric up to terms cubic in the gravitino field. The supersymmetry transformations contain the terms,

$$\delta e_\mu{}^a = \tfrac{1}{2}\bar\epsilon\,\Gamma^a \psi_\mu\,, \qquad \delta\psi_\mu = D_\mu(\omega)\epsilon\,, \tag{3.9}$$

where the gravitino variation is the extension to curved spacetime of the spinor gauge invariance of a Rarita-Schwinger field. Extending this Lagrangian to a fully supersymmetric one is not always possible. Usually it requires additional fields of lower spin, whose existence can be inferred from the knowledge of the possible underlying (massless) supermultiplets of states. When the spacetime dimension exceeds eleven, conventional supergravity no longer exists, as we shall discuss in the next section.

Let us now include a cosmological term into the above Lagrangian as well as a suitably chosen masslike term for the gravitino field,

$$\begin{aligned}\mathcal{L} \;=\;& -\tfrac{1}{2} e\, R(e,\omega) - \tfrac{1}{2} e \bar\psi_\mu\, \Gamma^{\mu\nu\rho} D_\nu(\omega)\psi_\rho \\ & + \tfrac{1}{4} g (D-2) e\, \bar\psi_\mu \Gamma^{\mu\nu}\psi_\nu + \tfrac{1}{2} g^2 (D-1)(D-2)\, e + \cdots \end{aligned} \tag{3.10}$$

As it turns out the corresponding action is still locally supersymmetric, up to terms that are cubic in the gravitino field, provided that we introduce an extra term to the transformation rules,

$$\delta e_\mu{}^a = \tfrac{1}{2}\bar\epsilon\,\Gamma^a \psi_\mu\,, \qquad \delta\psi_\mu = \big(D_\mu(\omega) + \tfrac{1}{2} g \Gamma_\mu\big)\,\epsilon\,. \tag{3.11}$$

The Lagrangian (3.10) was first written down in [31] in four space-time dimensions and the correct interpretation of the masslike term was given in [32]. Observe that the variation for ψ_μ may be regarded as a generalized covariant derivative, where ω_μ^{ab} and $e_\mu^{\ a}$ act as gauge fields[11].

Consistency requires that $g\Gamma_\mu\epsilon$ satisfies the same Majorana constraint as ψ_μ and ϵ. With the conventions that we have adopted this implies that g is real. The reality of g has important consequences, as it implies that the cosmological term is of definite sign. Hence supersymmetry does not forbid a cosmological term, but it must be of definite sign (at least, if the ground state is to preserve supersymmetry). This example does not cover all cases, as one does not always have a single Majorana spinor with the specified charge conjugation properties. Nevertheless the conclusion that the cosmological term must have this particular sign remains, unless one accepts "ghosts": fields whose kinetic terms are of the wrong sign. For an early discussion, see [33, 34] and references therein. We should point out that there are situations where a cosmological term is not consistent with supersymmetry. Assuming that the theory has an anti-de Sitter or de Sitter ground state, one may verify whether the Minkowski spinors have the right dimension to enable them to live in these spaces. For instance, a Majorana-Weyl spinor in $D = 10$ spacetime dimensions has only half the number of components as compared to a spinor in (anti-)de Sitter space of the same dimension. Therefore, simple supergravity in $D = 10$ dimensions cannot possibly have (anti-)de Sitter ground states. Such a counting argument does not exclude anti-de Sitter ground states in $D = 11$ spacetime dimensions, because $D = 11$ Lorentz spinors can exist in anti-de Sitter space. Here the argument may be invoked that no relevant supersymmetric extension of the anti-de Sitter algebra exists beyond $D = 7$ dimensions [3], but there are also explicit studies ruling out supersymmetric cosmological terms in 11 dimensions [35].

The Einstein equation corresponding to (3.10) reads (suppressing the gravitino field),

$$R_{\mu\nu} - \tfrac{1}{2}g_{\mu\nu}\,R + \tfrac{1}{2}g^2(D-1)(D-2)\,g_{\mu\nu} = 0\,, \tag{3.12}$$

which implies,

$$R_{\mu\nu} = g^2(D-1)\,g_{\mu\nu}\,, \qquad R = g^2 D(D-1)\,. \tag{3.13}$$

[11] The masslike term in (3.10) is consistent with that interpretation as it can be generated from the Rarita-Schwinger Lagrangian by the same change of the covariant derivative, *i.e.*,

$$-\tfrac{1}{2}e\bar\psi_\mu\Gamma^{\mu\nu\rho}(D_\nu(\omega) + \tfrac{1}{2}g\Gamma_\nu)\psi_\rho\,.$$

Hence we are dealing with a D-dimensional Einstein space. The maximally symmetric solution of this equation is an anti-de Sitter space, whose Riemann curvature equals

$$R_{\mu\nu}{}^{ab} = 2g^2\, e_\mu{}^{[a}\, e_\nu{}^{b]}\,. \tag{3.14}$$

This solution leaves all supersymmetries intact just as flat Minkowski space does. One can verify this directly by considering the supersymmetry variation of the gravitino field and by requiring that it vanishes in the bosonic background. This happens for spinors $\epsilon(x)$ satisfying

$$\big(D_\mu(\omega) + \tfrac{1}{2} g\Gamma_\mu\big)\,\epsilon = 0\,. \tag{3.15}$$

Spinors satisfying this equation are called Killing spinors. Since (3.15) is a first-order differential equation, one expects that it can be solved provided some integrability condition is satisfied. To see this one notes that also $(D_\mu(\omega) + \frac{1}{2} g\Gamma_\mu)(D_\nu(\omega) + \frac{1}{2} g\Gamma_\nu)\epsilon$ must vanish. Antisymmetrizing this expression in μ and ν then yields the (algebraic) integrability condition

$$\Big(-\tfrac{1}{4} R_{\mu\nu}{}^{ab}\,\Gamma_{ab} + \tfrac{1}{2} g^2\,\Gamma_{\mu\nu}\Big)\epsilon = 0\,. \tag{3.16}$$

Multiplication with Γ^ν yields

$$\Big(R_{\mu\nu} - g^2(D-1)\, g_{\mu\nu}\Big)\Gamma^\nu\epsilon = 0\,, \tag{3.17}$$

from which one derives that the Riemann tensor satisfies (3.13). Therefore supersymmetry requires an Einstein space. Requiring full supersymmetry, so that (3.15) holds for any spinor ϵ, implies (3.14) so that the spinor ϵ must live in anti-de Sitter space.

Hence we have seen that supersymmetry can be realized in anti-de Sitter space. We will return to this issue later in Section 6, where we discuss the (super)multiplet structure in anti-de Sitter space. We stress once more that, in this section, we have restricted ourselves to the graviton-gravitino sector. To construct the full theory usually requires more fields and important restrictions arise on the dimensionality of spacetime. For instance, while minimal supergravity in $D = 4$ dimensions does not require additional fields, in $D = 11$ dimensions an additional antisymmetric gauge field is necessary. The need for certain extra fields can be readily deduced from the underlying massless supermultiplets, which were extensively discussed in the previous section.

3.2 Maximal supersymmetry and supergravity

In Section 2 we restricted ourselves to supermultiplets based on $Q \leq 32$ supercharge components. From the general analysis it is clear that increasing the number of supercharges leads to higher and higher helicity representations. For instance, the maximal helicity, $|\lambda_{\max}|$, of a massless supermultiplet in $D = 4$ spacetime dimensions is larger than or equal to $\frac{1}{16}Q$. Therefore, when $Q > 8$ we have $|\lambda_{\max}| \geq 1$, so that theories for these multiplets must include vector gauge fields. When $Q > 16$ we have $|\lambda_{\max}| \geq \frac{3}{2}$, so that the theory should contain Rarita-Schwinger fields. In view of the supersymmetry algebra an interacting supersymmetric theory of this type should contain gravity, so that in this case we must include $\lambda = 2$ states for the graviton. Beyond $Q = 32$ one is dealing with states of helicity $\lambda > 2$. Those are described by gauge fields that are [illegible] Lorentz tensors. Symmetric tensor gauge fields for arbitrary helicity states can be constructed (in $D = 4$ dimensions, see, for instance [36]). However, it turns out that symmetric gauge fields cannot consistently couple, neither to themselves nor to other fields. An exception is the graviton field, which can interact with itself as well as with low-spin matter, but not with other fields of the same spin [37]. By consistent, we mean that the respective gauge invariances of the higher-spin fields (or appropriate deformations thereof) cannot be preserved at the interacting level. Most of the search for interacting higher-spin fields was performed in 4 spacetime dimensions [38], but in higher dimensional spacetimes one expects to arrive at the same conclusions, because otherwise, upon dimensional reduction, these theories would give rise to theories that are consistent in $D = 4$. There is also direct evidence in $D = 3$, where graviton and gravitini fields do not describe dynamic degrees of freedom. Hence, one can write down supergravity theories based on a graviton field and an arbitrary number of gravitino fields, which are topological. However, when coupling matter to this theory in the form of scalars and spinors, the theory does not support more than 32 supercharges. Beyond $Q = 16$ there are four unique theories with $Q = 18, 20, 24$ and 32 [8].

Hence the conclusion is that there is a restriction on the number Q of independent supersymmetries, as for $Q > 32$ no interacting field theories seem to exist. There have been many efforts to circumvent this bound of $Q = 32$ supersymmetries. It seems clear that one needs a combination of the following ingredients in order to do this (for a review, see [illegible] [39]): (i) an infinite tower of higher-spin gauge fields; (ii) interactions that are inversely proportional to the cosmological constant; (iii) extensions of the super-Poincaré or the super-de Sitter algebra with additional fermionic and bosonic charges. Indeed explicit theories have been constructed which demonstrate this. However, conventional supergravity theories are [illegible] of this kind. This is the reason why we avoided ([illegible] in Table 5) to list supermultiplets with states transforming in higher-helicity representations. The fact that an infinite number of fields can cure certain inconsistencies is by itself not new. While a massive spin-2 field cannot be coupled to gravity,

Table 10. Bosonic field content for maximal supergravities. The $p = 3$ gauge field in $D = 10$B has a self-dual field strength. The representations [1] and [28] (in $D = 8, 4$, respectively) are extended to $U(1)$ and $SU(8)$ representations through duality transformations on the field strengths. These transformations cannot be represented on the vector potentials. In $D = 3$ dimensions, the graviton does not describe propagating degrees of freedom. For $p > 0$ the fields can be assigned to representations of a bigger group than H_{R}. This will be discussed in due course.

D	H_{R}	graviton	$p=-1$	$p=0$	$p=1$	$p=2$	$p=3$
11	1	1	0	0	0	1	0
10A	1	1	1	1	1	1	0
10B	$SO(2)$	1	2	0	2	0	1^*
9	$SO(2)$	1	$2+1$	$2+1$	2	1	
8	$U(2)$	1	$5+1+\bar{1}$	$3+\bar{3}$	3	[1]	
7	$USp(4)$	1	14	10	5		
6	$USp(4)$ $\times USp(4)$	1	(5, 5)	(4, 4)	$(5,1)$ $+(1,5)$		
5	$USp(8)$	1	42	27			
4	$U(8)$	1	$35+\overline{35}$	[28]			
3	$SO(16)$	1	128				

the coupling of an infinite number of them can be consistent, as can be seen in Kaluza-Klein theory.

In this section we review the maximal supergravities in various dimensions. These theories have $Q = 32$ supersymmetries and we restrict our discussion to $3 \leq D \leq 11$. The highest dimension $D = 11$ is motivated by the fact that spinors have more that 32 components in flat Minkowski space for spacetime dimensions $D > 11$. Observe, however, that this argument assumes D-dimensional Lorentz invariance. As was stressed in [40,41], there are scenarios based on spacetime dimensions higher than $D = 11$, where the extra dimensions can not uniformly decompactify so that the no-go theorem is avoided. The fact that no uniform decompactification is possible is closely related to the T-duality between winding and momentum states that one knows from string theory.

The bosonic fields always comprise the metric tensor for the graviton and a number of $(p+1)$-rank antisymmetric gauge fields. For the antisymmetric gauge fields, it is unclear whether to choose a $(p+1)$-rank gauge field or its dual $(D-3-p)$-rank partner, but it turns out that the interactions often prefer the rank of the gauge field to be as small as possible. Therefore,

Table 11. Fermionic field content for maximal supergravities. For $D = 5, 6, 7$ the fermion fields are counted as symplectic Majorana spinors. For $D = 4, 8$ we include both chiral and antichiral spinor components, which transform in conjugate representations of $H_{\rm R}$. In $D = 3$ dimensions the gravitino does not describe propagating degrees of freedom.

D	$H_{\rm R}$	gravitini	spinors
11	1	1	0
10A	1	$1+1$	$1+1$
10B	$SO(2)$	2	2
9	$SO(2)$	2	$2+2$
8	$U(2)$	$2+\bar{2}$	$2+\bar{2}+4+\bar{4}$
7	$USp(4)$	4	16
6	$USp(4)\times USp(4)$	$(4,1)+(1,4)$	$(4,5)+(5,4)$
5	$USp(8)$	8	48
4	$U(8)$	$8+\bar{8}$	$56+\overline{56}$
3	$SO(16)$	16	128

in Table 10, we restrict ourselves to $p \leq 3$, as in $D = 11$ dimensions, $p = 3$ and $p = 4$ are each other's dual conjugates. This table presents all the field configurations for maximal supergravity in various dimensions. Obviously, the problematic higher-spin fields are avoided, because the only symmetric gauge field is the one describing the graviton. In Table 11 we also present the fermionic fields, always consisting of gravitini and simple spinors. All these fields are classified as representations of the R-symmetry group $H_{\rm R}$. Note that the simplest versions of supergravity (which depend on no other coupling constant than Newton's constant) are manifestly invariant under $H_{\rm R}$. Actually, as we will explain in a sequel, the maximal supergravity theories have symmetry groups that are much larger than $H_{\rm R}$.

3.3 $D = 11$ supergravity

Supergravity in 11 spacetime dimensions is based on an "elfbein" field $E_M^{\;A}$, a Majorana gravitino field Ψ_M and a 3-rank antisymmetric gauge field C_{MNP}. With chiral (2, 0) supergravity in 6 dimensions, it is the only $Q \geq 16$ supergravity theory without a scalar field. Its Lagrangian can be written as follows [11],

$$\mathcal{L}_{11} = \frac{1}{\kappa_{11}^2}\Big[-\tfrac{1}{2}E\,R(E,\Omega) - \tfrac{1}{2}E\bar\Psi_M\Gamma^{MNP}D_N(\Omega)\Psi_P - \tfrac{1}{48}E(F_{MNPQ})^2 \\ -\tfrac{1}{3456}\sqrt{2}\,\varepsilon^{MNPQRSTUVWX}\,F_{MNPQ}\,F_{RSTU}\,C_{VWX} \\ -\tfrac{1}{192}\sqrt{2}E\Big(\bar\Psi_R\Gamma^{MNPQRS}\Psi_S + 12\,\bar\Psi^M\Gamma^{NP}\Psi^Q\Big)F_{MNPQ} + \cdots\Big], \tag{3.18}$$

where the ellipses denote terms of order Ψ^4, $E = \det E_M^A$ and $\Omega_M{}^{AB}$ denotes the spin connection. The supersymmetry transformations are equal to

$$\begin{aligned} \delta E_M^A &= \tfrac{1}{2}\,\bar\epsilon\,\Gamma^A\Psi_M\,, \\ \delta C_{MNP} &= -\tfrac{1}{8}\sqrt{2}\,\bar\epsilon\,\Gamma_{[MN}\Psi_{P]}\,, \\ \delta\Psi_M &= D_M(\hat\Omega)\,\epsilon + \tfrac{1}{288}\sqrt{2}\,\Big(\Gamma_M{}^{NPQR} - 8\,\delta_M^N\,\Gamma^{PQR}\Big)\,\epsilon\,\hat F_{NPQR}\,. \end{aligned} \tag{3.19}$$

Here the covariant derivative is covariant with respect to local Lorentz transformation

$$D_M(\Omega)\,\epsilon = \Big(\partial_M - \tfrac{1}{4}\Omega_M{}^{AB}\Gamma_{AB}\Big)\epsilon\,, \tag{3.20}$$

and $\hat F_{MNPQ}$ is the supercovariant field strength

$$\hat F_{MNPQ} = 24\,\partial_{[M}C_{NPQ]} + \tfrac{3}{2}\sqrt{2}\,\bar\Psi_{[M}\Gamma_{NP}\Psi_{Q]}\,. \tag{3.21}$$

The supercovariant spin connection is the solution of the following equation,

$$D_{[M}(\hat\Omega)\,E_{N]}^A - \tfrac{1}{4}\bar\Psi_M\Gamma^A\Psi_N = 0\,. \tag{3.22}$$

The left-hand side is the supercovariant torsion tensor.

Note the presence of a Chern-Simons-like term $F\wedge F\wedge C$ in the Lagrangian, so that the action is only invariant under tensor gauge transformations up to surface terms. We also wish to point out that the quartic-Ψ terms can be included into the Lagrangian (3.18) by replacing the spin-connection field Ω by $(\Omega+\hat\Omega)/2$ in the covariant derivative of the gravitino kinetic term and by replacing F_{MNPQ} in the last line by $(\hat F_{MNPQ} + F_{MNPQ})/2$. These substitutions ensure that the field equations corresponding to (3.18) are supercovariant. The Lagrangian is derived in the context of the so-called "1.5-order" formalism, in which the spin connection is defined as a dependent field determined by its (algebraic) equation of motion, whereas its supersymmetry variation in the action is treated as if it were an independent field [42].

We have the following bosonic field equations and Bianchi identities,

$$\begin{aligned} R_{MN} &= \tfrac{1}{72}g_{MN}\,F_{PQRS}F^{PQRS} - \tfrac{1}{6}F_{MPQR}\,F_N{}^{PQR}\,, \\ \partial_M\Big(E\,F^{MNPQ}\Big) &= \tfrac{1}{1152}\sqrt{2}\,\varepsilon^{NPQRSTUVWXY}F_{RSTU}\,F_{VWXY}\,, \\ \partial_{[M}F_{NPQR]} &= 0\,, \end{aligned} \tag{3.23}$$

which no longer depend explicitly on the antisymmetric gauge field. An alternative form of the second equation is [43]

$$\partial_{[M} H_{NPQRSTU]} = 0\,, \tag{3.24}$$

where $H_{MNPQRST}$ is the dual field strength,

$$H_{MNPQRST} = \frac{1}{7!} E\, \varepsilon_{MNPQRSTUVWX} F^{UVWX} - \tfrac{1}{2}\sqrt{2}\, F_{[MNPQ}\, C_{RST]}\,. \tag{3.25}$$

One could imagine that the third equation of (3.23) and (3.24) receive contributions from charges that would give rise to source terms on the right-hand side of the equations. These charges are associated with the "flux"-integral of $H_{MNPQRST}$ and F_{MNPQ} over the boundary of an 8- and a 5-dimensional spatial volume, respectively. In analogy with the Maxwell theory, the integral $\oint H$ may be associated with electric flux and the integral $\oint F$ with magnetic flux. The spatial volumes are orthogonal to a $p = 2$ and a $p = 5$ brane configuration, respectively, and the corresponding charges are 2- and 5-rank Lorentz tensors. These are just the charges that can appear as central charges in the supersymmetry algebra (2.12). Solutions of 11-dimensional supergravity that contribute to these charges were considered in [44–46].

Finally, the constant $1/\kappa_{11}^2$ in front of the Lagrangian (3.18), which carries dimension $[\text{length}]^{-9} \sim [\text{mass}]^9$, is undetermined and depends on fixing some length scale. To see this consider a continuous rescaling of the fields,

$$E_M^{\,A} \to \mathrm{e}^{-\alpha} E_M^{\,A}\,, \qquad \Psi_M \to \mathrm{e}^{-\alpha/2} \Psi_M\,, \qquad C_{MNP} \to \mathrm{e}^{-3\alpha} C_{MNP}\,. \tag{3.26}$$

Under this rescaling the Lagrangian changes according to

$$\mathcal{L}_{11} \to \mathrm{e}^{-9\alpha} \mathcal{L}_{11}\,. \tag{3.27}$$

This change can then be absorbed into a redefinition of κ_{11}[12],

$$\kappa_{11}^2 \to \mathrm{e}^{-9\alpha} \kappa_{11}^2\,. \tag{3.28}$$

This simply means that the Lagrangian depends on only one dimensional coupling constant, namely κ_{11}. The same situation is present in many other supergravity theories. Concentrating on the Einstein-Hilbert action in D spacetime dimensions, the corresponding scaling property is

$$g_{\mu\nu}^D \to \mathrm{e}^{-2\alpha} g_{\mu\nu}^D\,, \qquad \mathcal{L}_D \to \mathrm{e}^{(2-D)\alpha} \mathcal{L}_D\,, \qquad \kappa_D^2 \to \mathrm{e}^{(2-D)\alpha} \kappa_D^2\,. \tag{3.29}$$

[12]Note that the rescalings also leave the supersymmetry transformation rules unchanged, provided the supersymmetry parameter ϵ is changed accordingly.

Of course, this implies that the physical value of Newton's constant, does not necessarily coincide with the parameter κ_D^2 in the Lagrangian but it also depends on the precise value adopted for the (flat) metric in the ground state of the theory.

3.4 Dimensional reduction and hidden symmetries

The maximal supergravities in various dimensions are related by dimensional reduction. In this reduction some of the spatial dimensions are compactified on a hypertorus and one retains only the fields that do not depend on the torus coordinates. This corresponds to the theory one obtains when the size of the torus is shrunk to zero. A subset of the gauge symmetries associated with the compactified dimensions survive as internal symmetries. The aim of the present discussion here is to elucidate a number of features related to these symmetries, mainly in the context of the reduction of $D = 11$ supergravity to $D = 10$ dimensions.

We denote the compactified coordinate by x^{10} which now parameterizes a circle of length L[13]. The fields are thus decomposed in a Fourier series as periodic functions in x^{10} on the interval $0 \leq x^{10} \leq L$. This results in a spectrum of massless modes and an infinite tower of massive modes with masses inversely proportional to the circle length L. The massless modes form the basis of the lower-dimensional supergravity theory. Because a toroidal background does not break supersymmetry, the resulting supergravity has the same number of supersymmetries as the original one. For compactifications on less trivial spaces than the hypertorus, this is usually not the case and the number of independent supersymmetries will be reduced. Fully supersymmetric compactifications are rare. For instance, 11-dimensional supergravity can be compactified to a 4-dimensional maximally symmetric spacetime in only two ways such that all supersymmetries remain unaffected [47]. One is the compactification on a torus T^7, the other one the compactification on a sphere S^7. In the latter case the resulting 4-dimensional supergravity theory acquires a cosmological term.

In the formulation of the compactified theory, it is important to decompose the higher-dimensional fields in such a way that they transform covariantly under the lower-dimensional gauge symmetries and under diffeomorphisms of the lower-dimensional spacetime. This ensures that various complicated mixtures of massless modes with the tower of massive modes will be avoided. It is a key element in ensuring that solutions of the lower-dimensional theory remain solutions of the original higher-dimensional one, which is an obvious requirement for having consistent truncations to the

[13]Throughout these lectures we enumerate spacetime coordinates by $0, 1, \ldots, D-1$.

massless states. Another point of interest concerns the nature of the massive supermultiplets. Because these originate from supermultiplets that are massless in higher dimensions, they are 1/2-BPS multiplets which are shortened by the presence of central charges corresponding to the momenta in the compactified dimension. Implications of these BPS supermultiplets will be discussed in more detail in Section 3.6.

The emergence of new internal symmetries in theories that originate from a higher-dimensional setting, is a standard feature of Kaluza-Klein theories [48]. Following the discussion in [49] we distinguish between symmetries that have a direct explanation in terms of the higher dimensional symmetries, and symmetries whose origin is obscure from a higher-dimensional viewpoint. Let us start with the symmetries associated with the metric tensor. The 11-dimensional metric can be decomposed according to

$$\mathrm{d}s^2 = g_{\mu\nu}\, \mathrm{d}x^\mu \mathrm{d}x^\nu + \mathrm{e}^{4\phi/3}(\mathrm{d}x^{10} + V_\mu \mathrm{d}x^\mu)(\mathrm{d}x^{10} + V_\nu \mathrm{d}x^\nu)\,, \tag{3.30}$$

where the indices μ, ν label the 10-dimensional coordinates and the factor multiplying ϕ is for convenience later on. The massless modes correspond to the x^{10}-independent parts of the 10-dimensional metric $g_{\mu\nu}$, the vector field V_μ and the scalar ϕ. Here the x^{10}-independent component of V_μ acts as a gauge field associated with reparametrizations of the circle coordinate x^{10} with an arbitrary function $\xi(x)$ of the 10 remaining spacetime coordinates x^μ. Specifically, we have $x^{10} \to x^{10} - \xi(x)$ and $x^\mu \to x^\mu$, leading to

$$V_\mu(x) \to V_\mu(x) + \partial_\mu \xi(x)\,. \tag{3.31}$$

The massive modes, which correspond to the nontrivial Fourier modes in x^{10}, couple to this gauge field with a charge that is a multiple of

$$e_{\rm KK} = \frac{2\pi}{L}\cdot \tag{3.32}$$

Another symmetry of the lower-dimensional theory is more subtle to identify[14]. In the previous section we noted the existence of certain scale transformations of the $D = 11$ fields, which did not leave the theory invariant but could be used to adjust the coupling constant κ_{10}. In the compactified situation we can also involve the compactification length into the dimensional scaling. The integration over x^{11} introduces an overall factor L in the action (we do not incorporate any L-dependent normalizations in the

[14]There are various discussions of this symmetry in the literature. Its existence in 10-dimensional supergravity was noted long ago (see, *e.g.* [9, 50]) and an extensive discussion can be found in [18]. Our derivation here was alluded to in [49], which deals with isometries in $N = 2$ supersymmetric Maxwell-Einstein theories in $D = 5, 4$ and 3 dimensions.

Fourier sums, so that the 10-dimensional and the 11-dimensional fields are directly proportional). Therefore, the coupling constant that emerges in the 10-dimensional theory equals

$$\frac{1}{\kappa_{10}^2} = \frac{L}{\kappa_{11}^2}\,, \tag{3.33}$$

and is of dimension [mass]^8. However, because of the invariance under diffeomorphisms, L itself has no intrinsic meaning. It simply expresses the length of the x^{10}-periodicity interval, which depends on the coordinatization. Stated differently, we can reparameterize x^{10} by some diffeomorphism, as long as we change L accordingly. In particular, we may rescale L according to

$$L \to \mathrm{e}^{-9\alpha} L\,, \tag{3.34}$$

corresponding to a reparametrization of the 11-th coordinate,

$$x^{10} \to \mathrm{e}^{-9\alpha} x^{10}\,, \tag{3.35}$$

so that κ_{10} remains invariant. Consequently we are then dealing with a [illegible] of the Lagrangian.

In the effective 10-dimensional theory, the scale transformations (3.26) are thus suitably combined with the diffeomorphism (3.35) to yield an invariance of the Lagrangian. For the fields corresponding to the 11-dimensional metric, these combined transformations are given by[15]

$$e_\mu^a \to \mathrm{e}^{-\alpha} e_\mu^a\,, \qquad \phi \to \phi + 12\alpha\,, \qquad V_\mu \to \mathrm{e}^{-9\alpha} V_\mu\,. \tag{3.36}$$

The tensor gauge field C_{MNP} decomposes into a 3- and a 2-rank tensor in 10 dimensions, which transform according to

$$C_{\mu\nu\rho} \to \mathrm{e}^{-3\alpha} C_{\mu\nu\rho}\,, \qquad C_{11\mu\nu} \to \mathrm{e}^{6\alpha} C_{11\mu\nu}\,. \tag{3.37}$$

The presence of the above scale symmetry is confirmed by the resulting 10-dimensional Lagrangian for the massless (, x^{10}-independent) modes. Its purely bosonic terms read

$$\begin{aligned} \mathcal{L}_{10} \;=\; \frac{1}{\kappa_{10}^2}\Big[& -\tfrac{1}{2} e\, \mathrm{e}^{2\phi/3} R(e,\omega) - \tfrac{1}{8} e\, \mathrm{e}^{2\phi} (\partial_\mu V_\nu - \partial_\nu V_\mu)^2 \\ & - \tfrac{1}{48} e\, \mathrm{e}^{2\phi/3} (F_{\mu\nu\rho\sigma})^2 - \tfrac{3}{4} e\, \mathrm{e}^{-2\phi/3} (H_{\mu\nu\rho})^2 \\ & + \tfrac{1}{1152}\sqrt{2}\, \varepsilon^{\mu_1 - \mu_{10}}\, C_{11\mu_1\mu_2}\, F_{\mu_3\mu_4\mu_5\mu_6}\, F_{\mu_7\mu_8\mu_9\mu_{10}} \Big]\,, \end{aligned} \tag{3.38}$$

[15] Note that these transformations apply uniformly to all Fourier modes, as those depend on x^{10}/L which is insensitive to the scale transformation. This does not imply that the Lagrangian remains invariant when retaining the higher Fourier modes, because the Kaluza-Klein charges (3.32) depend explicitly on L. This issue will be relevant in Section 3.6.

where $H_{\mu\nu\rho} = 6\,\partial_{[\mu}C_{\nu\rho]11}$ is the field strength tensor belonging to the 2-rank tensor gauge field.

The above example exhibits many of the characteristic features of dimensional reduction and of the symmetries that emerge as a result. When reducing to lower dimension one can follow the same procedure a number of times, consecutively reducing the dimension by unit steps, or one can reduce at once to lower dimensions. Before continuing our general discussion, let us briefly discuss an example of the latter based on gravity coupled to an antisymmetric tensor gauge field in $D+n$ spacetime dimensions,

$$\mathcal{L} \propto -\tfrac{1}{2}E\,R - \tfrac{9}{4}E\,(\partial_{[M}B_{NP]})^2\,. \tag{3.39}$$

After compactification on a torus T^n, the fields that are independent of the torus coordinates remain massless fields in D dimensions: the graviton, one tensor gauge field, $2n$ Abelian vector gauge fields, and n^2 scalar fields. The scalar fields originate from the metric and the antisymmetric gauge field with both indices taking values in T^n, so that they are parametrized by a symmetric tensor g_{ij} and an antisymmetric tensor B_{ij}. The diffeomorphisms acting on the torus coordinates x^i which are linear in x^i, , $x^i \to \mathsf{O}^i{}_j\,x^j$, act on g_{ij} and B_{ij} according to $g \to \mathsf{O}^T g\,\mathsf{O}$ and $B \to \mathsf{O}^T B\,\mathsf{O}$. The matrices O generate the group $GL(n)$, which can be regarded as a generalization of the scale transformations (3.36) and (3.37). The group $GL(n)$ contains the rotation group $SO(n)$; its remaining part depends on $\frac{1}{2}n(n+1)$ parameters, exactly equal to the number of independent fields g_{ij}. Special tensor gauge transformations with parameters proportional to $\Lambda_{ij}\,x^j$ induce a shift of the massless scalars B_{ij} proportional to the constants $\Lambda_{[ij]}$. There are thus $\frac{1}{2}n(n-1)$ independent shift transformations, so that, in total, we have now identified $\frac{1}{2}n(3n-1)$ isometries, which act transitively on the manifold (they leave no point on the manifold invariant). Therefore the manifold is homogeneous (for a discussion of such manifolds, see Sect. 4.1). However, it turns out that there exist $\frac{1}{2}n(n-1)$ additional isometries, whose origin is directly related to the higher-dimensional context, and which combine with the previous ones to generate the group $SO(n,n)$. The homogeneous space can then be identified as the coset space $SO(n,n)/(SO(n)\times SO(n))$.

According to the above, 11-dimensional supergravity reduced on a hypertorus thus leads to a Lagrangian for the massless sector in lower dimensions (the massive sector is discussed in Sect. 3.6), which exhibits a number of invariances that find their origin in the diffeomorphisms and gauge transformations related to the torus coordinates. As already explained, one must properly account for the periodicity intervals of the torus coordinates x^i, but the action for the massless fields remains invariant under continuous $GL(n)$ transformations. Furthermore, all the scalars that emerge from dimensional reduction of gauge fields are subject to constant shift transformations. These scalars and the scalars originating from the metric transform transitively under the isometry group. Since 11-dimensional supergravity has itself no scalar fields, the rank of the resulting symmetry group in lower dimensions is equal to the rank of $GL(n)$, and thus to the number of compactified dimensions, , $\mathbf{r} = 11 - D$, where D is the spacetime dimension to which we reduce. In general these extra symmetries are not necessarily symmetries of the full action. In even dimensions, the symmetries may not leave the Lagrangian, but only the field equations, invariant. The reason for this is that the isometries may act by means of duality transformations on field strengths associated with antisymmetric tensor gauge fields of rank $\frac{1}{2}D - 1$ which cannot be implemented on the gauge fields themselves. In 4 dimensions this phenomenon is known as electric-magnetic duality (for a recent review, see [51]); for $D = 6$ we refer to [52]. For supergravity, it is easy to see that the scalar manifold (as well as the rest of the theory) possesses additional symmetries beyond the ones that follow from higher dimensions, because the latter do not yet incorporate the full R-symmetry group of the underlying supermultiplet. We expect that H_{R} is also realized as a symmetry, because the maximal supergravity theory that one obtains from compactification on a hypertorus has no additional coupling constants (beyond Newton's constant) which could induce R-symmetry breaking. Therefore we expect that the target space for the scalar fields is an homogeneous space, with an isometry group whose generators belong to a solvable subalgebra associated with the shift transformations, to the subalgebra of $GL(11 - D)$ scale transformations and/or to the subalgebra associated with H_{R}. Of course, these subalgebras will partly overlap. Usually a counting argument (of the type first used in [53]) then readily indicates what the structure is of the corresponding homogeneous space that is parametrized by the scalar fields. In Table 12 we list the isometry group G and the isotropy group H_{R} of these scalar manifolds for maximal supergravity in dimensions $3 \leq D \leq 11$. Earlier versions of such tables can, for instance, be found in [9, 50].

A more recent discussion of these isometry groups from the perspectives of string theory and M-theory can be found in, for example [24, 54]. Here we merely stress a number of characteristic features of the group G. One of them is that H_{R} is always the maximal compact subgroup of G. As we mentioned already, another (noncompact) subgroup is the group $GL(11 - D)$, associated with the reduction on an $(11 - D)$-dimensional torus. Yet another subgroup is $SL(2) \times SO(n, n)$, where $n = 10 - D$. This group, which emerges for $D < 10$ can be understood within the string perspective; $SL(2)$ is the S-duality group and $SO(n, n)$ is the T-duality group. It also follows from the toroidal compactifications of IIB supergravity, which has a manifest $SL(2)$ in $D = 10$ dimensions. The group $SO(n, n)$ is associated with

Table 12. Homogeneous scalar manifolds G/H for maximal supergravities in various dimensions. The type-IIB theory cannot be obtained from reduction of 11-dimensional supergravity and is included for completeness. The difference of the dimensions of G and H equals the number of scalar fields, listed in Table 10.

D	G	H	$\dim[G] - \dim[H]$
11	1	1	$0 - 0 = 0$
10A	$SO(1,1)/\mathbf{Z}_2$	1	$1 - 0 = 1$
10B	$SL(2)$	$SO(2)$	$3 - 1 = 2$
9	$GL(2)$	$SO(2)$	$4 - 1 = 3$
8	$E_{3(+3)} \sim SL(3)\times SL(2)$	$U(2)$	$11 - 4 = 7$
7	$E_{4(+4)} \sim SL(5)$	$USp(4)$	$24 - 10 = 14$
6	$E_{5(+5)} \sim SO(5,5)$	$USp(4)\times USp(4)$	$45 - 20 = 25$
5	$E_{6(+6)}$	$USp(8)$	$78 - 36 = 42$
4	$E_{7(+7)}$	$SU(8)$	$133 - 63 = 70$
3	$E_{8(+8)}$	$SO(16)$	$248 - 120 = 128$

the invariance of toroidal compactifications that involve the metric and an antisymmetric tensor field ((3.39)).

Here we should add that it is generally possible to realize the group $H_{\rm R}$ as a symmetry of the Lagrangian. The corresponding connections are then composite connections, governed by the Cartan-Maurer equations. In such a formulation most fields (in particular, the fermions) do not transform under the group G, but only under the local $H_{\rm R}$ group. The scalars transform linearly under both the rigid duality group as well as under the local $H_{\rm R}$ group; the gauge fields cannot transform under the local group $H_{\rm R}$, as this would be in conflict with their own gauge invariance. After fixing a gauge, the G-transformations become realized nonlinearly (we discuss such nonlinear realizations in detail in Sects. 4 and 5). The fields which initially transform only under the local $H_{\rm R}$ group, will now transform under the duality group G through field-dependent $H_{\rm R}$ transformations. This phenomenon is also realized for the central charges, which transform under the group $H_{\rm R}$ as we have shown in Table 8. We discuss some of the consequences for the central charges and the BPS states in Section 3.6.

3.5 Frames and field redefinitions

The Lagrangian (3.38) does not contain the standard Einstein-Hilbert term for gravity, while a standard kinetic term for the scalar field ϕ is lacking. This does not pose a serious problem. In this form the gravitational field and the scalar field are entangled and one has to deal with the scalar-graviton system as a whole. To separate the scalar and gravitational degrees of freedom, one applies a so-called Weyl rescaling of the metric $g_{\mu\nu}$ by an appropriate function of ϕ. In the case that we include the massive modes, this rescaling may depend on the extra coordinate x^{10}. In the context of Kaluza-Klein theory this factor is known as the "warp factor". For these lectures two different Weyl rescalings are particularly relevant, which lead to the so-called Einstein and to the string frame, respectively. They are defined by

$$e_\mu^a = \mathrm{e}^{-\phi/12}\,[e_\mu^a]^{\mathrm{Einstein}}\,, \qquad e_\mu^a = \mathrm{e}^{-\phi/3}\,[e_\mu^a]^{\mathrm{string}}\,. \tag{3.40}$$

We already stressed that the the compactification length L is just a parameter length with no intrinsic meaning as a result of the fact that one can always apply general coordinate transformations which involve x^{10}. Of course, one may also consider the [illegible], which in the metric specified by (3.30) is equal to $L\,\exp[2\langle\phi\rangle/3]$. In the Einstein frame, the geodesic length of the 11-th dimension is invariant under the $SO(1,1)$ transformations.

After applying the first rescaling (3.40) to the Lagrangian (3.38) one obtains the Lagrangian in the Einstein frame. This frame is characterized by a standard Einstein-Hilbert term and by a graviton field that is invariant under the scale transformations (3.36, 3.37). The corresponding Lagrangian reads[16]

$$\begin{aligned}\mathcal{L}_{10}^{\mathrm{Einstein}} \;=\; \frac{1}{\kappa_{10}^2}\Big[& e\Big[-\tfrac{1}{2}R(e,\omega) - \tfrac{1}{4}(\partial_\mu\phi)^2\Big] - \tfrac{1}{8}e\,\mathrm{e}^{3\phi/2}(\partial_\mu V_\nu - \partial_\nu V_\mu)^2 \\ & -\tfrac{3}{4}e\,\mathrm{e}^{-\phi}(H_{\mu\nu\rho})^2 - \tfrac{1}{48}e\,\mathrm{e}^{\phi/2}(F_{\mu\nu\rho\sigma})^2 \\ & +\tfrac{1}{1152}\sqrt{2}\,\varepsilon^{\mu_1-\mu_{10}}\,C_{11\mu_1\mu_2}\,F_{\mu_3\mu_4\mu_5\mu_6}\,F_{\mu_7\mu_8\mu_9\mu_{10}}\Big]\,.\end{aligned} \tag{3.41}$$

Supergravity theories are usually formulated in this frame, where the isometries of the scalar fields do not act on the graviton.

[16]Note that under a local scale transformation $e_\mu^a \to \mathrm{e}^{\Lambda}e_\mu^a$, the Ricci scalar in D dimensions changes according to

$$R \to \mathrm{e}^{-2\Lambda}\Big[R + 2(D-1)D^\mu\partial_\mu\Lambda + (D-1)(D-2)g^{\mu\nu}\,\partial_\mu\Lambda\,\partial_\nu\Lambda\Big]\,.$$

Observe that gauge fields cannot be redefined by these local scale transformations because this would interfere with their own gauge invariance.

The second rescaling (3.40) leads to the Lagrangian in the string frame,

$$\begin{aligned}\mathcal{L}_{10}^{\text{string}} &= \frac{1}{\kappa_{10}^2}\Big[e\,\mathrm{e}^{-2\phi}\Big[-\tfrac{1}{2}R(e,\omega)+2(\partial_\mu\phi)^2-\tfrac{3}{4}(H_{\mu\nu\rho})^2\Big]\\ &\qquad -\tfrac{1}{8}e\,(\partial_\mu V_\nu-\partial_\nu V_\mu)^2-\tfrac{1}{48}e\,(F_{\mu\nu\rho\sigma})^2\\ &\qquad +\tfrac{1}{1152}\sqrt{2}\,\varepsilon^{\mu_1-\mu_{10}}\,C_{11\mu_1\mu_2}\,F_{\mu_3\mu_4\mu_5\mu_6}\,F_{\mu_7\mu_8\mu_9\mu_{10}}\Big]\,. \qquad (3.42)\end{aligned}$$

This frame is characterized by the fact that R and $(H_{\mu\nu\rho})^2$ have the same coupling to the scalar ϕ, or, equivalently, that $g_{\mu\nu}$ and $C_{11\mu\nu}$ transform with equal weights under the scale transformations (3.36, 3.37). In string theory ϕ coincides with the dilaton field that couples to the topology of the worldsheet and whose vacuum-expectation value defines the string coupling constant according to $g_{\mathrm{s}}=\exp(\langle\phi\rangle)$. The significance of the dilaton factors in the Lagrangian above is well known. The metric $g_{\mu\nu}$, the antisymmetric tensor $C_{\mu\nu 11}$ and the dilaton ϕ always arise in the Neveu-Schwarz sector and couple universally to $\mathrm{e}^{-2\phi}$. On the other hand the vector V_μ and the 3-form $C_{\mu\nu\rho}$ describe Ramond-Ramond (R-R) states and the specific form of their vertex operators forbids any tree-level coupling to the dilaton [18,55]. In particular the Kaluza-Klein gauge field V_μ corresponds in the string context to the R-R gauge field of type-II string theory. The infinite tower of massive Kaluza-Klein states carry a charge quantized in units of e_{KK}, defined in (3.32). In the context of 10-dimensional supergravity, states with R-R charge are solitonic. In string theory, R-R charges are carried by the D-brane states.

For later purposes let us note that the above discussion can be generalized to arbitrary spacetime dimensions. The Einstein frame in any dimension is defined by a gravitational action that is just the Einstein-Hilbert action, whereas in the string frame the Ricci scalar is multiplied by a dilaton term $\exp(-2\phi_D)$, as in (3.41) and (3.42), respectively. The Weyl rescaling which connects the two frames is given by,

$$[e_\mu^{\,a}]^{\text{string}}=\mathrm{e}^{2\phi_D/(D-2)}\,[e_\mu^{\,a}]^{\text{Einstein}}\,. \qquad (3.43)$$

Let us now return to 11-dimensional supergravity with the 11-th coordinate compactified to a circle so that $0\leq x^{10}\leq L$. As we stressed already, L itself has no intrinsic meaning and it is better to consider the geodesic radius of the 11-th dimension, which reads

$$R_{10}=\frac{L}{2\pi}\,\mathrm{e}^{2\langle\phi\rangle/3}\,. \qquad (3.44)$$

This result applies to the frame specified by the 11-dimensional theory[17].

[17]This is the frame specified by the metric given in (3.30), which leads to the Lagrangian (3.38).

In the string frame, the above result reads

$$(R_{10})^{\text{string}} = \frac{L}{2\pi}\,\mathrm{e}^{\langle\phi\rangle}\,. \tag{3.45}$$

It shows that a small 11-th dimension corresponds to small values of $\exp\langle\phi\rangle$ which in turn corresponds to a weakly coupled string theory. Observe that L is fixed in terms of κ_{10} and κ_{11} ((3.33)).

From the 11-dimensional expressions,

$$E_a^{\ M}\partial_M = e_a^{\ \mu}(\partial_\mu - V_\mu\,\partial_{10})\,, \qquad E_{10}^{\ M}\partial_M = \mathrm{e}^{-2\phi/3}\,\partial_{10}\,, \tag{3.46}$$

where a and μ refer to the 10-dimensional Lorentz and world indices, we infer that, in the frame specified by the 11-dimensional theory, the Kaluza-Klein masses are multiples of

$$M^{\text{KK}} = \frac{1}{R_{10}}\cdot \tag{3.47}$$

Hence Kaluza-Klein states have a mass and a Kaluza-Klein charge ((3.32)) related by

$$M^{\text{KK}} = |e_{\text{KK}}|\,\mathrm{e}^{-2\langle\phi\rangle/3}. \tag{3.48}$$

In the string frame, this result becomes

$$(M^{\text{KK}})^{\text{string}} = |e_{\text{KK}}|\,\mathrm{e}^{-\langle\phi\rangle}. \tag{3.49}$$

Massive Kaluza-Klein states are always BPS states, meaning that they are contained in supermultiplets that are "shorter" than the generic massive supermultiplets because of nontrivial central charges. The central charge here is just the 10-th component of the momentum, which is proportional to the Kaluza-Klein charge.

The surprising insight that emerged, is that the Kaluza-Klein features of 11-dimensional supergravity have a precise counterpart in string theory [54–56]. There one has nonperturbative (in the string coupling constant) states which carry R-R charges. On the supergravity side these states often appear as solitons.

3.6 Kaluza–Klein states and BPS-extended supergravity

In most of this section we restrict ourselves to pure supergravity. However, when compactifying dimensions one also encounters massive Kaluza-Klein states, which couple to the supergravity theory as massive matter supermultiplets. The presence of these BPS states introduces a number of qualitative changes to the theory which we discuss in this section. The most conspicuous change is that the continuous nonlinearly realized symmetry group G

is broken to an arithmetic subgroup, known as the U-duality group. This U-duality group has been conjectured to be the exact symmetry group of (toroidally compactified) M-theory [54]. The BPS states (which are contained in M-theory) should therefore be assigned to representations of the U-duality group. Here one naively assumes that the U-duality group acts on the central charges of the BPS states and it is simply defined as the arithmetic subgroup of G that leaves the central-charge lattice invariant. However, central charges are in principle assigned to representations of the group H_{R} and not of the group G (although the central charges will eventually, upon gauge fixing, transform nonlinearly under G, field-dependent H_{R} transformations). In most cases, these H_{R} representations can be elevated to representations of G, by multiplying with the representatives of the coset space G/H_{R} (representatives of coset spaces will be discussed in Sects. 4 and 5). In this way, the pointlike (field-dependent) central charges can be assigned to representations of G for spacetime dimensions $D \geq 4$. Similar observations exist for stringlike and membranelike central charges except that in these cases the dimension must be restricted even further [41].

Another aspect of the coupling of the BPS states to supergravity is that the central charges should be related to symmetries, in view of the fact that they appear in the anticommutator of two supercharges and supersymmetry is realized locally. Therefore nonzero central charges must couple to appropriate gauge fields in the supergravity theory. These gauge fields transform (with minor exceptions) linearly with constant matrices under the group G. Inspection of the tables that we have presented earlier, shows that the gauge fields usually appear in the G-representation required for gauging the corresponding central charge. Provided that the central charges can be assigned to the appropriate representations of the U-duality group and that the appropriate gauge fields are available, one may thus envisage a (possibly local field) theory of BPS states coupled to supergravity that is U-duality invariant. This theory would exhibit many of the features of M-theory and describe many of the relevant degrees of freedom.

The Kaluza-Klein states that we encounter in toroidal compactifications of supergravity are a subset of the 1/2-BPS states in M-theory. They carry pointlike central charges and they couple to the Kaluza-Klein photon fields, , the vector gauge fields that emerge from the higher-dimensional metric upon the toroidal compactification. However, they do constitute representations of the U-duality group, because the central charges that they carry are too restricted. This is the reason why retaining the Kaluza-Klein states in the dimensional compactification will lead to a breaking of the U-duality group. From the eleven-dimensional perspective it is easy to see why the central charges associated with the Kaluza-Klein states are too restricted, because under U-duality the central charges related to the momentum operator in the compactified dimensions combine with the two- and five-brane charges ((2.12)) in order to define representations of the U-duality group. However, conventional dimensional compactification does not involve any brane charges. Nevertheless, in certain cases one may still be able to extend the Kaluza-Klein states with other BPS states, so that a U-duality invariant theory is obtained. Such extended theories are called BPS-extended supergravity [40, 41].

The fact that some of the central charges are associated with extra spacetime dimensions (the charges carried by the Kaluza-Klein states) implies that the newly introduced states (associated with wrapped branes) may also have an interpretation in terms of extra dimensions. In this way, the number of spacetime dimensions could exceed eleven, although the theory would presumably not be able to decompactify uniformly to a flat spacetime of more than eleven dimensions. The aim of this section is to elucidate some of these ideas in the relatively simple context of $N = 2$ supergravity in $D = 9$ spacetime dimensions.

We start by considering the BPS multiplets that are relevant in 9 spacetime dimensions from the perspective of supergravity, string theory and (super)membranes. In 9 dimensions the R-symmetry group and the duality group are equal to $H_{\mathrm{R}} = SO(2)$ and $G = SO(1,1) \times SL(2;\mathbf{R})$, respectively. It is well known that the massive supermultiplets of IIA and IIB string theory coincide, whereas the massless states comprise inequivalent supermultiplets for the simple reason that they transform according to different representations of the $SO(8)$ helicity group. When compactifying the theory on a circle, IIA and IIB states that are massless in 9 spacetime dimensions, transform according to identical representations of the $SO(7)$ helicity group and constitute equivalent supermultiplets. The corresponding interacting field theory is the unique $N = 2$ supergravity theory in 9 spacetime dimensions. However, the BPS supermultiplets which carry momentum along the circle, remain inequivalent as they remain assigned to the inequivalent representations of the group $SO(8)$ which is now associated with the restframe (spin) rotations of the massive states. Henceforth the momentum states of the IIA and the IIB theories will be denoted as KKA and KKB states, respectively. The fact that they constitute inequivalent supermultiplets, has implications for the winding states in order that T-duality remains valid [19].

In 9 spacetime dimensions with $N = 2$ supersymmetry the Lorentz-invariant central charges are encoded in a two-by-two real symmetric matrix Z^{ij}, which can be decomposed as

$$Z^{ij} = b\,\delta^{ij} + a\,(\cos\theta\,\sigma_3 + \sin\theta\,\sigma_1)^{ij}\,. \tag{3.50}$$

Here $\sigma_{1,3}$ are the real symmetric Pauli matrices. We note that the central charge associated with the parameter a transforms as a doublet under the

$SO(2)$ R-symmetry group that rotates the two supercharge spinors, while the central charge proportional to the parameter b is $SO(2)$ invariant. Subsequently one shows that BPS states that carry these charges must satisfy the mass formula,

$$M_{\rm BPS} = |a| + |b|\,. \tag{3.51}$$

Here one can distinguish three types of BPS supermultiplets. One type has central charges $b = 0$ and $a \neq 0$. These are 1/2-BPS muliplets, because they are annihilated by half of the supercharges. The KKA supermultiplets that comprise Kaluza-Klein states of IIA supergravity are of this type. Another type of 1/2-BPS mulitplets has central charges $a = 0$ and $b \neq 0$. The KKB supermultiplets that comprise the Kaluza-Klein states of IIB supergravity are of this type. Finally there are 1/4-BPS multiplets (annihilated by one fourth of the supercharges) characterized by the fact that neither a nor b vanishes.

For type-II string theory one obtains these central charges in terms of the left- and right-moving momenta, $p_{\rm L}$, $p_{\rm R}$, that characterize winding and momentum along S^1. However, the result takes a different form for the IIA and the IIB theory as the following formula shows,

$$Z^{ij} = \begin{cases} \frac{1}{2}(p_{\rm L}+p_{\rm R})\delta^{ij} + \frac{1}{2}(p_{\rm L}-p_{\rm R})\sigma_3^{ij}\,, & \text{(for IIB)} \\ \frac{1}{2}(p_{\rm L}-p_{\rm R})\delta^{ij} + \frac{1}{2}(p_{\rm L}+p_{\rm R})\sigma_3^{ij}\,. & \text{(for IIA)} \end{cases} \tag{3.52}$$

The corresponding BPS mass formula is thus equal to

$$M_{\rm BPS} = \tfrac{1}{2}|p_{\rm L}+p_{\rm R}| + \tfrac{1}{2}|p_{\rm L}-p_{\rm R}|\,. \tag{3.53}$$

For $p_{\rm L} = p_{\rm R}$ we confirm the original identification of the momentum states, namely that IIA momentum states constitute KKA supermultiplets, while IIB momentum states constitute KKB supermultiplets. For the winding states, where $p_{\rm L} = -p_{\rm R}$, one obtains the opposite result: IIA winding states constitute KKB supermultiplets, while IIB winding states constitute KKA supermultiplets. The 1/4-BPS multiplets arise for string states that have either right- or left-moving oscillator states, so that either $M_{\rm BPS} = |p_{\rm L}|$ or $|p_{\rm R}|$ with $p_{\rm L}^2 \neq p_{\rm R}^2$. All of this is entirely consistent with T-duality [16, 17], according to which there exists a IIA and a IIB perspective, with decompactification radii are that inversely proportional and with an interchange of winding and momentum states. Observe that the 1/4-BPS states will never become massless, so that they don't play a role in what follows.

It is also possible to view the central charges from the perspective of the 11-dimensional (super)membrane [57]. Assuming that the two-brane charge takes values in the compact coordinates labelled by 9 and 10, which can be generated by wrapping the membrane over the corresponding T^2, one readily finds the expression,

$$Z^{ij} = Z_{9\,10}\,\delta^{ij} - (P_9\,\sigma_3^{ij} - P_{10}\,\sigma_1^{ij})\,. \tag{3.54}$$

When compactifying on a torus with modular parameter τ and area A, the BPS mass formula takes the form

$$\begin{aligned} M_{\rm BPS} &= \sqrt{P_9^2 + P_{10}^2} + |Z_{9\,10}| \\ &= \frac{1}{\sqrt{A\,\tau_2}}|q_1 + \tau\,q_2| + T_{\rm m}A\,|p|\,. \end{aligned} \tag{3.55}$$

Here $q_{1,2}$ denote the momentum numbers on the torus and p is the number of times the membrane is wrapped over the torus; $T_{\rm m}$ denotes the supermembrane tension. Clearly the KKA states correspond to the momentum modes on T^2 while the KKB states are associated with the wrapped membranes on the torus. Therefore there is a rather natural way to describe the IIA and IIB momentum and winding states starting from a (super)membrane in eleven spacetime dimensions. This point was first emphasized in [58].

This suggests to consider $N = 2$ supergravity in 9 spacetime dimensions and couple it to the simplest BPS supermultiplets corresponding to KKA and KKB states. As shown in Tables 8 and 10 there are three central charges and 9-dimensional supergravity possesses precisely three gauge fields that couple to these charges. From the perspective of 11-dimensional supergravity compactified on T^2, the Kaluza-Klein states transform as KKA multiplets. Their charges transform obviously with respect to an $SO(2)$ associated with rotations of the coordinates labelled by 9 and 10. Hence we have a “double” tower of these charges with corresponding KKA supermultiplets. On the other hand, from the perspective of IIB compactified on S^1, the Kaluza-Klein states constitute KKB multiplets and their charge is $SO(2)$ invariant. Here we have a “single” tower of KKB supermultiplets. However, from the perspective of 9-dimensional supergravity one is led to couple both towers of KKA and KKB supermultiplets simultaneously. In that case one obtains some dichotomic theory [19], which we refer to as BPS-extended supergravity. In the case at hand this new theory describes the ten-dimensional IIA and IIB theories in certain decompactification limits, as well as eleven-dimensional supergravity. But the theory is in some sense truly 12-dimensional with three compact coordinates, although there is no 12-dimensional Lorentz invariance, not even in a uniform decompactification limit, as the fields never depend on all the 12 coordinates! Whether this kind of BPS-extended supergravity offers a viable scheme in a more general context than the one we discuss here, is not known. In the case at hand we know a lot about these couplings from our knowledge of the T^2 compactification of $D = 11$ supergravity and the S^1 compactification of IIB supergravity.

The fields of 9-dimensional $N = 2$ supergravity are listed in Table 13, where we also indicate their relation with the fields of 11-dimensional and 10-dimensional IIA/B supergravity upon dimensional reduction. It is not necessary to work out all the nonlinear field redefinitions here, as the corresponding fields can be uniquely identified by their scaling weights under $SO(1,1)$, a symmetry of the massless theory that emerges upon dimensional reduction and is associated with scalings of the internal vielbeine. The scalar field σ is related to G_{99}, the IIB metric component in the compactified dimension, by $G_{99} = \exp(\sigma)$; likewise it is related to the determinant of the 11-dimensional metric in the compactified dimensions, which is equal to $\exp(-\frac{4}{3}\sigma)$. The precise relationship follows from comparing the $SO(1,1)$ weights through the dimensional reduction of IIB and eleven-dimensional supergravity. In 9 dimensions supergravity has two more scalars, which are described by a nonlinear sigma model based on $SL(2,\mathbf{R})/SO(2)$. The coset is described by the complex doublet of fields ϕ^α, which satisfy a constraint $\phi^\alpha \phi_\alpha = 1$ and are subject to a local $SO(2)$ invariance, so that they describe precisely two scalar degrees of freedom ($\alpha = 1, 2$). We expect that the local $SO(2)$ invariance can be incorporated in the full BPS-extended supergravity theory and can be exploited in the construction of the couplings of the various BPS supermultiplets to supergravity.

We already mentioned the three Abelian vector gauge fields which couple to the central charges. There are two vector fields A_μ^α, which are the Kaluza-Klein photons from the T^2 reduction of eleven-dimensional supergravity and which couple therefore to the KKA states. From the IIA perspective these correspond to the Kaluza-Klein states on S^1 and the D0 states. From the IIB side they originate from the tensor fields, which confirms that they couple to the IIB (elementary and D1) winding states. These two fields transform under $SL(2)$, which can be understood from the perspective of the modular transformation on T^2 as well as from the S-duality transformations that rotate the elementary strings with the D1 strings. The third gauge field, denoted by B_μ, is a singlet under $SL(2)$ and is the Kaluza-Klein photon on the IIB side, so that it couples to the KKB states. On the IIA side it originates from the IIA tensor field, which is consistent with the fact that the IIA winding states constitute KKB supermultiplets.

From the perspective of the supermembrane, the KKA states are the momentum states on T^2, while the KKB states correspond to the membranes wrapped around the torus. While it is gratifying to see how all these correspondences work out, we stress that, from the perspective of 9-dimensional $N = 2$ supergravity, the results follow entirely from supersymmetry.

The resulting BPS-extended theory incorporates 11-dimensional supergravity and the two type-II supergravities in special decompactification limits. But, as we stressed above, we are dealing with a 12-dimensional theory

Table 13. The bosonic fields of the eleven dimensional, type-IIA, nine-dimensional $N = 2$ and type-IIB supergravity theories. The 11-dimensional and 10-dimensional indices, respectively, are split as $\hat{M} = (\mu, 9, 10)$ and $M = (\mu, 9)$, where $\mu = 0, 1, \dots 8$. The last column lists the $SO(1,1)$ scaling weights of the fields.

$D = 11$	IIA	$D = 9$	IIB	$SO(1,1)$
$\hat{G}_{\mu\nu}$	$G_{\mu\nu}$	$g_{\mu\nu}$	$G_{\mu\nu}$	0
$\hat{A}_{\mu\, 9\, 10}$	$C_{\mu\, 9}$	B_μ	$G_{\mu\, 9}$	-4
$\hat{G}_{\mu\, 9},\ \hat{G}_{\mu\, 10}$	$G_{\mu\, 9}\ ,\ C_\mu$	A_μ^{α}	$A_{\mu\, 9}^{\alpha}$	3
$\hat{A}_{\mu\nu\, 9},\ \hat{A}_{\mu\nu\, 10}$	$C_{\mu\nu\, 9}, C_{\mu\nu}$	$A_{\mu\nu}^{\alpha}$	$A_{\mu\nu}^{\alpha}$	-1
$\hat{A}_{\mu\nu\rho}$	$C_{\mu\nu\rho}$	$A_{\mu\nu\rho}$	$A_{\mu\nu\rho\sigma}$	2
$\hat{G}_{9\, 10},\ \hat{G}_{9\, 9},\ \hat{G}_{10\, 10}$	$\phi,\ G_{9\, 9},\ C_9$	ϕ^{α}	ϕ^{α}	0
		$\exp(\sigma)$	$G_{9\, 9}$	7

here, although no field can depend nontrivially on all of these coordinates. The theory has obviously two mass scales associated with the KKA and KKB states. We return to them in a moment. Both S- and T-duality are manifest, although the latter has become trivial as the theory is not based on a specific IIA or IIB perspective. One simply has the freedom to view the theory from a IIA or a IIB perspective and interpret it accordingly.

We should discuss the fate of the group $G = SO(1,1) \times SL(2, \mathbf{R})$ of pure supergravity after coupling the theory to the BPS multiplets. The central charges of the BPS states form a discrete lattice, which is affected by this group. Hence, after coupling to the BPS states, we only have a discrete subgroup that leaves the charge lattice invariant. This is the group $SL(2, \mathbf{Z})$.

The KKA and KKB states and their interactions with the massless theory can be understood from the perspective of compactified 11-dimensional and IIB supergravity. In this way we are able to deduce the following BPS mass formula,

$$M_{\mathrm{BPS}}(q_1, q_2, p) = m_{\mathrm{KKA}}\, \mathrm{e}^{3\sigma/7}\, |q_\alpha \phi^\alpha| + m_{\mathrm{KKB}}\, \mathrm{e}^{-4\sigma/7}\, |p| \,, \tag{3.56}$$

where q_α and p refer to the integer-valued KKA and KKB charges, respectively, and m_{KKA} and m_{KKB} are two independent mass scales. This formula can be compared to the membrane BPS formula (3.55) in the 11-dimensional frame. One then finds that

$$m_{\mathrm{KKA}}^2\, m_{\mathrm{KKB}} \propto T_{\mathrm{m}} \,, \tag{3.57}$$

Table 14. Field content for maximal super-Maxwell theories in various dimensions. All supermultiplets contain a gauge field A_μ, scalars ϕ and spinors χ and comprises $8+8$ degrees of freedom. In $D=3$ dimensions the vector field is dual to a scalar. The 6^* representation of $SU(4)$ is a selfdual rank-2 tensor.

D	$H_{\rm R}$	A_μ	ϕ	χ
10	1	1	0	1
9	1	1	1	1
8	$U(1)$	1	$1+\bar{1}$	$1+\bar{1}$
7	$USp(2)$	1	3	2
6	$USp(2)\times USp(2)$	1	$(2,2)$	$(2,1)+(1,2)$
5	$USp(4)$	1	5	4
4	$U(4)$	1	6^*	$4+\bar{4}$
3	$SO(8)$		8	8

with a numerical proportionality constant. However, the most important conclusion to draw from (3.56) is that there is no limit in which the masses of both KKA and KKB states will tend to zero. In other words, there is no uniform decompactification limit. Therefore, in spite of the fact that we have more than 11 dimensions, there exists no theory with $Q=32$ supercharges in flat Minkowski spacetime of dimensions $D>11$.

3.7 Nonmaximal supersymmetry: $Q = 16$

For completeness we also summarize a number of results on nonmaximal supersymmetric theories with $Q=16$ supercharges, which are now restricted to dimensions $D\leq 10$. Table 14 shows the field representations for the vector multiplet in dimension $3\leq D\leq 10$. This multiplet comprises $8+8$ physical degrees of freedom. We also consider the $Q=16$ supergravity theories. The Lagrangian can be obtained by truncation of (3.38). However, unlike in the case of maximal supergravity, we now have the option of introducing additional matter fields. For $Q=16$ the matter will be in the form of vector supermultiplets, possibly associated with some non-Abelian gauge group. Table 15 summarizes $Q=16$ supergravity for dimensions $3\leq D\leq 10$. In $D=10$ dimensions the bosonic terms of the supergravity Lagrangian take the form [59],

$$\begin{aligned}\mathcal{L}_{10} &= \frac{1}{\kappa_{10}^2}\Big[-\tfrac{1}{2}e\,\mathrm{e}^{2\phi/3}R(e,\omega)\\ &\qquad -\tfrac{3}{4}e\,\mathrm{e}^{-2\phi/3}(H_{\mu\nu\rho})^2-\tfrac{1}{4}e(\partial_\mu A_\nu-\partial_\nu A_\mu)^2\Big],\end{aligned}\qquad(3.58)$$

Table 15. Bosonic fields of nonmaximal supergravity with $Q = 16$. In 6 dimensions type-A and type-B correspond to (1, 1) and (2, 0) supergravity. In the third column, # denotes the number of bosonic degrees of freedom. Note that, with the exception of the 6B and the 4-dimensional theory, all these theories contain precisely one scalar field. The tensor field in the 6B theory is selfdual. In $D = 4$ dimensions, the $SU(4)$ transformations cannot be implemented on the vector potentials, but act on the (Abelian) field strengths by duality transformations. In $D = 3$ dimensions supergravity is a topological theory and can be coupled to scalars and spinors. The scalars parametrize the coset space $SO(8,k)/(SO(8)\times SO(k))$, where k is an arbitrary integer.

D	$H_{\rm R}$	#	graviton	$p=-1$	$p=0$	$p=1$
10	1	64	1	1		1
9	1	56	1	1	1	1
8	$U(1)$	48	1	1	$1+\bar{1}$	1
7	$USp(2)$	40	1	1	3	1
6A	$USp(2)\times USp(2)$	32	1	1	(2, 2)	(1, 1)
6B	$USp(4)$	24	1			5^*
5	$USp(4)$	24	1	1	5+1	
4	$U(4)$	16	1	$1+\bar{1}$	[6]	
3	$SO(8)$	$8k$	1	$8k$		

where, for convenience, we have included a single vector gauge field A_μ, representing an Abelian vector supermultiplet. A feature that deserves to be mentioned, is that the field strength $H_{\mu\nu\rho}$ associated with the 2-rank gauge field contains a Chern-Simons term $A_{[\mu}\partial_\nu A_{\rho]}$. Chern-Simons terms play an important role in the anomaly cancellations of this theory. Note also that the kinetic term for the Kaluza-Klein vector field in (3.38), depends on ϕ, unlike the kinetic term for the matter vector field in the Lagrangian above. This reflects itself in the extension of the symmetry transformations noted in (3.36, 3.37),

$$\begin{aligned} e_\mu^a &\to \mathrm{e}^{-\alpha} e_\mu^a\,, \qquad & \phi &\to \phi + 12\alpha\,,\\ C_{11\mu\nu} &\to \mathrm{e}^{6\alpha} C_{11\mu\nu}\,, \qquad & A_\mu &\to \mathrm{e}^{3\alpha} A_\mu \end{aligned} \tag{3.59}$$

where A_μ transforms differently from the Kaluza-Klein vector field V_μ.

In this case there are three different Weyl rescalings that are relevant, namely

$$
\begin{aligned}
e_\mu^a &= \mathrm{e}^{-\phi/12}\,[e_\mu^a]^{\mathrm{Einstein}}\,, \\
e_\mu^a &= \mathrm{e}^{-\phi/3}\,[e_\mu^a]^{\mathrm{string}}\,, \\
e_\mu^a &= \mathrm{e}^{\phi/6}\,[e_\mu^a]^{\mathrm{string}'}\,.
\end{aligned}
\tag{3.60}
$$

It is straightforward to obtain the corresponding Lagrangians. In the Einstein frame, the graviton is again invariant under the isometries of the scalar field. The bosonic terms read

$$
\begin{aligned}
\mathcal{L}_{10}^{\mathrm{Einstein}} = \frac{1}{\kappa_{10}^2}\Big[&-\tfrac{1}{2}e\,R(e,\omega) - \tfrac{1}{4}e(\partial_\mu\phi)^2 \\
&-\tfrac{3}{4}e\,\mathrm{e}^{-\phi}(H_{\mu\nu\rho})^2 - \tfrac{1}{4}e\,\mathrm{e}^{-\phi/2}(\partial_\mu A_\nu - \partial_\nu A_\mu)^2\Big]\,.
\end{aligned}
\tag{3.61}
$$

The second Weyl rescaling leads to the following Lagrangian,

$$
\begin{aligned}
\mathcal{L}_{10}^{\mathrm{string}} = \frac{1}{\kappa_{10}^2}\mathrm{e}^{-2\phi}\Big[&-\tfrac{1}{2}e\,R(e,\omega) + 2e(\partial_\mu\phi)^2 \\
&-\tfrac{3}{4}e\,(H_{\mu\nu\rho})^2 - \tfrac{1}{4}e(\partial_\mu A_\nu - \partial_\nu A_\mu)^2\Big]\,,
\end{aligned}
\tag{3.62}
$$

which shows a uniform coupling to the dilaton. This is the low-energy effective Lagrangian relevant for the heterotic string. Eventually the matter gauge field has to be part of an non-Abelian gauge theory based on the groups $SO(32)$ or $E_8\times E_8$ in order to be anomaly-free.

Finally, the third Weyl rescaling yields

$$
\begin{aligned}
\mathcal{L}_{10}^{\mathrm{string}'} = \frac{1}{\kappa_{10}^2}\Big[&e\,\mathrm{e}^{2\phi}\Big[-\tfrac{1}{2}R(e,\omega) + 2(\partial_\mu\phi)^2\Big] \\
&-\tfrac{3}{4}e\,(H_{\mu\nu\rho})^2 - \tfrac{1}{4}e\,\mathrm{e}^{\phi}(\partial_\mu A_\nu - \partial_\nu A_\mu)^2\Big]\,.
\end{aligned}
\tag{3.63}
$$

Now the dilaton seems to appear with the wrong sign. As it turns out, this is the low-energy effective action of the type-I string, where the type-I dilaton must be associated with $-\phi$. This is related to the fact that the $SO(32)$ heterotic string theory is S-dual to type-I string theory [60].

4 Homogeneous spaces and nonlinear sigma models

This chapter offers an introduction to coset spaces and nonlinear sigma models based on such target spaces with their possible gaugings. The aim of this introduction is to facilitate the discussion in the next chapter, where we explain the gauging of maximal supergravity, concentrating on the maximal supergravities in $D = 4, 5$ spacetime dimensions. As discussed earlier, these theories have a nonlinearly realized symmetry group, equal to $E_{7(7)}$ and $E_{6(6)}$, respectively. It is possible to elevate the Abelian gauge group associated with the vector gauge fields to a non-Abelian group, which is a subgroup of these exceptional groups. The construction of these gaugings makes an essential use of the concepts and techniques discussed here.

We start by introducing the concept of a coset space G/H, where H is a subgroup of a group G. Most of this material is standard and can be found in textbooks, such as [61, 62]. Then we discuss the corresponding nonlinear sigma models, based on homogeneous target spaces, and present their description in a form that emphasizes a local gauge invariance associated with the group H. Finally we introduce the so-called [illegible] of this class of nonlinear sigma models. In this introduction we try to be as general as possible but in the examples we restrict ourselves to pseudo-orthogonal groups: $SO(n)$ or the noncompact versions $SO(p, q)$. The latter enable us to include some material on de Sitter and anti-de Sitter spacetimes, which we make use of in later chapters.

4.1 Nonlinearly realized symmetries

As an example consider the n-dimensional sphere S^n of unit radius which we may embed in an $(n+1)$-dimensional real vector space $\mathbf{R}^{n+1}$. The sphere is obviously invariant under $SO(n+1)$, the group of $(n+1)$-dimensional rotations. Such invariances are called isometries and $G = SO(n+1)$ is therefore known as the [illegible]. A [illegible] space is a space where every two points can be connected by an isometry transformation. Clearly the sphere is such a homogeneous space as every two points on S^n can be related by an $SO(n+1)$ isometry. However, the rotation connecting these two points is not unique as every point on S^n is invariant under an $SO(n)$ subgroup. This group is called the [illegible] (or stability subgroup), denoted by H. Obviously, for a homogeneous manifold, the isotropy groups for two arbitrary points are isomorphic (but not identical as one has to rotate between these points). It is convenient to choose a certain point on the sphere (let us call it the north pole) with coordinates in $\mathbf{R}^{n+1}$ given by $(0, \ldots, 0, 1)$. From the north pole, we can reach each point by a suitable rotation. However, the north pole itself is invariant under the $SO(n)$ isotropy group, consisting of the following orthogonal matrices embedded into $SO(n+1)$,

$$h = \begin{pmatrix} * & \cdots & * & 0 \\ \vdots & * & \vdots & \vdots \\ * & \cdots & * & 0 \\ 0 & \cdots & 0 & 1 \end{pmatrix} , \qquad h \in H \,. \tag{4.1}$$

Therefore, if a rotation $g_1 \in G$, with $G = SO(n+1)$ maps the north pole onto a certain point on the sphere, then the transformation $g_2 = g_1 \cdot h$ will do the same. Therefore points on the sphere can be associated with the class of group elements $g \in G$ that are equivalent up to multiplication by elements $h \in H$ from the . Such equivalence classes are called cosets, and therefore the space S^n is a coset space G/H with $G = SO(n+1)$ and $H = SO(n)$. The sphere is obviously just one particular example of a homogeneous space. Every such space can be described in terms of appropriate G/H cosets based on an isometry group G and a isotropy subgroup H.

A parametrization of the cosets of $SO(n+1)/SO(n)$, which assigns a single $SO(n+1)$ element to every coset, is thus equivalent to giving a parametrization of the sphere. It is not difficult to find such a parametrization. One first observes that every element of $SO(n+1)$ can be decomposed as the product of

$$g(\alpha) = \exp \begin{pmatrix} 0 & \alpha^i \\ -\alpha_j & 0 \end{pmatrix} , \tag{4.2}$$

with some element of H. Here $i, j = 1, \ldots, n$. In view of various applications we extend our notation to noncompact versions of the orthogonal group, so that we can also deal with noncompact spaces. The noncompact groups leave an indefinite metric invariant. For $SO(p,q)$ we have a diagonal metric with p eigenvalues $+1$ and q eigenvalues -1 (or as the overall sign is not relevant). Using the same decomposition as in (4.1), we choose this metric of the form diag$\,(\eta, 1)$, where η is again a diagonal metric with p (or q) eigenvalues equal to -1 and $q-1$ (or $p-1$) eigenvalues equal to $+1$. Elements of $SO(p,q)$ thus satisfy

$$g^{-1} = \begin{pmatrix} \eta & 0 \\ 0 & 1 \end{pmatrix} g^{\mathrm{T}} \begin{pmatrix} \eta & 0 \\ 0 & 1 \end{pmatrix} , \tag{4.3}$$

so that the metric η is obviously H-invariant. The coset representative (4.2) satisfies the condition (4.3), provided that

$$\alpha_i = \eta_{ij}\, \alpha^j \,. \tag{4.4}$$

When η equals the unit matrix, we are dealing with a compact space. When η has negative eigenvalues the matrix (4.2) is no longer orthogonal and the space will be noncompact[18]. Let us mention some examples. For $\eta = -\mathbf{1}$ we have the hyperbolic space[19] $SO(n,1)/SO(n)$, for $\eta = (-,+,\cdots,+)$ we have the de Sitter space $SO(n,1)/SO(n-1,1)$, and for $\eta = (-,\cdots,-,+)$, we have the anti-de Sitter space $SO(n-1,2)/SO(n-1,1)$. In this way we can thus treat a variety of spaces at the same time. However, note that η, which eventually will play the role of the tangent-space metric, is "mostly plus" for de Sitter, and "mostly minus" for the hyperbolic and the anti-de Sitter space. This aspect will be important later on when comparing the curvature for these spaces.

The elements $g(\alpha) \in G$ define a [illegible] of the G/H coset space, and therefore a parametrization of the corresponding space. Of course, the coset representative is not unique. We have decomposed the generators of G into generators h of H and generators k belonging to its complement; the latter have been used to generate the coset representative. In our example, the latter are associated with the last row and column of (4.2), so that they satisfy (schematically),

$$\begin{aligned} [\mathsf{h},\mathsf{h}] &= \mathsf{h}\,, \\ [\mathsf{h},\mathsf{k}] &= \mathsf{k}\,, \\ [\mathsf{k},\mathsf{k}] &= \mathsf{h}\,. \end{aligned} \tag{4.5}$$

The first commutation relation states that the h form a subalgebra. The second one implies that the generators k form a representation of H, which ensures that in an infinitesimal neighbourhood of a point invariant under H, the coordinates α^i rotate under H according to that representation. In the more general case the third commutator may also yield the generators k. When they do not, the homogeneous space is [illegible]. Obviously the above relations involve a choice of basis; the generators k are defined up to additive terms belonging to elements of the algebra associated with H.

[18] Noncompact groups are not fully covered by exponentiation, such as in the coset representatives (4.2), because there are disconnected components. We also note that, in the noncompact case, there are other decompositions than (4.1), which offer distinct advantages. We will not discuss these issues here.

[19] We are a little cavalier here with our terminology. Also the de Sitter and anti-de Sitter spaces are hyperbolic, as we shall see later (*cf.* (4.10)), but they are pseudo-Riemannian. We reserve the term hyperbolic for the Riemannian hyperbolic space. Unlike the de Sitter and anti-de Sitter spaces the former spaces are double-sheeded. In general, coset spaces where H is the maximal compact subgroup of a noncompact group G, have a positive or negative definite metric and are thus Riemannian spaces.

Let us now proceed and evaluate (4.2),

$$g(\alpha) = \begin{pmatrix} \delta^i_j + \alpha^i\alpha_j \dfrac{\cos\alpha - 1}{\alpha^2} & \dfrac{\sin\alpha}{\alpha}\alpha^i \\ \\ -\dfrac{\sin\alpha}{\alpha}\alpha_j & \cos\alpha \end{pmatrix} , \tag{4.6}$$

where $\alpha^2 = \eta_{ij}\,\alpha^i\alpha^j$. Obviously the space is compact when η is positive, because in that case the parameter space can be restricted to $0 \leq \alpha < \pi$. This corresponds to the sphere S^n. In all other cases the parameter space is obviously noncompact and the sine and cosine may change to the hyperbolic sine and cosine in those parts of the space where α^2 is negative. Observe that the appearance of η in the above formulae is the result of the fact that the generators k are normalized according to $\mathrm{tr}(\mathsf{k}_i\,\mathsf{k}_j) = -2\,\eta_{ij}$.

One may choose a different parametrization of the cosets by making a different decomposition than in (4.2). Different parametrizations are generally related through (coordinate-dependent) H transformations acting from the right. We may also chose different coordinates, such as, for instance[20],

$$y^i = \alpha^i \frac{\sin\alpha}{\alpha} , \tag{4.7}$$

so that the coset representative reads

$$g(y) = \begin{pmatrix} \delta^i_j + y^i y_j \,\dfrac{\pm\sqrt{1-y^2} - 1}{y^2} & y^i \\ \\ -y_j & \pm\sqrt{1-y^2} \end{pmatrix} \tag{4.8}$$

where, depending on the sign choice, we parametrize different parts of the space (for the sphere S^n, the upper or the lower hemisphere). For the sphere the range of the coordinates is restricted by $y^2 = \Sigma_i (y^i)^2 \leq 1$. Note that the $n\times n$ submatrix in (4.8) equals the square root of the matrix $\delta^i_j - y^i y_j$.

One may use the coset representative to sweep out the coset space from one point (the “north pole”) in the $(n{+}1)$-dimensional embedding space. Acting with (4.8) on the point $(0,\ldots,0,1)$ yields the following coordinates in the embedding space,

$$Y^A = (y^i, \pm\sqrt{1-y^2}) . \tag{4.9}$$

[20]The coordinates $(y^i, \pm\sqrt{1-y^2})$ are sometimes called *homogeneous* coordinates, because the G-transformations act linearly on these coordinates. Inhomogeneous coordinates are the ratios $y^i/\sqrt{1-y^2}$.

Using (4.3) one then shows that the coset space is embedded in $n+1$ dimensions according to

$$\eta_{ij}\, Y^i Y^j + (Y^{n+1})^2 = 1\,. \tag{4.10}$$

Since the $g(y)$ are contained in G one may examine the effect of G transformations acting on $g(y)$, which will induce corresponding transformations in the coset space. To see this, we multiply $g(y)$ by a constant element $o_G \in G$ from the left. After this multiplication the result is in general no longer compatible with the coset representative $g(y)$, but by applying a suitable y-dependent H transformation, $o_H(y)$, from the right, we can again bring $g(y)$ in the desired form. In other words, one has

$$g(y) \longrightarrow o_G\, g(y) = g(y')\, o_H(y)\,. \tag{4.11}$$

Hence the effect is a change of coordinates $y \to y'$ in a way that satisfies the group multiplication laws. The infinitesimal transformation $y^i \to y'^{\,i} = y^i + \xi^i(y)$ defines the so-called Killing vectors $\xi^i(y)$. Writing $o_G \approx \mathbf{1} + \hat{\mathsf{g}}$ and $o_H \approx \mathbf{1} + \hat{\mathsf{h}}(y)$, we find the relation,

$$\xi^i(y)\, \partial_i g(y) = \hat{\mathsf{g}}\, g(y) - g(y)\, \hat{\mathsf{h}}(y)\,. \tag{4.12}$$

We return to this result in the next section.

Applying this construction to the case at hand, one finds that there are two types of isometries. One corresponding to the group H, which changes the coordinates y^i by constant rotations. The other corresponds to n coordinate dependent shifts,

$$\delta y^i = \epsilon^i \sqrt{1-y^2}\,, \tag{4.13}$$

where the ϵ^i are n constant parameters. As the reader can easily verify, both types of transformations take the form of a constant G transformation on the embedding coordinates Y^A which leaves the embedding condition (4.10) invariant.

Coset representatives can be defined in different representations of the group G. The most interesting one is the spinor representation. Assume that we have a representation of the Clifford algebra $\mathcal{C}(p,q)$. The representation transforms (not necessarily irreducibly) under $SO(p,q)$ generated by the matrices $\frac{1}{2}\Gamma_{ij}$, but in fact it transforms also as a spinor under $SO(p,q+1)$, as one can verify by including extra generators equal to the matrices $\frac{1}{2}\Gamma_i$. Consequently we can define a representative of $SO(p,q+1)/SO(p,q)$ in the spinor representation,

$$g(\alpha) = \exp[\tfrac{1}{2}\Gamma_i\, \alpha^i] = \cos(\alpha/2)\, \mathbf{1} + i\, \frac{\sin(\alpha/2)}{\alpha}\, \alpha^i \Gamma_i\,, \tag{4.14}$$

with α defined as before, $\alpha^2 = \sqrt{\alpha_i \alpha^i}$, and $\{\Gamma_i, \Gamma_j\} = -2\eta_{ij}\,\mathbf{1}$. This construction can applied as well to cosets of other (pseudo-)orthogonal groups. In terms of the coordinates y^i the representative reads

$$g(y) = \tfrac{1}{2}\Big[\sqrt{1+y} + \sqrt{1-y}\,\Big]\,\mathbf{1} + \frac{y^i \Gamma_i}{\sqrt{1+y} + \sqrt{1-y}}\cdot \tag{4.15}$$

One can act with this representative on a constant spinor, specified at the north pole, we define $\psi(y) = g^{-1}(y)\,\psi(0)$. The resulting y-dependent spinor $\psi(y)$ is a so-called Killing spinor of the coset space. We shall exhibit this below. Obviously, similar results can be obtained in other representations of G.

4.2 Geometrical quantities

Geometrical quantities of the homogeneous space are defined from the left-invariant one-forms $g^{-1}\,\mathrm{d}g$, where $g(y) \in G$, so that the one-forms take their value in the Lie algebra associated with G. It is convenient to use the language of differential forms,but by no means essential. The exterior derivative $\mathrm{d}g(y)$ describes the change of g induced by an infinitesimal variation of the coset-space coordinates y^i. The one-forms $g^{-1}\,\mathrm{d}g$ are called left-invariant, because they are invariant under left multiplication of g with constant elements of G. The significance of this fact will be clear in a sequel. Because the g's themselves are elements of G, the one-form $g^{-1}\,\mathrm{d}g$ takes its values in the Lie algebra associated with G. Therefore the one-forms can be decomposed into the generators h and k, introduced earlier, ,

$$g^{-1}\,\mathrm{d}g = \omega + e\,, \tag{4.16}$$

where ω is decomposable into the generators h and e into the generators k. Hence e defines a square matrix, with indices i that label the coordinates and indices a that label the generators k. These one-forms e are thus related to the vielbeine of the coset space, which define a tangent frame at each point of the space. The one-forms ω define the spin connection[21], associated with tangent-space rotations that belong to the group H. Equation (4.16) is of central importance for the geometry of the coset spaces. As a first consequence we note that the spinor $\psi(y)$, defined with the help of the representative (4.15) at the end of the previous subsection, satisfies the equation,

$$\Big(\mathrm{d} + \omega + e\Big)\psi(y) = 0\,. \tag{4.17}$$

[21] Observe that in supergravity we have defined the spin connection field with opposite sign.

Upon writing this out in terms of the gamma matrices, one recovers precisely the so-called Killing spinor equation ((3.15)).

Let us now proceed and investigate the properties of the one-forms ω and e. In general it is not necessary to specify the coset representative, as different representatives are related by y-dependent H transformation acting from the right on g, ,

$$g(y) \longrightarrow g(y)\, h(y)\,, \qquad h(y) \in H\,. \tag{4.18}$$

This leads to a different parametrization of the coset space. It is straightforward to see how ω and e transform under (4.18)

$$(\omega + e) \longrightarrow h^{-1}(\omega + e)h + h^{-1}\,\mathrm{d}h\,. \tag{4.19}$$

This equation can again be decomposed (using the first two relations (4.5) in terms of the generators h and k, which yields

$$\begin{aligned} \omega_i &\longrightarrow h^{-1}\,\omega_i h + h^{-1}\,\partial_i h\,, \\ e_i &\longrightarrow h^{-1}\,e_i h\,. \end{aligned} \tag{4.20}$$

Obviously, ω acts as a gauge connection for the local H transformations. Furthermore it follows from (4.27) that ω_i and e_i transform as covariant vectors under coordinate transformations,

$$\begin{aligned} y^i &\longrightarrow y^i + \xi^i\,, \\ \omega_i &\longrightarrow \omega_i - \partial_i\xi^j\omega_j - \xi^j\partial_j\omega_i\,, \\ e_i &\longrightarrow e_i - \partial_i\xi^j e_j - \xi^j\partial_j e_i\,. \end{aligned} \tag{4.21}$$

We can also define Lie-algebra valued curvatures associated with ω_i and e_i,

$$\begin{aligned} R_{ij}(H) &= \partial_i\omega_j - \partial_j\omega_i + [\omega_i, \omega_j]\,, \\ R_{ij}(G/H) &= \partial_i e_j - \partial_j e_i + [\omega_i, e_j] - [\omega_j, e_i]\,. \end{aligned} \tag{4.22}$$

Introducing H-covariant derivatives, we note the relations

$$\begin{aligned} [D_i, D_j] &= -R_{ij}(H)\,, \\ R_{ij}(G/H) &= D_i e_j - D_j e_i\,. \end{aligned} \tag{4.23}$$

The values of these curvatures follow from the Cartan-Maurer equations. To derive these equations we take the exterior derivative of the defining relation (4.16),

$$\mathrm{d}(g^{-1}\,\mathrm{d}g) = -(g^{-1}\,\mathrm{d}g) \wedge (g^{-1}\,\mathrm{d}g)\,, \tag{4.24}$$

or, in terms of ω and e,

$$\mathrm{d}(\omega + e) = -(\omega + e) \wedge (\omega + e)\,. \tag{4.25}$$

Decomposing this equation in terms of the Lie algebra generators, using the relations (4.5), we find

$$R_{ij}(H) = -[e_i, e_j]\,, \qquad R_{ij}(G/H) = 0\,. \tag{4.26}$$

Note that the vanishing of $R_{ij}(G/H)$ is a consequence of the fact that we assumed that the coset space was [illegible] (see the text below (4.5)).

As we already alluded to earlier, the fields e_i can be decomposed into the generators k and thus define a set of vielbeine $e_i^{\,a}$ that specify a tangent frame at each point in the coset space. In the context of differential geometry the indices i are called world indices, because they refer to the coordinates of a manifold, whereas the indices $a, b, \ldots$ that label the generators k are called tangent-space indices (or local Lorentz indices in the context of general relativity). Because the generators k form a representation of H, this group rotates the tangent frames. Usually the group H can be embedded into $SO(n)$ (or a noncompact version thereof) and leaves some target-space metric invariant (we will see the importance of this fact shortly). The quantity ω_i thus acts as the connection associated with rotations of the tangent frames, and therefore we call it the spin connection.

These aspects are easily recognized in the examples we are discussing, because the group H was precisely the (pseudo)orthogonal group. Hence, using the same matrix decomposition as before, we find explicit expressions for the vielbein and the spin connection,

$$g^{-1}\,\mathrm{d}g = \begin{pmatrix} \omega_i(y)\,\mathrm{d}y^i & e_i(y)\,\mathrm{d}y^i \\ & \\ -e_i(y)\,\mathrm{d}y^i & 0 \end{pmatrix}. \tag{4.27}$$

From (4.8) one readily obtains

$$\begin{aligned} \omega_i^{\,ab} &= \Big(y^a\,\delta_i^b - y^b\,\delta_i^a\Big)\frac{1 \mp \sqrt{1-y^2}}{y^2}\,, \\ e_i^{\,a} &= \delta_i^a + \frac{y_i\,y^a}{y^2}\Big(\pm\frac{1}{\sqrt{1-y^2}} - 1\Big)\,, \end{aligned} \tag{4.28}$$

where, as before, indices are raised and lowered with η. Note that $\omega_i^{\,ab}$ is antisymmetric in a, b, which follows from the (pseudo)orthogonality of H. The inverse vielbein reads

$$e_a^{\,i} = \delta_a^i + \frac{y_a\,y^i}{y^2}\Big(\pm\sqrt{1-y^2} - 1\Big). \tag{4.29}$$

Furthermore, the curvatures introduced before, are readily identified with the curvature of the spin connection and with the torsion tensor. From the

Cartan-Maurer equations, explained above, we thus find in components,

$$R_{ij}^{ab}(\omega) = 2\, e_{[i}^{a}\, e_{j]}^{b}\,, \qquad D_i\, e_j^{\,a} - D_j e_i^{\,a} = 0\,. \tag{4.30}$$

To define a metric g_{ij} one contracts an H-invariant symmetric rank-2 tensor with the vielbeine. The obvious invariant tensor is η_{ab}, so that

$$g_{ij} = \eta_{ab}\, e_i^{\,a} e_j^{\,b}\,. \tag{4.31}$$

When there are several H-invariant tensors, there is a more extended class of metrics that one may consider, but in the case at hand the metric is unique up to a proportionality factor. In the parametrization (4.28) one obtains for the metric and its inverse,

$$g_{ij} = \eta_{ij} + \frac{y_i\, y_j}{1-y^2}\,, \qquad g^{ij} = \eta^{ij} - y^i\, y^j\,. \tag{4.32}$$

Given the fact that we have already made a choice for η previously, a “mostly plus” metric requires to include a minus sign in the definition (4.31) for the hyperbolic and anti-de Sitter spaces. This sign is important when comparing to spheres or de Sitter spaces.

From the vielbein postulate, we know that the affine connection is equal to $\Gamma_{ij}{}^{k} = e_a{}^{k}\, D_i e_j{}^{a}$, from which we can define the Riemann curvature. For the examples at hand, this leads to

$$\Gamma_{ij}{}^{k} = y^k g_{ij}. \tag{4.33}$$

Because the torsion is zero, the connection coincides with Christoffel symbol. Because ω_i differs in sign as compared to the spin connection used in Section 3.1, the Riemann tensor is equal to minus the curvature $R_{ij}^{ab}(\omega)$, upon contraction with $\eta_{ac}\, e_k^{c}\, e_b^{l}$, and we find the following result for the Riemann curvature tensor,

$$R_{ijk}{}^{l} = -g_{ki}\, \delta_j^l + g_{kj}\, \delta_i^l\,, \tag{4.34}$$

where g_{ij} is the metric tensor defined by (4.31). Thus the curvature is of definite sign, but we stress that this is related to the signature choice that we made for the metric, as we have discussed above. The Riemann curvature (4.34) is proportional to the metric, which indicates that we are dealing with a maximally symmetric space. This means that the maximal number of isometries (equal to $\frac{1}{2}n(n+1)$) is realized for this space.

All coset spaces have isometries corresponding to the group G. The diffeomorphisms associated with these isometries are generated by Killing vectors $\xi^i(y)$, which we introduced earlier. Combining (4.12) with (4.16), we obtain,

$$\xi^i(y)\,(\omega_i(y) + e_i(y)) = g^{-1}(y)\hat{\mathsf{g}}\, g(y) - \hat{\mathsf{h}}(y)\,. \tag{4.35}$$

Decomposing this equation according to the Lie algebra, we find

$$\begin{aligned} \xi^i(y)\, e_i(y) &= \tilde{\mathsf{g}}(y)\,, \\ \hat{\mathsf{h}}(y) &= -\xi^i(y)\,\omega_i(y) + \tilde{\mathsf{h}}(y)\,, \end{aligned} \tag{4.36}$$

where

$$\begin{aligned} \tilde{\mathsf{h}}(y) &= \Big[g^{-1}(y)\,\hat{\mathsf{g}}\,g(y)\Big]_H\,, \\ \tilde{\mathsf{g}}(y) &= \Big[g^{-1}(y)\,\hat{\mathsf{g}}\,g(y)\Big]_{G/H}\,. \end{aligned} \tag{4.37}$$

The contribution $\hat{\mathsf{h}}(y)$ is only relevant for those quantities that live in the tangent space.

Now we return to the observation that the left-invariant forms, from which e and ω were constructed, are invariant under the group G. Therefore, it follows that e and ω are both invariant as well. Moreover, we established ((4.11)) that the G-transformation acting on the left can be decomposed into a diffeomorphism combined with a coordinate-dependent H-transformation. Therefore, the vielbein e and the spin connection ω are invariant under these combined transformations. Since the metric is H-invariant by construction, it thus follows that the metric is invariant under the diffeomorphism associated with G. Hence,

$$\delta g_{ij} = D_i\xi_j(y) + D_j\xi_i(y) = 0\,, \tag{4.38}$$

where $\xi_i = g_{ij}\xi^j$ and ξ^i is the so-called Killing vector defined by (4.36). For the vielbein and spin connection, which transform under H, we find

$$\begin{aligned} \partial_i\xi^j\, e_j + \xi^j\,\partial_j e_i + [e_i, \hat{\mathsf{h}}(y)] &= 0\,, \\ \partial_i\xi^j\,\omega_j + \xi^j\,\partial_j\omega_i + \partial_i\hat{\mathsf{h}}(y) + [\omega_i, \hat{\mathsf{h}}(y)] &= 0\,. \end{aligned} \tag{4.39}$$

In terms of $\tilde{\mathsf{h}}(y)$ these results take a more covariant form,

$$\begin{aligned} \partial_i\xi^j\, e_j + \xi^j\, D_j e_i + [e_i, \tilde{\mathsf{h}}(y)] &= 0\,, \\ D_i\tilde{\mathsf{h}}(y) &= R_{ij}(H)\,\xi^j\,. \end{aligned} \tag{4.40}$$

Combining the first equation with the first equation (4.36) yields

$$R_{ij}(G/H)\,\xi^j + [e_i, \tilde{\mathsf{g}}(y)]_{G/H} = 0\,. \tag{4.41}$$

Observe that both terms vanish separately for a symmetric space.

The diffeomorphisms generated by the Killing vector fields will give rise to the group G. This follows from (4.12). Let us label the generators of the group G by indices $\alpha, \beta, \ldots$, and introduce structure constants by

$$[\hat{\mathsf{g}}_\alpha, \hat{\mathsf{g}}_\beta] = f_{\alpha\beta}{}^\gamma\,\hat{\mathsf{g}}_\gamma\,. \tag{4.42}$$

The Killing vectors and the corresponding H-transformations then satisfy corresponding group multiplication properties,

$$\begin{aligned} \xi^j_\beta\,\partial_j\xi^i_\alpha - \xi^j_\alpha\,\partial_j\xi^i_\beta &= f_{\alpha\beta}{}^\gamma\,\xi^i_\gamma\,, \\ [\tilde{\mathsf{h}}_\alpha, \tilde{\mathsf{h}}_\beta] &= f_{\alpha\beta}{}^\gamma\,\tilde{\mathsf{h}}_\gamma + \xi^i_\alpha\,\xi^j_\beta\,R_{ij}(H)\,. \end{aligned} \tag{4.43}$$

One can consider fields on the coset space, which are functions of the coset space coordinates assigned to a representation of the group H. On such fields the isometries are generated by the operators,

$$-\xi^i_\alpha D_i + \tilde{\mathsf{h}}_\alpha\,. \tag{4.44}$$

On the basis of the results above one can show that these operators satisfy the commutation relations of the Lie algebra associated with the isometry group G. To show this we note the identity

$$\begin{aligned} (-\xi^i_\alpha D_i + \tilde{\mathsf{h}}_\alpha)(-\xi^i_\beta D_i + \tilde{\mathsf{h}}_\beta) &= \xi^i_\alpha\xi^j_\beta(D_iD_j - R_{ij}(H)) + \tilde{\mathsf{h}}_\alpha\,\tilde{\mathsf{h}}_\beta \\ &\quad + [\xi^i_\alpha(D_i\xi^j_\beta) + \tilde{\mathsf{h}}_\alpha\,\xi^j_\beta + \tilde{\mathsf{h}}_\beta\,\xi^j_\alpha]D_j\,. \end{aligned} \tag{4.45}$$

4.3 Nonlinear sigma models with homogeneous target space

It is now rather straightforward to describe a nonlinear sigma model based on a homogeneous target space by making use of the above framework. One starts from scalar fields which take their values in the homogeneous space, so that the fields $\phi^i(x)$ define a map from the spacetime to the coset space. Hence we may follow the same procedure as before and define a coset representative $\mathcal{V}(\phi^i(x)) \in G$, which now depends on n fields. Subsequently, one uses the analogue of (4.16), to define Lie-algebra valued quantities $\mathcal{Q}_\mu$ and $\mathcal{P}_\mu$,

$$\mathcal{V}^{-1}\partial_\mu\mathcal{V} = \mathcal{Q}_\mu + \mathcal{P}_\mu\,, \tag{4.46}$$

where $\mathcal{Q}_\mu$ is decomposable into the generators h and $\mathcal{P}_\mu$ into the generators k. Obviously one has the relations

$$\mathcal{Q}_\mu(\phi) = \omega_i(\phi)\,\partial_\mu\phi^i\,, \qquad \mathcal{P}_\mu(\phi) = e_i(\phi)\,\partial_\mu\phi^i\,. \tag{4.47}$$

The above expressions show that $\mathcal{Q}_\mu$ and $\mathcal{P}_\mu$ are just the pull backs of the target space connection and vielbein to the spacetime.

The local H transformations depend on the fields $\phi(x)$ and thus indirectly on the spacetime coordinates. Therefore one may elevate these transformations to transformations that depend arbitrarily on x^μ. Under such transformations we have

$$\mathcal{V}(\phi) \to \mathcal{V}(\phi)\,h(x)\,. \tag{4.48}$$

By allowing ourselves to perform such local gauge transformations, we introduced new degrees of freedom into $\mathcal{V}$ associated with the group H. Eventually we will fix this gauge freedom, but until that point $\mathcal{V}$ will just be an unrestricted spacetime dependent element of the group G. After imposing the gauge condition on $\mathcal{V}(x)$ one obtains the coset representative $\mathcal{V}(\phi(x))$. From (4.48) we derive the following local H-transformations,

$$\begin{aligned} \mathcal{Q}_\mu(x) &\longrightarrow h^{-1}(x)\,\mathcal{Q}_\mu(x)\,h(x) + h^{-1}(x)\,\partial_\mu h(x)\,, \\ \mathcal{P}_\mu(x) &\longrightarrow h^{-1}(x)\,\mathcal{P}_\mu(x)\,h(x)\,. \end{aligned} \tag{4.49}$$

Hence $\mathcal{Q}_\mu$ acts as a gauge field associated with the local H transformations. Furthermore both $\mathcal{P}_\mu$ and $\mathcal{Q}_\mu$ are invariant under rigid G-transformations. It is convenient to introduce a corresponding H-covariant derivative,

$$D_\mu\mathcal{V} = \partial_\mu\mathcal{V} - \mathcal{V}\mathcal{Q}_\mu\,, \tag{4.50}$$

so that (4.47) reads

$$\mathcal{V}^{-1}D_\mu\mathcal{V} = \mathcal{P}_\mu\,. \tag{4.51}$$

Just as before, one derives the Cartan-Maurer equations (4.26),

$$\begin{aligned} F_{\mu\nu}(\mathcal{Q}) &= \partial_\mu\mathcal{Q}_\nu - \partial_\nu\mathcal{Q}_\mu + [\mathcal{Q}_\mu,\mathcal{Q}_\nu] = -[\mathcal{P}_\mu,\mathcal{P}_\nu]\,, \\ D_{[\mu}\mathcal{P}_{\nu]} &= \partial_{[\mu}\mathcal{P}_{\nu]} + [\mathcal{Q}_{[\mu},\mathcal{P}_{\nu]}] = 0\,. \end{aligned} \tag{4.52}$$

Here we made use of the commutation relations (4.5)

There are several ways to write down the Lagrangian of the corresponding nonlinear sigma model. Obviously the Lagrangian must be invariant under both the rigid G transformations and the local H transformations. Hence we write

$$\mathcal{L} = \tfrac{1}{2}\mathrm{tr}\left[D_\mu\mathcal{V}^{-1}\,D^\mu\mathcal{V}\right]. \tag{4.53}$$

One can interpret this result in a first- and in a second-order form. In the first one regards the gauge field $\mathcal{Q}_\mu$ as an independent field, whose field equations are algebraic and are solved by (4.51). After substituting the result one obtains the second-order form, which presupposes (4.51) from the beginning. The result can be written as,

$$\mathcal{L} = -\tfrac{1}{2}\mathrm{tr}\left[\mathcal{P}_\mu\,\mathcal{P}^\mu\right]. \tag{4.54}$$

Clearly this Lagrangian is invariant under the group G. At this stage one still has the full gauge invariance with respect to local H transformations and one can impose a gauge restricting $\mathcal{V}$ to a coset representative. When this is not done, the theory is invariant under $G_{\mathrm{rigid}}\times H_{\mathrm{local}}$ with both groups acting [illegible]. However, as soon as one imposes a gauge and restricts $\mathcal{V}$ to a coset representative parametrized by certain fields ϕ^i, the residual subgroup is such that the H transformations are linked to the G transformations and depend on the fields ϕ^i. This combined subgroup still generates a representation of the group G, but it is now realized in a nonlinear fashion. In order to deal with complicated supergravity theories that involve homogeneous spaces, the strategy is to postpone this gauge choice till the end, so that one is always dealing with a manifest linearly realized symmetry group $G_{\rm rigid} \times H_{\rm local}$. As we intend to demonstrate, this strategy allows for a systematic approach, whereas the gauge-fixed approach leads to unsurmountable difficulties (at least, for the spaces of interest). It is straightforward to demonstrate that (4.54) leads to the standard form of the nonlinear sigma model,

$$\mathcal{L} = -\tfrac{1}{2} g_{ij}(\phi)\, \partial_\mu \phi^i\, \partial^\mu \phi^j\ , \tag{4.55}$$

where the target space metric is given by (4.31). In this form the local H invariance is absent, but the invariance under G is still there and realized as target space isometries generated by corresponding Killing vectors.

In this form it is easy to see how to couple matter fields to the sigma model in a way that the invariance under the isometries remains unaffected. Matter fields are assigned to a representation of the local H group, so that they couple to the sigma model fields through the connection $\mathcal{Q}_\mu$ that appears in the covariant derivatives. Usually the fields will remain invariant under the group G as long as one does not fix the gauge and choose a specific coset representative. Also here we can proceed in first- or second-order formalism. In first-order form the equation (4.47) will acquire some extra terms that depend on the matter fields. Upon choosing a gauge, the matter fields transform nonlinearly under the group G with transformations that take the form of ϕ-dependent H-transformations, determined by (4.36). However, gauge fields cannot couple in this way as their gauge invariance would be in conflict with the local invariance under the group H. Therefore, gauge fields have to transform under the rigid group G.

We emphasize that the presentation that we followed so far was rather general; the maximally symmetric spaces that we considered served only as an example. In supergravity we are often dealing with sigma models based on homogeneous, symmetric target spaces. These target spaces are usually noncompact and Riemannian, so that H is the maximally compact subgroup of G. Later in this chapter we will be dealing with the $E_{7(7)}/SU(8)$ and $E_{6(6)}/USp(8)$ coset spaces. The exceptional groups are noncompact and are divided by their maximal compact subgroups. The corresponding spaces have dimension 70 and 42, respectively.

4.4 Gauged nonlinear sigma models

Given a nonlinear sigma model with certain isometries, one can gauge some or all of these isometries in the usual way: one elevates the parameters of the isometry group (of a subgroup thereof) to arbitrary functions of the spacetime coordinates and introduces the necessary gauge fields (with their standard gauge-invariant Lagrangian containing a kinetic term) and corresponding covariant derivatives. As explained above, for sigma models based on homogeneous target spaces one can proceed in a way in which all transformations remain linearly realized. To adopt this approach is extremely important for the construction of gauged supergravity theories, as we will discuss in the next section. We will always use the second-order formalism so that the H-connection $\mathcal{Q}_\mu$ will not be an independent field.

Since these new gauge transformations involve the isometry group they must act on the group element $\mathcal{V}$ as a subgroup of G. Hence the covariant derivative of $\mathcal{V}$ is now changed by the addition of the corresponding (dynamical) gauge fields A_μ which take their values in the corresponding Lie algebra (which is a subalgebra of the Lie algebra associated with G). Hence,

$$D_\mu \mathcal{V}(x) = \partial_\mu \mathcal{V}(x) - \mathcal{V}(x)\, \mathcal{Q}_\mu(x) - g\, A_\mu(x)\, \mathcal{V}(x)\,, \tag{4.56}$$

where we have introduced a coupling constant g to keep track of the new terms introduced by the gauging. With this change, the expressions for $\mathcal{Q}_\mu$ and $\mathcal{P}_\mu$ will change. They remain expressed by (4.51), but the derivative is now covariantized and modified by the terms depending on the new gauge fields A_μ. The consistency of this procedure is obvious as (4.51) is fully covariant. Of course, the original rigid invariance under G transformations from the left is now broken by the embedding of the new gauge group into G.

The modifications caused by the new minimal couplings are minor and the effects can be concisely summarized by the Cartan-Maurer equations,

$$\begin{aligned} \mathcal{F}_{\mu\nu}(\mathcal{Q}) &= \partial_\mu \mathcal{Q}_\nu - \partial_\nu \mathcal{Q}_\mu + [\mathcal{Q}_\mu, \mathcal{Q}_\nu]\,, \\ &= [\mathcal{P}_\mu, \mathcal{P}_\nu] - g\Big[\mathcal{V}^{-1} F_{\mu\nu}(A) \mathcal{V}\Big]_H\,, \\ D_{[\mu}\mathcal{P}_{\nu]} &= -\tfrac{1}{2} g \Big[\mathcal{V}^{-1} F_{\mu\nu}(A) \mathcal{V}\Big]_{G/H}\,. \end{aligned} \tag{4.57}$$

Because $\mathcal{Q}_\mu$ and $\mathcal{P}_\mu$ now depend on the gauge connections A_μ, according to

$$\mathcal{Q}_\mu = \mathcal{Q}^{(0)}_\mu - g[\mathcal{V}^{-1} A_\mu \mathcal{V}]_H\,, \qquad \mathcal{P}_\mu = \mathcal{P}^{(0)}_\mu - g[\mathcal{V}^{-1} A_\mu \mathcal{V}]_{G/H}\,. \tag{4.58}$$

When imposing a gauge condition, the last result for $\mathcal{P}_\mu$ exhibits precisely the Killing vectors (4.36) (in the gauge where $\mathcal{V}$ equals the coset representative). When gauging isometries in a generic nonlinear sigma model

((4.55)), one replaces the derivatives according to $\partial_\mu \phi \to \partial_\mu \phi^i - A_\mu \xi^i(\phi)$, where for simplicity we assumed a single isometry. The modifications in the matter sector arise through the order g contributions to $\mathcal{Q}_\mu$. Note that $\mathcal{P}_\mu$ and $\mathcal{Q}_\mu$ are invariant under the new gauge group (but transform under local H-transformations, as before). In the next section we will discuss the application of this formulation to gauged supergravity.

5 Gauged maximal supergravity in 4 and 5 dimensions

The maximally extended supergravity theories introduced in Section 3 were obtained by dimensional reduction from 11-dimensional supergravity on a hypertorus. In these theories the scalar fields parametrize a G/H coset space (Table 12) and the group G is also realized as a symmetry of the full theory. Generically the fields transform as follows. The graviton is invariant, the (Abelian) gauge fields transform linearly under G and the fermions transform linearly under the group H. However, in some dimensions the G-invariance is not realized at the level of the action, but at the level of the combined field equations and Bianchi identities. For example, in 4 dimensions the 28 Abelian vector fields do not constitute a representation of the group $E_{7(7)}$. In this case the group G is realized by electric-magnetic duality and acts on the field strengths, rather than on the vector fields. We return to electric-magnetic duality in Section 5.3.

It is an obvious question whether these theories allow an extension in which the Abelian gauge fields are promoted to non-Abelian ones. This turns out to be possible and the corresponding theories are known as gauged supergravities. They contain an extra parameter g, which is the non-Abelian coupling constant. Supersymmetry requires the presence of extra terms of order g and g^2 in the Lagrangian. Apart from the gauge field interactions there are fermionic masslike terms of order g and a scalar potential of order g^2. The latter may give rise to groundstates with nonzero cosmological constant. To explain the construction of gauged supergravity theories, we concentrate on the maximal gauged supergravities in $D = 4$ and 5 spacetime dimensions. An obvious gauging in $D = 4$ dimensions is based on the group $SO(8)$, as the Lagrangian has a manifest $SO(8)$ invariance and there are precisely 28 vector fields [63]. This gauging has an obvious Kaluza-Klein origin, and arises when compactifying seven coordinates of $D = 11$ supergravity on the sphere S^7, which has an $SO(8)$ isometry group. The group emerges as the gauge group of the compactified theory formulated in 4 dimensions. In this theory the $E_{7(7)}$ invariance group is broken to a local $SO(8)$ group so that the resulting theory is invariant under $SU(8)_{\text{local}} \times SO(8)_{\text{local}}$. In this compactification the four-index field strength acquires a nonzero values when all its indices are in the four-dimensional spacetime.

Table 16. Representation assignments for the various supergravity fields with respect to the groups G and H. In $D = 4$ dimensions these groups are $E_{7(7)}$ and $SU(8)$, respectively. In $D = 5$ dimensions they are $E_{6(6)}$ and $USp(8)$. Note that the tensors $\mathcal{F}_{\mu\nu}$, $\mathcal{G}_{\mu\nu}$ and/or $\mathcal{H}_{\mu\nu}$ denote the field strengths of the vector fields and/or (for $D = 5$) possible tensor fields.

	e_μ^a	ψ_μ^i	$\mathcal{F}_{\mu\nu};\mathcal{G}_{\mu\nu};\mathcal{H}_{\mu\nu}$	χ^{ijk}	${u_{ij}}^{IJ};v^{ijIJ}$
$SU(8)$	**1**	**8**	**1**	**56**	$\mathbf{28}+\overline{\mathbf{28}}$
$E_{7(7)}$	**1**	**1**	**56**	**1**	**56**
$USp(8)$	**1**	**8**	**1**	**48**	$\mathbf{27}+\overline{\mathbf{27}}$
$E_{6(6)}$	**1**	**1**	**27**	**1**	$\mathbf{27}+\overline{\mathbf{27}}$

However, it turns out that many other subgroups of $E_{7(7)}$ can be gauged.

In $D = 5$ dimensions the possible gaugings are not immediately clear, as there is no obvious 27-dimensional gauge group. Again the Kaluza-Klein scenario can serve as a guide. While $D = 11$ supergravity has no obvious compactification to five dimensions, type-IIB supergravity has a compactification on the sphere S^5. In this solution the five-index (self-dual) field strength acquires a nonzero value whenever the five indices take all values in either S^5 or in the five-dimensional spacetime. Type-IIB supergravity has a manifest $SL(2)$ invariance and the isometry group of S^5 is $SO(6)$, so that the symmetry group of the Lagrangian equals the $SL(2) \times SO(6)$ subgroup of $E_{6(6)}$, where $SO(6)$ is realized as a local gauge group. This implies that 15 of the 27 gauge fields become associated with the non-Abelian group $SO(6)$, which leaves 12 Abelian gauge fields which are charged with respect to the same group. This poses an obvious problem, as the Abelian gauge transformations of these 12 fields will be in conflict with their transformations under the $SO(6)$ gauge group. The solutions is to convert these 12 gauge fields to antisymmetric tensor fields. The Lagrangian can thus be written in a form that is invariant under $USp(8)_{\text{local}} \times SO(6)_{\text{local}} \times SL(2)$ [29]. Also in 5 dimensions other gauge groups are possible. We will briefly comment on this issue at the end of the section.

Before continuing with supergravity we first discuss some basic features of the two coset spaces $E_{7(7)}/SU(8)$ and $E_{6(6)}/USp(8)$. Both these exceptional Lie groups can be introduced in terms of 56-dimensional matrices[22].

[22]Strictly speaking the isotropy groups are $SU(8)/\mathbf{Z}_2$ and $USp(8)/\mathbf{Z}_2$.

5.1 On $E_{7(7)}/SU(8)$ and $E_{6(6)}/USp(8)$ cosets

We discuss the $E_{7(7)}$ and $E_{6(6)}$ on a par for reasons that will become obvious. To define the groups we consider the fundamental representation, acting on a pseudoreal vector (z_{IJ}, z^{KL}) with $z^{IJ} = (z_{IJ})^*$, where the indices are antisymmetrized index pairs $[IJ]$ and $[KL]$ and $I, J, K, L = 1, \ldots, 8$. Hence the (z_{IJ}, z^{KL}) span a 56-dimensional vector space. Consider now infinitesimal transformations of the form,

$$\begin{aligned} \delta z_{IJ} &= \Lambda_{IJ}{}^{KL}\, z_{KL} + \Sigma_{IJKL}\, z^{KL} \,, \\ \delta z^{IJ} &= \Lambda^{IJ}{}_{KL}\, z^{KL} + \Sigma^{IJKL}\, z_{KL} \end{aligned} \tag{5.1}$$

where $\Lambda_{IJ}{}^{KL}$ and Σ_{IJKL} are subject to the conditions

$$(\Lambda_{IJ}{}^{KL})^* = \Lambda^{IJ}{}_{KL} = -\Lambda_{KL}{}^{IJ} \,, \qquad (\Sigma_{IJKL})^* = \Sigma^{KLIJ} \,. \tag{5.2}$$

The corresponding group elements constitute the group $Sp(56; \mathbf{R})$ in a pseudoreal basis. This group is the group of electric-magnetic dualities of maximal supergravity in $D = 4$ dimensions. The matrices $\Lambda_{IJ}{}^{KL}$ are associated with its maximal compact subgroup, which is equal to $U(28)$. The defining properties of elements E of $Sp(56; \mathbf{R})$ are

$$E^* = \omega\, E\, \omega \,, \qquad E^{-1} = \Omega\, E^\dagger\, \Omega \,, \tag{5.3}$$

where ω and Ω are given by

$$\omega = \begin{pmatrix} 0 & \mathbf{1} \\ \mathbf{1} & 0 \end{pmatrix} \,, \qquad \Omega = \begin{pmatrix} \mathbf{1} & 0 \\ 0 & -\mathbf{1} \end{pmatrix} \,. \tag{5.4}$$

The above properties ensure that the sequilinear form,

$$(z_1, z_2) = z_1^{IJ}\, z_{2IJ} - z_{1IJ}\, z_2^{IJ} \,, \tag{5.5}$$

is invariant. In passing we note that the real subgroup (in this pseudoreal representation) is equal to the group $GL(28)$.

Let us now consider the $E_{7(7)}$ subgroup, for which the Σ^{IJKL} is fully antisymmetric and the generators are further restricted according to

$$\begin{aligned} &\Lambda_{IJ}{}^{KL} = \delta^{[K}_{[I}\, \Lambda_{J]}{}^{L]} \,, \qquad \Lambda_I{}^J = -\Lambda^J{}_I \,, \\ &\Lambda_I{}^I = 0 \,, \qquad \Sigma_{IJKL} = \tfrac{1}{24} \varepsilon_{IJKLMNPQ}\, \Sigma^{MNPQ} \,. \end{aligned} \tag{5.6}$$

Obviously the matrices $\Lambda_I{}^J$ generate the group $SU(8)$, which has dimension 63; since Σ_{IJKL} comprise 70 real parameters, the dimension of $E_{7(7)}$ equals $63 + 70 = 133$. Because $SU(8)$ is the maximal compact subgroup, the number of the noncompact generators minus the number of compact ones is equal to $70 - 63 = 7$. It is straightforward to show that these matrices close under commutation and generate the group $E_{7(7)}$. To show this one needs a variety of identities for selfdual tensors [64]; one of them is that the contraction $\Sigma_{IKLM}\,\Sigma^{JKLM}$ is traceless.

However, $E_{7(7)}$ has another maximal 63-dimensional subgroup, which is not compact. This is the group $SL(8)$. It is possible to choose conventions in which the $E_{7(7)}$ matrices have a different block decomposition than (5.1) and where the diagonal blocks correspond to the group $SL(8)$, rather than to $SU(8)$. We note that the subgroup generated by (5.6) with $\Lambda_I{}^J$ and Σ^{IJKL} real, defines the group $SL(8;\mathbf{R})$.

The group $E_{7(7)}$ has a quartic invariant,

$$\begin{aligned} J_4(z) &= z_{IJ}z^{JK}z_{KL}z^{LI} - \tfrac{1}{4}(z_{IJ}z^{IJ})^2 \\ &+\tfrac{1}{96}\Big[\varepsilon_{IJKLMNPQ}\, z^{IJ}z^{KL}z^{MN}z^{PQ} + \varepsilon^{IJKLMNPQ}\, z_{IJ}z_{KL}z_{MN}z_{PQ}\Big]\,, \end{aligned} \tag{5.7}$$

which, however, plays no role in the following. For further information the reader is encouraged to read the appendices of [53].

Another subgroup is the group $E_{6(6)}$, for which the restrictions are rather similar. Here one introduces a skew-symmetric tensor Ω_{IJ}, satisfying

$$\Omega_{IJ} = -\Omega_{JI}\,, \quad (\Omega_{IJ})^* = \Omega^{IJ}\,, \quad \Omega_{IK}\Omega^{KJ} = -\delta_I^J\,. \tag{5.8}$$

Now we restrict ourselves to the subgroup of $U(8)$ that leaves Ω_{IJ} invariant. This is the group $USp(8)$. The other restrictions on the generators concern Σ_{IJKL}. Altogether we have the conditions,

$$\begin{aligned} &\Lambda_{IJ}{}^{KL} = \delta_{[I}^{[K}\,\Lambda_{J]}{}^{L]}\,, \qquad \Lambda_I{}^J = -\Lambda^J{}_I\,, \\ &\Lambda_{[I}{}^K\Omega_{J]K} = 0\,, \qquad \Omega_{IJ}\Sigma^{IJKL} = 0\,, \\ &\Sigma_{IJKL} = \Omega_{IM}\,\Omega_{JN}\,\Omega_{KP}\,\Omega_{LQ}\,\Sigma^{MNPQ}\,. \end{aligned} \tag{5.9}$$

The maximal compact subgroup $USp(8)$ thus has dimension $64 - 28 = 36$, while there are $70 - 28 = 42$ generators associated with Σ_{IJKL}. Altogether we thus have $36 + 42 = 78$ generators, while the difference between the numbers of noncompact and compact generators equals $42 - 36 = 6$. These numbers confirm that we are indeed dealing with $E_{6(6)}$ and its maximal compact subgroup $USp(8)$. Because of the constraints (5.9) the 56-dimensional representation defined by (5.1) is reducible and decomposes into two singlets and a $\mathbf{27}$ and a $\overline{\mathbf{27}}$ representation. To see this we observe that the following restrictions are preserved by the group,

$$\Omega_{IJ}z^{IJ} = 0\,, \qquad z_{IJ} = \pm\Omega_{IK}\Omega_{JL}z^{KL}\,. \tag{5.10}$$

The first one suppresses the singlet representation and the second one projects out the $\mathbf{27}$ or the $\overline{\mathbf{27}}$ representation.

The group $E_{6(6)}$ has a cubic invariant, defined by

$$J_3(z) = z^{IJ} z^{KL} z^{MN}\, \Omega_{JK} \Omega_{LM} \Omega_{NI}\,, \tag{5.11}$$

which plays a role in the $E_{6(6)}$ invariant Chern-Simons term in the supergravity Lagrangian.

There is another maximal subgroup of $E_{6(6)}$, which is noncompact, that will be relevant in the following. This is the group $SL(6) \times SL(2)$, which has dimension $35 + 3 = 38$, and which plays a role in many of the known gaugings, where the gauge group is embedded into the group $SL(6)$, so that $SL(2)$ remains as a rigid invariance group of the Lagrangian.

5.2 On ungauged maximal supergravity Lagrangians

An important feature of pure extended supergravity theories is that the spinless fields take their values in a homogeneous target space (Table 12, where we have listed these spaces). Because the spinless fields always appear in nonpolynomial form, it is vital to exploit the coset structure explained in the previous section in the construction of the supersymmetric action and transformation rules, as well as in the gauging. We will not be complete here but sketch a number of features of the maximal supergravity theories in $D = 4, 5$ where the coset structure plays an important role. We will be rather cavalier about numerical factors, spinor conventions, etc. In this way we will, hopefully, be able to bring out the main features of the G/H structure, without getting entangled in issues that depend on the spacetime dimension. For those and other details we refer to the original literature [29, 63].

One starts by introducing a so-called 56-bein $\mathcal{V}(x)$, which is a 56×56 matrix that belongs to the group $E_{7(7)}$ or $E_{6(6)}$, depending on whether we are in $D = 4$, or 5 dimensions. A coset representative is obtained by exponentiation of the generators defined in (5.1). Schematically,

$$\mathcal{V}(x) = \exp \begin{pmatrix} 0 & \overline{\Sigma}(x) \\ \Sigma(x) & 0 \end{pmatrix}, \tag{5.12}$$

where the rank-4 antisymmetric tensor Σ satisfies the algebraic restrictions appropriate for the exceptional group. As explained in the previous section, the 56-bein is reducible for $E_{6(6)}$, but we will use the reducible version in order to discuss the two theories on a par. Our notation will be based on a description in terms of right cosets, just as in the previous sections, which may differ from the notations used in the original references where one sometimes uses left cosets. Hence, we assume that the 56-bein transforms under the exceptional group from the left and under the local $SU(8)$ (or

$USp(8)$) from the right. The 56-bein can be decomposed in block form according to

$$\mathcal{V}(x) = \begin{pmatrix} u^{ij}{}_{IJ}(x) & -v_{klIJ}(x) \\ \\ -v^{ijKL}(x) & u_{kl}{}^{KL}(x) \end{pmatrix} , \qquad (5.13)$$

with the usual conventions $u^{ij}{}_{IJ} = (u_{ij}{}^{IJ})^*$ and $v_{ijIJ} = (v^{ijIJ})^*$. Observe that the indices of the matrix are antisymmetrized index pairs $[IJ]$ and $[ij]$. In the above the row indices are $([IJ],[KL])$, and the column indices are $([ij],[kl])$. The latter are the indices associated with the local $SU(8)$ or $USp(8)$. The notation of the submatrices is chosen such as to make contact with [63], where left cosets were chosen, upon interchanging $\mathcal{V}$ and $\mathcal{V}^{-1}$. Observe also that (5.12) is a coset representative, we have fixed the gauge with respect to local $SU(8)$ or $USp(8)$, whereas in (5.13) gauge fixing is not assumed. According to (5.3) the inverse $\mathcal{V}^{-1}$ can be expressed in terms of the complex conjugates of the submatrices of $\mathcal{V}$,

$$\mathcal{V}^{-1}(x) = \begin{pmatrix} u_{ij}{}^{IJ}(x) & v_{ijKL}(x) \\ \\ v^{klIJ}(x) & u^{kl}{}_{KL}(x) \end{pmatrix} . \qquad (5.14)$$

Consequently we derive the identities, for $E_{7(7)}$,

$$\begin{aligned} u^{ij}{}_{IJ}\, u_{kl}{}^{IJ} - v^{ijIJ}\, v_{klIJ} &= \delta^{ij}_{kl} , \\ u^{ij}{}_{IJ}\, v^{klIJ} - v^{ijIJ}\, u^{kl}{}_{IJ} &= 0 , \end{aligned} \qquad (5.15)$$

or, conversely,

$$\begin{aligned} u^{ij}{}_{IJ}\, u_{ij}{}^{KL} - v_{ijIJ}\, v^{ijKL} &= \delta^{IJ}_{KL} , \\ u^{ij}{}_{IJ} v_{ijKL} - v_{ijIJ} u^{ij}{}_{KL} &= 0 . \end{aligned} \qquad (5.16)$$

The corresponding equations for $E_{6(6)}$ are identical, except that the antisymmetrized Kronecker symbols on the right-hand sides are replaced according to

$$\delta^{ij}_{kl} \to \delta^{ij}_{kl} - \tfrac{1}{8}\Omega_{kl}\,\Omega^{ij} , \qquad \delta^{IJ}_{KL} \to \delta^{IJ}_{KL} - \tfrac{1}{8}\Omega_{KL}\,\Omega^{IJ} . \qquad (5.17)$$

Furthermore the matrices u and v vanish when contracted with the invariant tensor Ω and they are pseudoreal, ,

$$u_{ij}{}^{IJ}\Omega_{IJ} = 0 , \qquad u_{ij}{}^{KL}\Omega_{IK}\Omega_{JL} = \Omega_{ik}\Omega_{jl}u^{kl}{}_{IJ} , \qquad (5.18)$$

with similar identities for the v^{ijIJ}. In this case the (pseudoreal) matrices $u^{ij}{}_{IJ} \pm \Omega_{IK}\,\Omega_{JL}\,v^{ijKL}$ and their complex conjugates define (irreducible) elements of $E_{6(6)}$ corresponding to the **27** and $\overline{\mathbf{27}}$ representations. We note the identity

$$\Big(u^{ij}{}_{IJ} + \Omega_{IK}\,\Omega_{JL}\,v^{ijKL}\Big)\Big(u_{kl}{}^{IJ} - \Omega^{IM}\,\Omega^{JN}\,v_{klMN}\Big) = \delta^{ij}_{kl} - \tfrac{1}{8}\Omega_{kl}\,\Omega^{ij}\,. \tag{5.19}$$

In this case we can thus decompose the 56-bein in terms of a 27-bein and a $\overline{27}$-bein.

Subsequently we evaluate the quantities $\mathcal{Q}_\mu$ and $\mathcal{P}_\mu$,

$$\mathcal{V}^{-1}\partial_\mu\mathcal{V} = \begin{pmatrix} \mathcal{Q}_{\mu\,ij}{}^{mn} & \mathcal{P}_{\mu\,ijpq} \\ & \\ \mathcal{P}^{klmn}_\mu & \mathcal{Q}_\mu{}^{kl}{}_{pq} \end{pmatrix}, \tag{5.20}$$

which leads to the expressions,

$$\begin{aligned} \mathcal{Q}_{\mu\,ij}{}^{kl} &= u_{ij}{}^{IJ}\,\partial_\mu u^{kl}{}_{IJ} - v_{ijIJ}\,\partial_\mu v^{klIJ}\,, \\ \mathcal{P}^{ijkl}_\mu &= v^{ijIJ}\,\partial_\mu u^{kl}{}_{IJ} - u^{ij}{}_{IJ}\,\partial_\mu v^{klIJ}\,. \end{aligned} \tag{5.21}$$

The important observation is that $\mathcal{Q}_{\mu ij}{}^{kl}$ and $\mathcal{P}^{ijkl}_\mu$ are subject to the same constraints as the generators of the exceptional group listed in the previous section. Hence, $\mathcal{P}^{ijkl}_\mu$ is fully antisymmetric and subject to a reality constraint. Therefore it transforms according to the 70-dimensional representation of $SU(8)$, with the reality condition,

$$\mathcal{P}^{ijkl}_\mu = \tfrac{1}{24}\,\varepsilon^{ijklmnpq}\,\mathcal{P}_{\mu\,mnpq}\,, \tag{5.22}$$

or, to the 42-dimensional representation of $USp(8)$, with the reality condition,

$$\mathcal{P}^{ijkl}_\mu = \Omega^{im}\Omega^{jn}\Omega^{kp}\Omega^{lq}\,\mathcal{P}_{\mu\,mnpq}. \tag{5.23}$$

Likewise $\mathcal{Q}_\mu$ transforms as a connection associated with $SU(8)$ or $USp(8)$, respectively. Hence $\mathcal{Q}_{\mu\,ij}{}^{kl}$ must satisfy the decomposition,

$$\mathcal{Q}_{\mu\,ij}{}^{kl} = \delta^{[k}_{[i}\,\mathcal{Q}_{\mu\,j]}{}^{l]}\,, \tag{5.24}$$

so that $\mathcal{Q}_{\mu\,i}{}^j$ equals

$$\mathcal{Q}_{\mu\,i}{}^j = \tfrac{2}{3}\Big[u_{ik}{}^{IJ}\,\partial_\mu u^{jk}{}_{IJ} - v_{ikIJ}\,\partial_\mu v^{jkIJ}\Big]\,. \tag{5.25}$$

Because of the underlying Lie algebra the connections $\mathcal{Q}_{\mu\,i}{}^j$ satisfy $\mathcal{Q}_\mu{}^i{}_j = -\mathcal{Q}_{\mu j}{}^i$ and $\mathcal{Q}_{\mu i}{}^i = 0$, as well as an extra symmetry condition in the case of $USp(8)$ ((5.9)).

Furthermore we can evaluate the Maurer-Cartan equations (4.52),

$$\begin{aligned} F_{\mu\nu}(\mathcal{Q})_i{}^j &= \partial_\mu \mathcal{Q}_{\nu\, i}{}^j - \partial_\nu \mathcal{Q}_{\mu\, i}{}^j + \mathcal{Q}_{[\mu\, i}{}^k \mathcal{Q}_{\nu]k}{}^j = -\tfrac{4}{3}\, \mathcal{P}_{[\mu}{}^{jklm}\, \mathcal{P}_{\nu]iklm}\,, \\ D_{[\mu}\mathcal{P}_{\nu]}^{ijkl} &= \partial_{[\mu}\mathcal{P}_{\nu]}^{ijkl} + 2\mathcal{Q}_{[\mu\, m}{}^{[i}\, \mathcal{P}_{\nu]}^{jkl]m} = 0\,. \end{aligned} \qquad (5.26)$$

Observe that the use of the Lie algebra decomposition for G/H is crucial in deriving these equations. Such decompositions are an important tool for dealing with the spinless fields in this nonlinear setting. Before fixing a gauge, we can avoid the nonlinearities completely and carry out the calculations in a transparent way. Fixing the gauge prematurely and converting to a specific coset representative for G/H would lead to unsurmountable difficulties.

Continuing along similar lines we turn to a number of other features that are of interest for the Lagrangian and transformation rules. The first one is the observation that variation of the 56-bein can be written, up to a local H-transformation, as

$$\delta\mathcal{V} = \mathcal{V} \begin{pmatrix} 0 & \overline{\Sigma} \\ \Sigma & 0 \end{pmatrix}, \qquad (5.27)$$

or, in terms of submatrices,

$$\delta u_{ij}{}^{IJ} = -\Sigma_{ijkl}\, v^{klIJ}\,, \qquad \delta v_{ijIJ} = -\Sigma_{ijkl}\, u^{kl}{}_{IJ}\,. \qquad (5.28)$$

where Σ^{ijkl} is the rank-four antisymmetric tensor corresponding to the generators associated with G/H (, the generators denoted by k in the previous section). Because Σ takes the form of an H-covariant tensor, the variation (5.28) is consistent with both groups G and H. Under this variation the quantities $\mathcal{Q}_\mu$ and $\mathcal{P}_\mu$ transform systematically,

$$\begin{aligned} \delta Q_{\mu\, i}{}^j &= \tfrac{2}{3}\Big(\Sigma^{jklm}\, \mathcal{P}_{\mu\, iklm} - \Sigma_{iklm}\, \mathcal{P}_\mu^{jklm}\Big)\,, \\ \delta\mathcal{P}_\mu^{ijkl} &= D_\mu\Sigma^{ijkl} = \partial_\mu\Sigma^{ijkl} + 2\mathcal{Q}_{\mu\, m}{}^{[i}\Sigma^{jkl]m}\,. \end{aligned} \qquad (5.29)$$

Observe that, this establishes that the $SU(8)$ tensors $\mathcal{Q}_\mu$ and $\mathcal{P}_\mu$ can be assigned to the adjoint representation of the group G, as is also obvious from the decomposition (5.20).

As was stressed above, any variation of $\mathcal{V}$ can be decomposed into (5.27), up to a local H-transformation. In particular this applies to supersymmetry transformations. The supersymmetry variation can be written in the form (5.27), where Σ is an H-covariant expression proportional to the supersymmetry parameter ϵ^i and the fermion fields χ^{ijk}. Hence it must be of the form $\Sigma^{ijkl} \propto \bar\epsilon^{[i}\chi^{jkl]}$, up to complex conjugation and possible contractions with H-covariant tensors, Furthermore Σ must satisfy the restrictions associated with the exceptional group, (5.6) or (5.9).

The supersymmetry variation of the spinor χ^{ijk} contains the quantity $\mathcal{P}_\mu^{ijkl}$, which incorporates the spacetime derivatives of the spinless fields, so that up to proportionality constants we must have a variation,

$$\delta\chi^{ijk} \propto \mathcal{P}_\mu^{ijkl}\,\gamma^\mu\epsilon_l\,. \tag{5.30}$$

The verification of the supersymmetry algebra on $\mathcal{V}$ is rather easy. Under two consecutive (field-dependent) variations (5.28) applied in different orders on the 56-bein, we have

$$[\delta_1,\delta_2]\,\mathcal{V} = \mathcal{V}\begin{pmatrix} 0 & 2\,\delta_{[1}\overline{\Sigma}_{2]} \\ 2\,\delta_{[1}\Sigma_{2]} & 0 \end{pmatrix} + \mathcal{V}\left[\begin{pmatrix} 0 & \overline{\Sigma}_1 \\ \Sigma_1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & \overline{\Sigma}_2 \\ \Sigma_2 & 0 \end{pmatrix}\right]. \tag{5.31}$$

The last term is just an infinitesimal H-transformations. For the first term we note that $\delta_1\Sigma_2$ leads to a term proportional to $\mathcal{P}_\mu^{ijkl}$, combined with two supersymmetry parameters, ϵ_1 and ϵ_2, of the form $(\bar\epsilon_1^{[i}\gamma^\mu\epsilon_{2\,m})\,\mathcal{P}_\mu^{jkl]m}$. Taking into account the various H-covariant combinations in the actual expressions implied by (5.6) or (5.9), respectively, this contribution can be written in the form

$$[\delta_1,\delta_2]\,\mathcal{V} \propto (\bar\epsilon_1^i\gamma^\mu\epsilon_{2\,i} - \bar\epsilon_2^i\gamma^\mu\epsilon_{1\,i})\;\mathcal{V}\begin{pmatrix} 0 & \overline{\mathcal{P}}_\mu \\ \mathcal{P}_\mu & 0 \end{pmatrix}. \tag{5.32}$$

This is precisely a spacetime diffeomorphism, up to a local H-transformation proportional to $\mathcal{Q}_\mu$, as follows from (5.20). Hence up to a number of field-dependent H-transformations, the supersymmetry commutator closes on $\mathcal{V}$ into a spacetime diffeomorphism (up to terms of higher-order in the spinors that we suppressed).

Let us now turn to the action. Apart from higher-order spinor terms, the terms in the Lagrangian pertaining to the graviton, gravitini, spinors and scalars take the following form,

$$\begin{aligned} \mathrm{e}^{-1}\mathcal{L}_1 \;=\;& -\tfrac12 R(e,\omega) - \tfrac12\bar\psi_\mu^i\gamma^{\mu\nu\rho}\Big[(\partial_\nu - \tfrac12\omega_\nu^{ab}\,\gamma_{ab})\delta_i^j + \tfrac12\mathcal{Q}_{\nu\,i}{}^j\Big]\psi_{\rho j} \\ & -\tfrac{1}{12}\bar\chi^{ijk}\gamma^\mu\Big[(\partial_\mu - \tfrac12\omega_\mu^{ab}\,\gamma_{ab})\delta_k^l + \tfrac32\mathcal{Q}_{\mu\,k}{}^l\Big]\chi_{ijl} - \tfrac{1}{96}\mathcal{P}_\mu^{ijkl}\,\mathcal{P}^\mu_{ijkl} \\ & -\tfrac{1}{12}\bar\chi_{ijk}\gamma^\nu\gamma^\mu\psi_{\nu l}\,\mathcal{P}_\mu^{ijkl}\,. \end{aligned} \tag{5.33}$$

This Lagrangian is manifestly invariant with respect to $E_{7(7)}$ or $E_{6(6)}$. Here we distinguish the Einstein-Hilbert term for gravity, the Rarita-Schwinger Lagrangian for the gravitini, the Dirac Lagrangian and the nonlinear sigma model associated with the G/H target space. The last term represents the Noether coupling term for the spin-0/spin-$\frac12$ system. For $D=4$ the fermion fields are chiral spinors and we have to add the contributions from the spinors of opposite chirality; for $D = 5$ we are dealing with so-called symplectic Majorana spinors. Here we disregard such details and concentrate on the symmetry issues.

The vector fields bring in new features, which are different for space-time dimensions $D = 4$ and 5. In $D = 5$ dimensions the vector fields B_μ^{IJ} transform as the **27** representation of $E_{6(6)}$, so that they satisfy the reality constraint $B_{\mu\,IJ} = \Omega_{IK}\Omega_{JL}\,B_\mu^{KL}$, and the Lagrangian is manifestly invariant under the corresponding transformations. It is impossible to construct an invariant action just for the vector fields and one has to make use of the scalars, which can be written in terms of the $\overline{27}$-beine, $u^{ij}{}_{IJ}+v^{ijKL}\Omega_{IK}\Omega_{JL}$, and which can be used to convert $E_{6(6)}$ to $USp(8)$ indices. Hence we define a $USp(8)$ covariant field strength for the vector fields, equal to

$$\mathsf{F}^{ij}_{\mu\nu} = (u^{ij}{}_{IJ} - v^{ijKL}\Omega_{IK}\Omega_{JL})(\partial_\mu B_\nu^{IJ} - \partial_\nu B_\mu^{IJ})\,. \tag{5.34}$$

The invariant Lagrangian of the vector fields then reads,

$$\begin{aligned}\mathcal{L}_2 \;\; = \;\; & -\tfrac{1}{16}e\,\mathsf{F}^{ij}_{\mu\nu}\,\mathsf{F}^{\mu\nu kl}\,\Omega_{ik}\Omega_{jl} \\ & -\tfrac{1}{12}\varepsilon^{\mu\nu\rho\sigma\lambda}\,B_\mu^{IJ}\,\partial_\nu B_\rho^{KL}\,\partial_\sigma B_\lambda^{MN}\,\Omega_{JK}\Omega_{LM}\Omega_{NI} \\ & +\tfrac{1}{4}e\,\mathsf{F}^{ij}_{\mu\nu}\,\mathcal{O}^{\mu\nu}_{ij}\,,\end{aligned} \tag{5.35}$$

where we distinguish the kinetic term, a Chern-Simons interaction associated with the $E_{6(6)}$ cubic invariant (5.11) and a moment coupling with the fermions. Here $\mathcal{O}^{\mu\nu}_{ij}$ denotes a covariant tensor antisymmetric in both space-time and $USp(8)$ indices and quadratic in the fermion fields, ψ_μ^i and χ^{ijk}. Observe that the dependence on the spinless fields is completely implicit. Any additional dependence would affect the invariance under $E_{6(6)}$. The result obtained by combining the Lagrangians (5.33) and (5.35) gives the full supergravity Lagrangian invariant under rigid $E_{6(6)}$ and local $USp(8)$ transformations, up to terms quartic in the fermion fields. We continue the discussion of the $D = 4$ theory in the next section, as this requires to first introduce the concept of electric-magnetic duality.

5.3 Electric–magnetic duality and $E_{7(7)}$

For $D = 4$ the Lagrangian is not invariant under $E_{7(7)}$ but under a smaller group, which acts on the vector fields (but not necessarily on the 56-bein) according to a 28-dimensional subgroup of $GL(28)$. However, the combined equations of motion and the Bianchi identities are invariant under the group $E_{7(7)}$. This situation is typical for $D = 4$ theories with Abelian vector fields, where the symmetry group of field equations and Bianchi identities can be bigger than that of the Lagrangian, and where different Lagrangians not related by local field redefinitions, can lead to an equivalent set of field equations and Bianchi identities. However, the phenomenon is not restricted to

4 dimensions and can occur for antisymmetric tensor gauge fields in any even number of spacetime dimensions (see, [52]). The 4-dimensional version has been known for a long time and is commonly referred to as electric-magnetic duality (for a recent review of this duality in supergravity, see, [51]). Its simplest form arises in Maxwell theory in four-dimensional (flat or curved) Minkowski space, where one can perform (Hodge) duality rotations, which commute with the Lorentz group and rotate the electric and magnetic fields and inductions according to $\mathbf{E} \leftrightarrow \mathbf{H}$ and $\mathbf{B} \leftrightarrow \mathbf{D}$.

This duality can be generalized to any $D = 4$ dimensional field theory with Abelian vector fields and no charged fields, so that the gauge fields enter the Lagrangian only through their (Abelian) field strengths. These field strengths (in the case at hand we have 28 of them, labelled by antisymmetric index pairs $[IJ]$, but for the moment we will remain more general and label the field strengths by $\alpha, \beta, \ldots$) are decomposed into selfdual and anti-selfdual components $F^{\pm\alpha}_{\mu\nu}$ (which, in Minkowski space, are related by complex conjugation) and so are the field strengths $G^{\pm}_{\mu\nu\,\alpha}$ that appear in the field equations, which are defined by

$$G^{\pm}_{\mu\nu\,\alpha} = \pm\frac{4i}{e}\,\frac{\partial\mathcal{L}}{\partial F^{\pm\,\alpha\mu\nu}}. \tag{5.36}$$

Together $F^{\pm\alpha}_{\mu\nu}$ and $G^{\pm}_{\mu\nu\,\alpha}$ comprise the electric and magnetic fields and inductions. The Bianchi identities and equations of motion for the Abelian gauge fields take the form

$$\partial^\mu\big(F^+ - F^-\big)^{\alpha}_{\mu\nu} = \partial^\mu\big(G^+ - G^-\big)_{\mu\nu\,\alpha} = 0\,, \tag{5.37}$$

which are obviously invariant under ., ..., rotations of the field strengths $F^{\pm}$ and $G^{\pm}$

$$\begin{pmatrix} F^{\pm\alpha}_{\mu\nu} \\ G^{\pm}_{\mu\nu\,\beta} \end{pmatrix} \longrightarrow \begin{pmatrix} U & Z \\ W & V \end{pmatrix} \begin{pmatrix} F^{\pm\alpha}_{\mu\nu} \\ G^{\pm}_{\mu\nu\,\beta} \end{pmatrix}, \tag{5.38}$$

where U^{α}_{β}, $V_{\alpha}{}^{\beta}$, $W_{\alpha\beta}$ and $Z^{\alpha\beta}$ are constant, real, $n \times n$ submatrices and n denotes the number of independent gauge potentials. In $N = 8$ supergravity we have 56 such field strengths of each duality, so that the rotation is associated with a 56×56 matrix. The relevant question is whether the rotated equations (5.37) can again follow from a Lagrangian. More precisely, does there exist a new Lagrangian depending on the new, rotated, field strengths, such that the new tensors $G_{\mu\nu}$ follow from this Lagrangian as in (5.36). This condition amounts to an integrability condition, which can only have a solution (for generic Lagrangians) provided that the matrix is an element of the group $Sp(2n;\mathbf{R})$[23]. This implies that the submatrices satisfy the constraint

$$\begin{aligned} &U^{\mathrm{T}}V - W^{\mathrm{T}}Z = VU^{\mathrm{T}} - WZ^{\mathrm{T}} = \mathbf{1}\,, \\ &U^{\mathrm{T}}W = W^{\mathrm{T}}U\,, \qquad Z^{\mathrm{T}}V = V^{\mathrm{T}}Z\,. \end{aligned} \tag{5.39}$$

We distinguish two subgroups of $Sp(2n;\mathbf{R})$. One is the invariance group of the combined field equations and Bianchi identities, which usually requires the other fields in the Lagrangian to transform as well. Of course, a generic theory does not have such an invariance group, but maximal supergravity is known to have an $E_{7(7)} \subset Sp(56;\mathbf{R})$ invariance. However, this invariance group is not necessarily realized as a symmetry of the Lagrangian. The subgroup that is a symmetry of the Lagrangian, is usually smaller and restricted by $Z=0$ and $U^{-1}=V^{\mathrm{T}}$; the subgroup associated with the matrices U equals $GL(n)$. Furthermore the Lagrangian is not uniquely defined (it can always be reparametrized an electric-magnetic duality transformation) and neither is its invariance group. More precisely, there exist different Lagrangians with different symmetry groups, whose Bianchi identities and equations of motion are the same (modulo a linear transformation) and are invariant under the same group (which contains the symmetry groups of the various Lagrangians as subgroups). These issues are extremely important when gauging a subgroup of the invariance group, as this requires the gauge group to be contained in the invariance group of the Lagrangian.

Given the fact that we can rotate the field strengths by electric-magnetic duality transformations, we assign different indices to the field strengths and the underlying gauge groups than to the 56-bein $\mathcal{V}$. Namely, we label the fields strengths by independent index pairs $[AB]$, which are related to the index pairs $[IJ]$ of the 56-bein ((5.13)) in a way that we will discuss below. Furthermore, to remain in the context of the pseudoreal basis used previously, we form the linear combinations,

$$\mathsf{F}^{+}_{1\mu\nu\,AB} = \tfrac{1}{2}(i\,G^{+}_{\mu\nu\,AB} + F^{+AB}_{\mu\nu})\,, \quad \mathsf{F}^{+AB}_{2\mu\nu} = \tfrac{1}{2}(i\,G^{+}_{\mu\nu\,AB} - F^{+AB}_{\mu\nu})\,. \tag{5.40}$$

[23] Without any further assumptions, one can show that in Minkowski spaces of dimensions $D=4k$, the rotations of the field equations and Bianchi identities associated with n rank-$(k-1)$ antisymmetric gauge fields that are described by a Lagrangian, constitute the group $Sp(2n;\mathbf{R})$. For rank-k antisymmetric gauge fields in $D=2k+2$ dimensions, this group equals $SO(n,n;\mathbf{R})$. Observe that these groups do not constitute an invariance of the theory, but merely characterize an equivalence class of Lagrangians. The fact that the symplectic redefinitions of the field strengths constitute the group $Sp(2n;\mathbf{R})$ was first derived in [65], but in the context of a duality *invariance* rather than of a *reparametrization*. In this respect our presentation is more in the spirit of a later treatment in [66] for $N=2$ vector multiplets coupled to supergravity (duality invariances for these theories were introduced in [67]).

Anti-selfdual field strengths $(\mathsf{F}^{-AB}_{1\mu\nu}, \mathsf{F}^{-}_{2\mu\nu\,AB})$ follow from complex conjugation. On this basis the field strengths rotate under $Sp(56;\mathbf{R})$ according to the matrices E specified in (5.3); the real $GL(28)$ subgroup is induced by corresponding linear transformations of the vector fields.

To exhibit how one can deal with a continuous variety of Lagrangians, which are manifestly invariant under different subgroups of $E_{7(7)}$, let us remember that the tensors $F^{AB}_{\mu\nu}$ and $G_{\mu\nu\,AB}$ are related by (5.36) and this relationship must be consistent with $E_{7(7)}$. In order to establish this consistency, the 56-bein plays a crucial role. The relation involves a constant $Sp(56;\mathbf{R})$ matrix E (so that it satisfies the conditions (5.3)),

$$\mathsf{E} = \begin{pmatrix} \mathsf{U}_{IJ}{}^{AB} & \mathsf{V}_{IJCD} \\ \mathsf{V}^{KLAB} & \mathsf{U}^{KL}{}_{CD} \end{pmatrix} . \tag{5.41}$$

On the basis of $E_{7(7)}$ and $SU(8)$ covariance, the relation among the field strengths must have the form,

$$\mathcal{V}^{-1}\,\mathsf{E}\begin{pmatrix} \mathsf{F}^{+}_{1\mu\nu\,AB} \\ \mathsf{F}^{+AB}_{2\mu\nu} \end{pmatrix} = \begin{pmatrix} \mathsf{F}^{+}_{\mu\nu\,ij} \\ \mathcal{O}^{+kl}_{\mu\nu} \end{pmatrix} , \tag{5.42}$$

where $\mathcal{O}^{+}_{\mu\nu}$ is an $SU(8)$ covariant tensor quadratic in the fermion fields and independent of the scalar fields, which appears in the moment couplings in the Lagrangian. Without going into the details we mention that the chirality and duality of $\mathcal{O}^{+}_{\mu\nu}$ is severely restricted so that the structure of (5.42) is unique ([63]). The tensor $\mathsf{F}^{+}_{\mu\nu\,ij}$ is an $SU(8)$ covariant field strength that appears in the supersymmetry transformation rules of the spinors, which is simply defined by the above condition.

Hence the matrix E allows the field strengths and the 56-bein to transform under $E_{7(7)}$ in an equivalent but nonidentical way. One could consider absorbing this matrix into the definition of the field strengths $(\mathsf{F}_1, \mathsf{F}_2)$, but such a redefinition cannot be carried out at the level of the Lagrangian, unless it belongs to a $GL(28)$ subgroup which can act on the gauge fields themselves. In the basis (5.3) the generators of $GL(28)$ have a block decomposition with $SO(28)$ generators in both diagonal blocks and identical real, symmetric, 28×28 matrices in the off-diagonal blocks. On the other hand, when $\mathsf{E} \in E_{7(7)}$, it can be absorbed into the 56-bein $\mathcal{V}$. The various Lagrangians are thus encoded in $Sp(56;\mathbf{R})$ matrices E, up to multiplication by $GL(28)$ from the right and multiplication by $E_{7(7)}$ from the left, in elements of $E_{7(7)}\backslash Sp(56;\mathbf{R})/GL(28)$.

From (5.42) one can straightforwardly determine the relevant terms in the Lagrangian. For convenience, we redefine the 56-bein by absorbing the matrix E,

$$\hat{\mathcal{V}}(x) = \mathsf{E}^{-1}\,\mathcal{V}(x) = \begin{pmatrix} u^{ij}{}_{AB}(x) & -v^{kl\,AB}(x) \\ & \\ -v^{ij\,CD}(x) & u_{klCD}(x) \end{pmatrix}\,, \tag{5.43}$$

where we have to remember that $\hat{\mathcal{V}}$ is now no longer a group element of $E_{7(7)}$. Note, however, that the $E_{7(7)}$ tensors $\mathcal{Q}_\mu$ and $\mathcal{P}_\mu$ are not affected by the matrix E and have identical expressions in terms of $\mathcal{V}$ and $\hat{\mathcal{V}}$. This is not the case for the terms in the Lagrangian that contain the Abelian field strengths,

$$F_{\mu\nu}^{AB} = \partial_\mu A_\nu^{AB} - \partial_\nu A_\mu^{AB}\,, \tag{5.44}$$

and which take the form,

$$\begin{aligned} \mathcal{L}_3 \;=\; & -\tfrac{1}{8} e\, F_{\mu\nu}^{+AB}\, F^{+CD\,\mu\nu}\,[(u+v)^{-1}]^{AB}{}_{ij}\,(u^{ij}{}_{CD} - v^{ijCD}) \\ & -\tfrac{1}{2} e\, F_{\mu\nu}^{+AB}\,[(u+v)^{-1}]^{AB}{}_{ij}\,\mathcal{O}^{+\mu\nu\,ij} \\ & + \text{h.c.}\,, \end{aligned} \tag{5.45}$$

where the 28×28 matrices satisfy $[(u+v)^{-1}]^{AB}{}_{ij}\,(u^{ij}{}_{CD} + v^{ijCD}) = \delta^{AB}_{CD}$. The $SU(8)$ covariant field strength $\mathsf{F}^+_{\mu\nu ij}$ will appear in the supersymmetry transformation rules for the fermions, and is equal to

$$F_{\mu\nu}^{+AB} = (u^{ij}{}_{AB} + v^{ijAB})\,\mathsf{F}^+_{\mu\nu ij} - (u_{ij}{}^{AB} + v_{ijAB})\,\mathcal{O}^{+ij}_{\mu\nu}\,. \tag{5.46}$$

Clearly the Lagrangian depends on the matrix E. Because the matrix $\mathsf{E}^{-1}\mathcal{V}$ is an element of $Sp(56;\mathbf{R})$, the matrix multiplying the two field strengths in (5.45) is symmetric under the interchange of $[AB]\leftrightarrow[CD]$[24].

In order that the Lagrangian be invariant under a certain subgroup of $E_{7(7)}$, one has to make a certain choice for the matrix E. According to the analysis leading to (5.38) and (5.39), this subgroup is generated on $\hat{\mathcal{V}}$ by matrices Λ and Σ, just as in (5.1), but with indices $A,B,\ldots$, rather than with $I,J,\ldots$, satisfying

$$\mathrm{Im}\,\Big(\Sigma_{ABCD} + \Lambda_{AB}{}^{CD}\Big) = 0\,. \tag{5.47}$$

In order to be a subgroup of $E_{7(7)}$ as well, they must also satisfy (5.6), but only after a proper conversion of the $I,J,\ldots$ to $A,B,\ldots$ indices. The gauge fields transform under the real subgroup (, the imaginary parts of

[24]Such symmetry properties follow from the symmetry under interchanging index pairs in the products $(u^{ij}{}_{AB} - v^{ijAB})\,(u^{kl}{}_{AB} + v^{klAB})$ and $(u^{ij}{}_{AB} + v^{ijAB})\,(u_{ij}{}^{CD} + v_{ijCD})$.

the generators act exclusively on the 56-bein). A large variety of symmetry groups exists, as one can deduce from the symmetry groups that are realized in maximal supergravity in higher dimensions. The biggest group whose existence can be inferred in this way, is $E_{6(6)} \times SO(1,1)$, which is the group that one obtains from the $D = 5$ Lagrangian upon reduction to $D = 4$ dimensions.

5.4 Gauging maximal supergravity; the T-tensor

The gauging of supergravity is effected by switching on the gauge coupling constant, after assigning the various fields to representations of the gauge group embedded in $E_{7(7)}$ or $E_{6(6)}$. Only the gauge fields themselves and the spinless fields can transform under this gauge group. Hence the Abelian field strengths are changed to non-Abelian ones and derivatives of the scalars are covariantized according to

$$\partial_\mu \mathcal{V} \to \partial_\mu \mathcal{V} - g A_\mu^{AB}\, T_{AB} \mathcal{V}\,, \tag{5.48}$$

where the gauge group generators T_{AB} are 56×56 matrices which span a subalgebra of maximal dimension equal to the number of vector fields, embedded in the Lie algebra of $E_{7(7)}$ or $E_{6(6)}$. The structure constants of the gauge group are given by

$$[T_{AB}, T_{CD}] = f_{AB,CD}{}^{EF}\, T_{EF}\,. \tag{5.49}$$

It turns out that the viability for a gauging depends sensitively on the choice of the gauge group and its corresponding embedding. This aspect is most nontrivial for the $D = 4$ theory, in view of electric-magnetic duality. Therefore, we will mainly concentrate on this theory. In $D = 4$ dimensions, one must start from a Lagrangian that is symmetric under the desired gauge group, which requires one to make a suitable choice of the matrix E. In $D = 5$ dimensions, the Lagrangian is manifestly symmetric under $E_{6(6)}$, so this subtlety does not arise. When effecting the gauging the vector fields may decompose into those associated with the non-Abelian gauge group and a number of remaining gauge fields. When the latter are charged under the gauge group, then there is a potential obstruction to the gauging as the gauge invariance of these gauge fields cannot coexist with the non-Abelian gauge transformations. However, in $D = 5$ this obstruction can be avoided, because (charged) vector fields can alternatively be described as antisymmetric rank-2 tensor fields. For instance, the gauging of $SO(p, 6-p)$ involves 15 non-Abelian gauge fields and 12 antisymmetric tensor fields. The latter can transform under the gauge group, because they are not realized as tensor fields. Typically this conversion of vector into tensor fields leads to terms that are inversely proportional to the gauge coupling [68]. However, to write down a corresponding Lagrangian requires an even number of tensor fields.

Introducing the gauging leads directly to a loss of supersymmetry, because the new terms in the Lagrangian lead to new variations. For convenience we now restrict ourselves to $D = 4$ dimensions. The leading variations are induced by the modification (5.48) of the Cartan-Maurer equations. This modification was already noted in (4.57) and takes the form

$$\begin{aligned} F_{\mu\nu}(\mathcal{Q})_i{}^j &= -\tfrac{4}{3}\,\mathcal{P}_{[\mu}{}^{jklm}\,\mathcal{P}_{\nu]iklm} - g\,F_{\mu\nu}^{AB}\,\mathcal{Q}_{AB\,i}{}^j\,, \\ D_{[\mu}\mathcal{P}_{\nu]}^{ijkl} &= -\tfrac{1}{2}g\,F_{\mu\nu}^{AB}\,\mathcal{P}_{AB}^{ijkl}\,, \end{aligned} \tag{5.50}$$

where

$$\mathcal{V}^{-1}T_{AB}\mathcal{V} = \begin{pmatrix} \mathcal{Q}_{AB\,ij}{}^{mn} & \mathcal{P}_{AB\,ijpq} \\ & \\ \mathcal{P}_{AB}^{klmn} & \mathcal{Q}_{AB}{}^{kl}{}_{pq} \end{pmatrix}. \tag{5.51}$$

These modifications are the result of the implicit dependence of $\mathcal{Q}_\mu$ and $\mathcal{P}_\mu$ on the vector potentials A_μ^{AB}. The fact that the matrices T_{AB} generate a subalgebra of the algebra associated with $E_{7(7)}$, in the basis appropriate for $\mathcal{V}$, implies that the quantities $\mathcal{Q}_{AB}$ and $\mathcal{P}_{AB}$ satisfy the constraints,

$$\begin{aligned} \mathcal{P}_{AB}^{ijkl} &= \tfrac{1}{24}\,\varepsilon^{ijklmnpq}\,\mathcal{P}_{AB\,mnpq}\,, \\ \mathcal{Q}_{AB\,ij}{}^{kl} &= \delta_{[i}^{[k}\,\mathcal{Q}_{AB\,j]}{}^{l]}\,, \end{aligned} \tag{5.52}$$

while $\mathcal{Q}_{ABi}{}^j$ is antihermitean and traceless. It is straightforward to write down the explicit expressions for $\mathcal{Q}_{AB}$ and $\mathcal{P}_{AB}$,

$$\begin{aligned} \mathcal{Q}_{AB\,i}{}^j &= \tfrac{2}{3}\Big[u_{ik}{}^{IJ}\,(\Delta_{AB}u^{jk}{}_{IJ}) - v_{ikIJ}\,(\Delta_{AB}v^{jkIJ})\Big]\,, \\ \mathcal{P}_{AB}^{ijkl} &= v^{ijIJ}\,(\Delta_{AB}u^{kl}{}_{IJ}) - u^{ij}{}_{IJ}(\Delta_{AB}v^{klIJ}) \end{aligned} \tag{5.53}$$

where $\Delta_{AB}u$ and $\Delta_{AB}v$ indicate the change of submatrices in $\mathcal{V}$ induced by multiplication with the generator T_{AB}. Note that we could have expressed the above quantities in terms of the modified 56-bein $\hat{\mathcal{V}}$, on which the $E_{7(7)}$ transformations act in the basis that is appropriate for the field strengths, provided we change the generators T_{AB} into

$$\hat{T}_{AB} = \mathsf{E}^{-1}T_{AB}\,\mathsf{E}\,. \tag{5.54}$$

This is done below.

When establishing supersymmetry of the action one needs the Cartan-Maurer equations at an early stage to cancel variations from the gravitino kinetic terms and the Noether term (the term in the Lagrangian proportional to $\bar{\chi}\psi_\mu\mathcal{P}_\nu$). The order-$g$ terms in the Maurer-Cartan equation yield the leading variations of the Lagrangian. They are linearly proportional to the fermion fields and read,

$$\begin{aligned}\delta\mathcal{L} &= \tfrac{1}{4}g(\bar\epsilon_j\gamma^\rho\gamma^{\mu\nu}\psi^i_\rho - \bar\epsilon^i\gamma^\rho\gamma^{\mu\nu}\psi_{\rho j})\,\mathcal{Q}_{AB\,i}{}^j\,(u^{kl}{}_{AB}+v^{klAB})\,\mathsf{F}^+_{\mu\nu kl}\\ &\quad +\tfrac{1}{288}\varepsilon^{ijklmnpq}\,\bar\chi_{ijk}\gamma^{\mu\nu}\epsilon_l\,\mathcal{P}_{AB\,mnpq}\,(u^{rs}{}_{AB}+v^{rsAB})\,\mathsf{F}^+_{\mu\nu rs}\\ &\quad + \text{h.c.} \end{aligned}\tag{5.55}$$

The first variation is proportional to an $SU(8)$ tensor T_i^{jkl}, which is known as the T-tensor,

$$\begin{aligned} T_i^{jkl} &= \tfrac{3}{4}\mathcal{Q}_{AB\,i}{}^j\,(u^{kl}{}_{AB}+v^{klAB}) \qquad (5.56)\\ &= \tfrac{1}{2}\Big[u_{im}{}^{CD}\,(\hat\Delta_{AB}u^{jm}{}_{CD}) - v_{imCD}\,(\hat\Delta_{AB}v^{jmCD})\Big](u^{kl}{}_{AB}+v^{klAB})\,, \end{aligned}$$

where $\hat\Delta_{AB}u$ and $\hat\Delta_{AB}v$ are the submatrices of $\hat T_{AB}\hat{\mathcal{V}}$. Another component of the T-tensor appears in the second variation and is equal to

$$\begin{aligned} T_{ijkl}^{mn} &= \tfrac{1}{2}\mathcal{P}_{AB\,ijkl}\,(u^{mn}{}_{AB}+v^{mnAB}) \qquad (5.57)\\ &= \tfrac{1}{2}\Big[v_{ijCD}\,(\hat\Delta_{AB}u_{kl}{}^{CD}) - u_{ij}{}^{CD}(\hat\Delta_{AB}v_{klCD})\Big]\,(u^{mn}{}_{AB}+v^{mnAB})\,. \end{aligned}$$

The T-tensor is thus a cubic product of the 56-bein $\hat{\mathcal{V}}$ which depends in a nontrivial way on the embedding of the gauge group into $E_{7(7)}$. It satisfies a number of important properties. Some of them are rather obvious (such as $T_i^{ijk}=0$), and follow rather straightforwardly from the definition. We will concentrate on properties which are perhaps less obvious. Apart from the distinction between $\mathcal{V}$ and $\hat{\mathcal{V}}$, which is a special feature of $D=4$ dimensions, these properties are generic.

First we observe that $SU(8)$ covariantized variations of the T-tensor are again proportional to the T-tensor. These variations are induced by (5.27) and (5.28). Along the same lines as before we can show that the $SU(8)$ tensors $\mathcal{Q}_{AB}$ and $\mathcal{P}_{AB}$ transform according to the adjoint representation of $E_{7(7)}$, which allows one to derive,

$$\begin{aligned} \delta T_i^{jkl} &= \Sigma^{jmnp}\,T^{kl}_{imnp} - \tfrac{1}{24}\varepsilon^{jmnpqrst}\,\Sigma_{imnp}\,T^{kl}_{qrst} + \Sigma^{klmn}\,T^{j}_{imn}\,,\\ \delta T^{mn}_{ijkl} &= \tfrac{4}{3}\Sigma_{p[ijk}\,T^{pmn}_{l]} - \tfrac{1}{24}\varepsilon_{ijklpqrs}\,\Sigma^{mntu}\,T^{pqrs}_{tu}\,. \end{aligned}\tag{5.58}$$

This shows that the $SU(8)$ covariant T-tensors can be assigned to a representation of $E_{7(7)}$. This property will play an important role below.

Before completing the analysis leading to a consistent gauging we stress that all variations are from now on expressed in terms of the T-tensor, as its variations yield again the same tensor. This includes the $SU(8)$ covariant derivative of the T-tensor, which follows directly from (5.58) upon the substitutions $\delta \to D_\mu$ and $\Sigma \to \mathcal{P}_\mu$. A viable gauging requires that the T-tensor satisfies a number of rather nontrivial identities, as we will discuss shortly, but the new terms in the Lagrangian and transformation rules have a universal form, irrespective of the gauge group. Let us first describe these new terms. First of all, to cancel the variations (5.55) we need masslike terms in the Lagrangian,

$$\begin{aligned}\mathcal{L}_{\rm masslike} \;\; = \;\; & g\, e\Big\{\tfrac{1}{2}\sqrt{2}\, A_{1ij}\bar\psi^{\,i}_{\mu}\gamma^{\mu\nu}\psi^j_\nu + \tfrac{1}{6}A_{2i}^{jkl}\,\bar\psi^i_\mu\gamma^\mu\chi_{jkl} \\ & \qquad + A_3^{ijk,lmn}\,\bar\chi_{ijk}\chi_{lmn} + \text{h.c.}\Big\}\,, \end{aligned} \tag{5.59}$$

whose presence necessitates corresponding modifications of the supersymmetry transformations of the fermion fields,

$$\begin{aligned}\delta_g\bar\psi^i_\mu &= -\sqrt{2}g\, A_1^{ij}\,\bar\epsilon_j\gamma_\mu\,, \\ \delta_g\chi^{ijk} &= -2g\, A_{2l}{}^{ijk}\,\bar\epsilon^l\,. \end{aligned} \tag{5.60}$$

Finally at order g^2 one needs a potential for the spinless fields,

$$P(\mathcal{V}) = g^2\Big\{\tfrac{1}{24}|A_{2i}{}^{jkl}|^2 - \tfrac{1}{3}|A_1^{ij}|^2\Big\}. \tag{5.61}$$

These last three formulae will always apply, irrespective of the gauge group. Note that the tensors A_1^{ij}, $A_{2i}{}^{jkl}$ and $A_3^{ijk,lmn}$ have certain symmetry properties dictated by the way they appear in the Lagrangian (5.59). To be specific, A_1 is symmetric in (ij), A_2 is fully antisymmetric in $[jkl]$ and A_3 is antisymmetric in $[ijk]$ as well as in $[lmn]$ and symmetric under the interchange $[ijk] \leftrightarrow [lmn]$. This implies that these tensors transform under $SU(8)$ according to the representations

$$\begin{aligned} A_1\!: &\quad \mathbf{36}\,, \\ A_2\!: &\quad \mathbf{28} + \mathbf{420}\,, \\ A_3\!: &\quad \overline{\mathbf{28}} + \overline{\mathbf{420}} + \overline{\mathbf{1176}} + \overline{\mathbf{1512}}\,. \end{aligned}$$

The three $SU(8)$ covariant tensors, A_1, A_2 and A_3, which depend only on the spinless fields, must be linearly related to the T-tensor, because they were introduced for the purpose of cancelling the variations (5.55). To see how this can be the case, let us analyze the $SU(8)$ content of the T-tensor. As we mentioned already, the T-tensor is cubic in the 56-bein, and as such is constitutes a certain tensor that transforms under $E_{7(7)}$. The transformation properties were given in (5.58), where we made use of the fact that the T-tensor consists of a product of the fundamental times the adjoint representation of $E_{7(7)}$. Hence the T-tensor comprises the representations,

$$\mathbf{56}\times\mathbf{133} = \mathbf{56} + \mathbf{912} + \mathbf{6480}\,. \tag{5.62}$$

The representations on the right-hand side can be decomposed under the action of $SU(8)$, with the result

$$\begin{aligned}
\mathbf{56} &= \mathbf{28} + \overline{\mathbf{28}}\,, \\
\mathbf{912} &= \mathbf{36} + \overline{\mathbf{36}} + \mathbf{420} + \overline{\mathbf{420}}\,, \\
\mathbf{6480} &= \mathbf{28} + \overline{\mathbf{28}} + \mathbf{420} + \overline{\mathbf{420}} + \mathbf{1280} + \overline{\mathbf{1280}} + \mathbf{1512} + \overline{\mathbf{1512}}\,.
\end{aligned} \tag{5.63}$$

Comparing these representations to the $SU(8)$ representations to which the tensors $A_1 - A_3$ (and their complex conjugates) belong, we note that there is a mismatch between (5.63) and (5.62). In view of (5.58) the T-tensor must be restricted by suppressing complete representations of $E_{7(7)}$ in order that its variations and derivatives remain consistent. This proves that the T-tensor cannot contain the entire **6480** representation of $E_{7(7)}$, so that it must consist of the $\mathbf{28}+\mathbf{36}+\mathbf{420}$ representation of $SU(8)$ (and its complex conjugate). This implies that the T-tensor is decomposable into A_1 and A_2, whereas A_3 is not an independent tensor and can be expressed in terms of A_2. Indeed this was found by explicit calculation, which gave rise to

$$\begin{aligned}
T_i^{jkl} &= -\tfrac{3}{4} A_{2i}{}^{jkl} + \tfrac{3}{2} A_1^{j[k}\,\delta_i^{l]}\,, \\
T_{ijkl}^{mn} &= -\tfrac{4}{3}\delta_{[i}^{[m} T_{jkl]}^{n]}\,, \\
A_3^{ijk,lmn} &= -\tfrac{1}{108}\sqrt{2}\,\varepsilon^{ijkpqr[lm} T_{pqr}^{n]}\,.
\end{aligned} \tag{5.64}$$

Note that these conditions are necessary, but not sufficient as one also needs nontrivial identities quadratic in the T-tensors in order to deal with the variations of the Lagrangian of order g^2. One then finds that there is yet another constraint, which suppresses the **28** representation of the T-tensor,

$$T_i^{[jk]i} = 0\,. \tag{5.65}$$

Observe that a contraction with the first upper index is also equal to zero, as follows from the definition (5.56). Hence the T-tensor transforms under $E_{7(7)}$ according to the **912** representation which decomposes into the **36** and **420** representations of $SU(8)$ and their complex conjugates residing in the tensors A_1 and A_2, respectively,

$$A_1^{ij} = \tfrac{4}{21} T_k^{ikj}\,, \qquad A_{2i}^{jkl} = -\tfrac{4}{3} T_i^{[jkl]}\,. \tag{5.66}$$

Although we concentrated on the $D=4$ theory, we should stress once more that many of the above features are generic and apply in other dimensions. For instance, the unrestricted T-tensors in $D=5$ and 3 dimensions belong to the following representations of $E_{6(6)}$ and $E_{8(8)}$, respectively[25]

$$
\begin{aligned}
D &= 5\colon\ \mathbf{27}\times\mathbf{78}=\mathbf{27}+\mathbf{351}+\mathbf{1728}\,,\\
D &= 3\colon\ \mathbf{248}\times\mathbf{248}=\mathbf{1}+\mathbf{248}+\mathbf{3875}+\mathbf{27\,000}+\mathbf{30\,380}\,.
\end{aligned}
\tag{5.67}
$$

In these cases a successful gauging requires the T-tensor to be restricted to the **351** and the $\mathbf{1}+\mathbf{3875}$ representations, respectively, which decompose as follows under the action of $USp(8)$ and $SO(16)$,

$$
\begin{aligned}
\mathbf{351} &= \mathbf{36}+\mathbf{315}\,,\\
\mathbf{1}+\mathbf{3875} &= \mathbf{1}+\mathbf{135}+\mathbf{1820}+\mathbf{1920}\,.
\end{aligned}
\tag{5.68}
$$

These representations correspond to the tensors A_1 and A_2; for $D=5$ A_3 is again dependent while for $D=3$ there is an independent tensor A_3 associated with the **1820** representation of $SO(16)$.

We close with a few comments regarding the various gauge groups that have been considered. As we mentioned at the beginning of this section, the first gaugings were to some extent motivated by corresponding Kaluza-Klein compactifications. The S^7 and the S^4 [70] compactifications of 11-dimensional supergravity and the S^5 compactification of IIB supergravity, gave rise to the gauge groups $SO(8)$, $SO(5)$ and $SO(6)$, respectively. Non-compact gauge groups were initiated in [71] for the 4-dimensional theory; for the 5-dimensional theory they were also realized in [29] and in [72]. In $D=3$ dimensions there is no guidance from Kaluza-Klein compactifications and one has to rely on the group-theoretical analysis described above. In that case there exists a large variety of gauge groups of rather high dimension [69]. Gaugings can also be constructed a so-called Scherk-Schwarz reduction from higher dimensions [73]. To give a really exhaustive classification remains cumbersome. For explorations based on the group-theoretical analysis explained above, see [74, 75].

6 Supersymmetry in anti-de Sitter space

In Section 3.1 we presented the first steps in the construction of a generic supergravity theory, starting with the Einstein-Hilbert Lagrangian for gravity and the Rarita-Schwinger Lagrangians for the gravitino fields. We established the existence of two supersymmetric gravitational backgrounds, namely flat Minkowski space and anti-de Sitter space with a cosmological

[25] The $D=3$ theory has initially no vector fields, but those can be included by adding Chern-Simons terms. These terms lose their topological nature when gauging some of the $E_{8(8)}$ isometries [69].

constant proportional to g^2, where g was some real coupling constant proportional to the the inverse anti-de Sitter radius. The two cases are clearly related and flat space is obtained in the limit $g \to 0$, as can for instance be seen from the expression of the Riemann curvature ((3.14)),

$$R_{\mu\nu\rho}{}^{\sigma} = g^2(g_{\mu\rho}\,\delta^\sigma_\nu - g_{\nu\rho}\,\delta^\sigma_\mu)\,. \tag{6.1}$$

Because both flat Minkowski space and anti-de Sitter space are maximally symmetric, they have $\frac{1}{2}D(D+1)$ independent isometries which comprise the Poincaré group or the group $SO(D-1,2)$, respectively. The algebra of the combined bosonic and fermionic symmetries is called the anti-de Sitter superalgebra. Note again that the derivation in Section 3.1 was incomplete and in general one will need to introduce additional fields.

In this section we will mainly be dealing with simple anti-de Sitter supersymmetry and we will always assume that $3 < D \leq 7$. In that case the bosonic subalgebra coincides with the anti-de Sitter algebra. In $D = 3$ spacetime dimensions the anti-de Sitter group $SO(2,2)$ is not simple. There exist N-extended versions where one introduces N supercharges, each transforming as a spinor under the anti-de Sitter group. These N supercharges transform under a compact R-symmetry group, whose generators will appear in the $\{Q,\bar{Q}\}$ anticommutator. As we discussed in Section 2.5, the R-symmetry group is in general not the same as in Minkowski space; according to Table 9, we have $H_{\rm R} = SO(N)$ for $D = 4$, $H_{\rm R} = U(N)$ for $D = 5$, and $H_{\rm R} = USp(2N)$ for $D = 6, 7$. For $D > 7$ the superalgebra is no longer simple [3]; its bosonic subalgebra can no longer be restricted to the sum of the anti-de Sitter algebra and the R-symmetry algebra, but one needs extra bosonic generators that transform as high-rank antisymmetric tensors under the Lorentz group (see also [76]). In contrast to this, there exists an (N-extended) super-Poincaré algebra associated with flat Minkowski space of any dimension, whose bosonic generators correspond to the Poincaré group, possibly augmented with the R-symmetry generators associated with rotations of the supercharges.

Anti-de Sitter space is isomorphic to $SO(D-1,2)/SO(D-1,1)$ and thus belongs to the coset spaces that were discussed extensively in Section 4. According to (4.10) it is possible to describe anti-de Sitter space as a hypersurface in a $(D+1)$-dimensional embedding space. Denoting the extra coordinate of the embedding space by Y^-, so that we have coordinates Y^A with $A = -, 0, 1, 2, \ldots, D-1$, this hypersurface is defined by

$$-(Y^-)^2 - (Y^0)^2 + \vec{Y}^{\,2} = \eta_{AB}\,Y^AY^B = -g^{-2}\,. \tag{6.2}$$

Obviously, the hypersurface is invariant under linear transformations that leave the metric $\eta_{AB} = \mathrm{diag}\,(-,-,+,+,\ldots,+)$ invariant. These transformations constitute the group $SO(D-1,2)$. The $\frac{1}{2}D(D+1)$ generators denoted by M_{AB} act on the embedding coordinates by

$$M_{AB} = Y_A \frac{\partial}{\partial Y^B} - Y_B \frac{\partial}{\partial Y^A}\,, \tag{6.3}$$

where we lower and raise indices by contracting with η_{AB} and its inverse η^{AB}. It is now easy to evaluate the commutation relations for the M_{AB}

$$[M_{AB}, M_{CD}] = \eta_{BC}\, M_{AD} - \eta_{AC}\, M_{BD} - \eta_{BD}\, M_{AC} + \eta_{AD}\, M_{BC}\,. \tag{6.4}$$

Anti-de Sitter space has the topology of S^1[time] $\times$ $\mathbf{R}^{D-1}$[space] and has closed timelike curves. These curves can be avoided by unwrapping S^1, so that one finds the universal covering space denoted by CadS, which has the topology of $\mathbf{R}^D$. There exist no Cauchy surfaces in this space. Any attempt to determine the outcome of some evolution or wave equation from a spacelike surface requires fresh information coming from a timelike infinity which takes a finite amount of time to arrive [77, 78]. Spatial infinity is a timelike surface which cannot be reached by timelike geodesics. There are many ways to coordinatize anti-de Sitter space, but we will avoid using explicit coordinates.

For later use we record the (simple) anti-de Sitter superalgebra, which in addition to (6.4) contains the (anti-)commutation relations

$$\begin{aligned} \{Q_\alpha, \bar Q_\beta\} &= -\tfrac{1}{2}(\Gamma_{AB})_{\alpha\beta}\, M^{AB}\,, \\ [M_{AB}, \bar Q_\alpha] &= \tfrac{1}{2}(\bar Q\, \Gamma_{AB})_\alpha\,. \end{aligned} \tag{6.5}$$

Here the matrices Γ_{AB}, which will be defined later, are the generators of $SO(D-1,2)$ group in the spinor representation. As we alluded to earlier, this algebra changes its form when considering N supercharges which rotate under R-symmetry, because the R-symmetry generators will appear on the right-hand side of the $\{Q, \bar Q\}$ anticommutator.

The relation with the Minkowski case proceeds by means of a so-called Wigner-Inönü contraction. Here one rescales the generators according to $M_{-A} \to g^{-1} P_A$, $Q \to g^{-1/2} Q$, keeping the remaining generators M_{AB} corresponding to the Lorentz subalgebra unchanged. In the limit $g \to 0$, the generators P_A will form a commuting subalgebra and the full algebra contracts to the super-Poincaré algebra.

On spinors, the anti-de Sitter algebra can be realized by the following combination of gamma matrices Γ_a in D-dimensional Minkowski space,

$$M_{AB} = \tfrac{1}{2}\Gamma_{AB} = \begin{cases} \tfrac{1}{2}\Gamma_{ab} & \text{for} \quad A, B = a, b\,, \\[2ex] \tfrac{1}{2}\Gamma_a & \text{for} \quad A = -\,, B = a \end{cases} \tag{6.6}$$

with $a, b = 0, 1, \ldots, D-1$. Our gamma matrices satisfy the Clifford property $\{\Gamma^a\,,\, \Gamma^b\} = 2\,\eta^{ab}\,\mathbf{1}$, where $\eta^{ab} = \mathrm{diag}\,(-, +, \ldots, +)$ is the D-dimensional

Lorentz-invariant metric[26]. Concerning the R-symmetry group in anti-de Sitter space, the reader is advised to consult Section 2.5.

Of central importance is the quadratic Casimir operator of the isometry group $SO(D-1,2)$, defined by

$$\mathcal{C}_2 = -\tfrac{1}{2}\, M^{AB}\, M_{AB}\,. \tag{6.7}$$

The group $SO(D-1,2)$ has more Casimir operators when $D>3$, but these are of higher order in the generators and will not play a role in the following. To make contact between the masslike terms in the wave equations and the properties of the irreducible representations of the anti-de Sitter group, which we will discuss in Section 6.1, it is important that we establish the relation between the wave operator for fields that live in anti-de Sitter space, which involves the appropriately covariantized D'Alembertian $\Box_{adS}$, and the quadratic Casimir operator $\mathcal{C}_2$. We remind the reader that fields in anti-de Sitter space are multi-component functions of the anti-de Sitter coordinates that rotate irreducibly under the action of the Lorentz group $SO(D-1,1)$. The appropriate formulae were given at the end of Section 4.2 ((4.44) and (4.45)) and from them one can derive,

$$\mathcal{C}_2 = \Box_{adS}\Big|_{g=1} + \mathcal{C}_2^{\rm Lorentz}\,, \tag{6.8}$$

where $\mathcal{C}_2^{\rm Lorentz}$ is the quadratic Casimir operator for the representation of the Lorentz group to which the fields have been assigned. This result can be proven for any symmetric, homogeneous, space (see, for example [79]). For scalar fields, the second term in (6.8) vanishes and the proof is elementary (see, [80]).

Let us now briefly return to the supersymmetry algebra as it is realized on the vielbein field. Using the transformation rules (3.11) The commutator of two supersymmetry transformations yields an infinitesimal general-coordinate transformation and a tangent-space Lorentz transformation. For example, we obtain for the vielbein,

$$\begin{aligned}
[\delta_1,\delta_2]\, e_\mu{}^a &= \tfrac{1}{2}\bar\epsilon_2\,\Gamma^a\,\delta_1\psi_\mu - \tfrac{1}{2}\bar\epsilon_1\,\Gamma^a\,\delta_2\psi_\mu \\
&= D_\mu(\tfrac{1}{2}\bar\epsilon_2\,\Gamma^a\epsilon_1) + \tfrac{1}{2}g\,(\bar\epsilon_2\,\Gamma^{ab}\epsilon_1)\,e_{\mu b}\,.
\end{aligned} \tag{6.9}$$

The first term corresponds to a spacetime diffeomorphism and the second one to a tangent space (local Lorentz) transformation. Here we consider only the gravitational sector of the theory; for a complete theory there are additional contributions, but nevertheless the above terms remain and

[26]Note that when the gravitino is a Majorana spinor, the quantities $\Gamma_{AB}\epsilon$ should satisfy the same Majorana constraint.

(6.9) should be realized uniformly on all the fields. In the anti-de Sitter background, were the gravitino field vanishes, the parameters of the supersymmetry transformations are Killing spinors satisfying (3.15) so that the gravitino field remains zero under supersymmetry. Therefore both the gravitino and the vielbein are left invariant under supersymmetry, so that the combination of symmetries on the right-hand side of (6.9) should vanish when ϵ_1 and ϵ_2 are Killing spinors. Indeed, the diffeomorphism with parameter $\xi^\mu = \frac{1}{2}\bar\epsilon_2\,\Gamma^\mu\epsilon_1$, is an anti-de Sitter Killing vector (, it satisfies (4.38)), because $D_\mu(g\,\bar\epsilon_2\Gamma_{\nu\rho}\epsilon_1) = -g^2\,g_{\mu[\rho}\xi_{\nu]}$ is antisymmetric in μ and ν. As for all Killing vectors, higher derivatives can be decomposed into the Killing vector and its first derivative. Indeed, we find $D_\mu(g\,\bar\epsilon_2\Gamma_{\nu\rho}\epsilon_1) = -g^2\,g_{\mu[\rho}\xi_{\nu]}$ in the case at hand. The Killing vector can be decomposed into the $\frac{1}{2}D(D+1)$ Killing vectors of the anti-de Sitter space. The last term in (6.9) is a compensating target space transformation of the type we have been discussing extensively in Section 4.2 for generic coset spaces.

6.1 Anti-de Sitter supersymmetry and masslike terms

In flat Minkowski space all fields belonging to a supermultiplet are subject to field equations with the same mass, because the momentum operators commute with the supersymmetry charges, so that P^2 is a Casimir operator. For supermultiplets in anti-de Sitter space this is no longer the case, so that masslike terms will not necessarily be the same for different fields belonging to the same multiplet. We have already discussed the interpretation of masslike terms for the gravitino, following (3.10). This phenomenon will be now illustrated below in a specific example, namely a scalar chiral supermultiplet in $D=4$ spacetime dimensions. Further clarification from an algebraic viewpoint will be given later in Section 6.3.

A scalar chiral supermultiplet in 4 spacetime dimensions consists of a scalar field A, a pseudoscalar field B and a Majorana spinor field ψ. In anti-de Sitter space the supersymmetry transformations of the fields are proportional to a spinor parameter $\epsilon(x)$, which is a Killing spinor in the anti-de Sitter space, , $\epsilon(x)$ must satisfy the Killing spinor equation (3.15). In the notation of this section, this equation reads

$$\Big(\partial_\mu - \tfrac{1}{4}\omega^{ab}\gamma_{ab} + \tfrac{1}{2}g\,e_\mu{}^a\gamma_a\Big)\epsilon = 0\,, \tag{6.1}$$

where we made the anti-de Sitter vierbein and spin connection explicit. We allow for two constants a and b in the supersymmetry transformations, which we parametrize as follows,

$$\begin{aligned} \delta A &= \tfrac{1}{4}\bar\epsilon\psi\,, \qquad \delta B = \tfrac{1}{4}i\bar\epsilon\gamma_5\psi\,,\\ \delta\psi &= \partial\!\!\!/(A + i\gamma_5 B)\epsilon - (a\,A + ib\,\gamma_5 B)\epsilon\,. \end{aligned} \tag{6.2}$$

In this expression the anti-de Sitter vierbein field has been used to contract the gamma matrix with the derivative. The coefficient of the first term in $\delta\psi$ has been chosen such as to ensure that $[\delta_1, \delta_2]$ yields the correct coordinate transformation $\xi^\mu D_\mu$ on the fields A and B. To determine the constants a and b and the field equations of the chiral multiplet, we consider the closure of the supersymmetry algebra on the spinor field. After some Fierz reordering we obtain the result,

$$[\delta_1, \delta_2]\psi = \xi^\mu D_\mu\psi + \tfrac{1}{16}(a-b)\,\bar\epsilon_2\gamma^{ab}\epsilon_1\,\gamma_{ab}\psi - \tfrac{1}{2}\xi^\rho\gamma_\rho[\not{D}\psi + \tfrac{1}{2}(a+b)\psi]\,. \qquad (6.3)$$

We point out that derivatives acting on $\epsilon(x)$ occur in this calculation at an intermediate stage and should not be suppressed in view of (6.1). However, they produce terms proportional to g which turn out to cancel in the above commutator. Now we note that the right-hand side should constitute a coordinate transformation and a Lorentz transformation, possibly up to a field equation. Obviously, the coordinate transformation coincides with (6.9) but the correct Lorentz transformation is only reproduced provided that $a-b=2g$. If we now define $m=\frac{1}{2}(a+b)$, so that the last term is just the Dirac equation with mass m, we find

$$a = m+g\,, \qquad b = m-g\,. \qquad (6.4)$$

Consequently, the supersymmetry transformation of ψ equals

$$\delta\psi = \not{D}(A+i\gamma_5B)\epsilon - m(A+i\gamma_5B)\epsilon - g(A-i\gamma_5B)\,\epsilon\,, \qquad (6.5)$$

and the fermionic field equation equals $(\not{D}+m)\psi=0$. The second term in (6.5), which is proportional to m, can be accounted for by introducing an auxiliary field to the supermultiplet. The third term, which is proportional to g, can be understood as a compensating S-supersymmetry transformation associated with auxiliary fields in the supergravity sector (see, [81]). In order to construct the corresponding field equations for A and B, we consider the variation of the fermionic field equation. Again we have to take into account that derivatives on the supersymmetry parameter are not equal to zero. This yields the following second-order differential equations

$$\begin{aligned}
[\Box_{adS} + 2g^2 - m(m-g)]\,A &= 0\,,\\
[\Box_{adS} + 2g^2 - m(m+g)]\,B &= 0\,,\\
[\Box_{adS} + 3g^2 - m^2]\,\psi &= 0\,.
\end{aligned} \qquad (6.6)$$

The last equation follows from the Dirac equation. Namely, one evaluates $(\not{D}-m)(\not{D}+m)\psi$, which gives rise to the wave operator $\Box_{adS}+\frac{1}{2}[\not{D},\not{D}]-m^2$. The commutator yields the Riemann curvature of the anti-de Sitter space. In an anti-de Sitter space of arbitrary dimension D this equation then reads

$$[\Box_{adS} + \tfrac{1}{4}D(D-1)g^2 - m^2]\psi = 0\,, \qquad (6.7)$$

which, for $D = 4$ agrees with the last equation of (6.6). A striking feature of the above result is that the field equations (6.6) all have different mass terms, in spite of the fact that they belong to the same supermultiplet [82]. Consequently, the role of mass is quite different in anti-de Sitter space as compared to flat Minkowski space. This will be elucidated later in Section 6.2.

For future applications we also evaluate the Proca equation for a massive vector field,

$$D^\mu(\partial_\mu A_\nu - \partial_\nu A_\mu) - m^2\, A_\nu = 0\,. \qquad (6.8)$$

This leads to[27] $D^\mu A_\mu = 0$, so that (6.8) reads $D^2 A_\nu - [D^\mu, D_\nu]A_\mu - m^2\, A_\nu = 0$ or, in anti-de Sitter space

$$[\Box_{adS} + (D-1)g^2 - m^2]\, A_\mu = 0\,. \qquad (6.9)$$

This can be generalized to an antisymmetric tensor of rank n, whose field equation reads (antisymmetrizing over indices $\nu_1, \ldots, \nu_n$)

$$(n+1)\, D^\mu \partial_{[\mu} C_{\nu_1\cdots\nu_n]} - m^2\, C_{\nu_1\cdots\nu_n} = 0\,. \qquad (6.10)$$

In the same way as before, this leads to

$$[\Box_{adS} + n(D-n)g^2 - m^2]\, C_{\nu_1\cdots\nu_n} = 0\,. \qquad (6.11)$$

The g^2 term in the field equations for the scalar fields can be understood from the observation that the scalar D'Alembertian (in an arbitrary gravitational background) can be extended to a conformally invariant operator (see, [81]),

$$\Box + \frac{1}{4}\frac{D-2}{D-1}\, R = \Box + \tfrac{1}{4}D(D-2)\, g^2\,, \qquad (6.12)$$

which seems the obvious candidate for a massless wave operator for scalar fields. Indeed, for $D = 4$, we do reproduce the g^2 dependence in the first two equations (6.6). Observe that the Dirac operator $\not{D}$ is also conformally invariant and so is the wave equation associated with the Maxwell field.

Using (6.8) we can now determine the values for the quadratic Casimir operator for the representations described by scalar, spinor, vector and tensor fields. The quadratic Casimir operator of the Lorentz group takes the values 0, $\frac{1}{8}D(D-1)$, $D-1$ and $n(D-n)$ for scalar, spinor, vector and tensor fields respectively. Combining this result with (6.12, 6.7, 6.9) and

[27] When $m \neq 0$, otherwise one can impose this equation as a gauge condition.

(6.10, 6.8) yields the following values for the quadratic Casimir operators,

$$\begin{aligned}
\mathcal{C}_2^{\text{scalar}} &= -\tfrac{1}{4}D(D-2)+m^2\,,\\
\mathcal{C}_2^{\text{spinor}} &= -\tfrac{1}{8}D(D-1)+m^2\,,\\
\mathcal{C}_2^{\text{vector}} &= m^2\,,\\
\mathcal{C}_2^{\text{tensor}} &= m^2\,.
\end{aligned}\tag{6.13}$$

For spinless fields, m^2 is . . . the coefficient in the mass term of the Klein-Gordon equation, as that is given by the value for $\mathcal{C}_2^{\text{scalar}}$, while, for spinor and vector fields, m and m^2 do correspond to the mass terms in the Dirac and Proca equations, respectively. We thus derive specific values for $\mathcal{C}_2$ for massless scalar, spinor, vector and tensor fields upon putting $m=0$. The fact that the value of $\mathcal{C}_2$ does not depend on the rank of the tensor field is in accord with the fact that, for $m^2=0$, a rank-n and a rank-$(D-n-2)$ tensor gauge field are equivalent on shell (also in curved space).

Hence, we see that the interpretation of the mass parameter is not straightforward in the context of anti-de Sitter space. In the next section we will derive a rather general lower bound on the value of $\mathcal{C}_2$ for the lowest-weight representations of the anti-de Sitter algebra ((6.26)), which implies that the masslike terms for scalar fields can have a negative coefficient μ^2 subject to the inequality,

$$\mu^2 \geq -\tfrac{1}{4}(D-1)^2\,. \tag{6.14}$$

This result is known as the the Breitenlohner-Freedman bound [82], which ensures the stability of an anti-de Sitter background against small fluctuations of the scalar fields. For spin-$\frac{1}{2}$ the bound on $\mathcal{C}_2$ implies that $m^2 \geq 0$, whereas for spin-0 we find that $m^2 \geq -\frac{1}{4}$, with m^2 as defined in (6.13).

In the next section we study unitary representations of the anti-de Sitter algebra and this study will confirm some of the results found above. For massless representations of higher spin there is a decoupling of degrees of freedom, which uniquely identifies the massless representations and their values of $\mathcal{C}_2$. In a number of cases this decoupling is more extreme and one obtains a so-called singleton representation which does not have a smooth Poincaré limit. In those cases there is no decoupling of a representation that could be identified as massless and therefore there remains a certain ambiguity in the definition of "massless" representations. This can also be seen from the observation that (massless) antisymmetric tensor gauge fields of rank $n=D-2$ are on-shell equivalent to massless scalar fields. While we concluded above that these tensor fields lead to $\mathcal{C}_2=0$, massless scalar fields have $\mathcal{C}_2=-\frac{1}{4}D(D-2)$ according to (6.13). The difference may not be entirely surprising in view of the fact that the antisymmetric tensor Lagrangian is not conformally invariant for arbitrary values of D, contrary to the scalar field Lagrangian. At any rate, we have established the existence of two different field representations that describe massless, spinless states which correspond to different values for the anti-de Sitter Casimir operator $\mathcal{C}_2$. At the end of the next section, where we discuss unitary representations of the anti-de Sitter algebra, we briefly return to the issue of massless representations. The connection between the local field theory description and the anti-de Sitter representations tends to be subtle.

6.2 Unitary representations of the anti-de Sitter algebra

In this section we discuss unitary representations of the anti-de Sitter algebra. We refer to [83] for some of the original work, and to [80,84,85] where part of this work was reviewed. In order to underline the general features we will stay as much as possible in general spacetime dimensions $D > 3$[28]. The anti-de Sitter isometry group, $SO(D-1,2)$, is noncompact, which implies that unitary representations will be infinitely dimensional. For these representations the generators are anti-hermitean,

$$M^\dagger_{AB} = -M_{AB}\,. \tag{6.15}$$

Here we note that the cover group of $SO(D-1,2)$ has the generators $\frac{1}{2}\Gamma_{\mu\nu}$ and $\frac{1}{2}\Gamma_\mu$, acting on spinors which are finite-dimensional objects. These generators, however, have different hermiticity properties.

The compact subgroup of the anti-de Sitter group is $SO(2)\times SO(D-1)$ corresponding to rotations of the compact anti-de Sitter time and spatial rotations. It is convenient to decompose the $\frac{1}{2}D(D+1)$ generators as follows. First, the generator M_{-0} is related to the energy operator when the radius of the anti-de Sitter space is taken to infinity. The eigenvalues of this generator, associated with motions along the circle, are quantized in integer units in order to have single-valued functions, unless one passes to the covering space CadS. The energy operator H will thus be defined as

$$H = -iM_{-0}\,. \tag{6.16}$$

Obviously the generators of the spatial rotations are the operators M_{ab} with $a,b=1,\ldots,D-1$. Note that we have changed notation: here and henceforth in this section the indices $a,b,\ldots$ refer to space indices. The remaining generators M_{-a} and M_{0a} are combined into $D-1$ pairs of mutually conjugate operators,

$$M^\pm_a = -iM_{0a} \pm M_{-a}\,, \tag{6.17}$$

[28] The case of $D=3$ is special because $SO(2,2)\cong (SL(2,R)\times SL(2,R))/\mathbf{Z}_2$.

satisfying $(M_a^+)^\dagger = M_a^-$. The anti-de Sitter commutation relations then read

$$\begin{aligned}[H, M_a^\pm] &= \pm M_a^\pm\,,\\ [M_a^\pm, M_b^\pm] &= 0\,,\\ [M_a^+, M_b^-] &= -2(H\,\delta_{ab} + M_{ab})\,.\end{aligned} \tag{6.18}$$

Obviously, the $M_a^\pm$ play the role of raising and lowering operators: when applied to an eigenstate of H with eigenvalue E, application of $M_a^\pm$ yields a state with eigenvalue $E \pm 1$. We also give the Casimir operator in this basis,

$$\begin{aligned}\mathcal{C}_2 &= -\tfrac{1}{2}M^{AB}M_{AB}\\ &= H^2 - \tfrac{1}{2}\{M_a^+, M_a^-\} - \tfrac{1}{2}(M_{ab})^2\\ &= H(H-D+1) + J^2 - M_a^+ M_a^-\,,\end{aligned} \tag{6.19}$$

where J^2 is the total spin operator: the quadratic Casimir operator of the rotation group $SO(D-1)$, defined by

$$J^2 = -\tfrac{1}{2}(M_{ab})^2\,. \tag{6.20}$$

In simple cases, its value is well known. For $D = 4$ it is expressed in terms of the "spin" s which is an integer for bosons and a half-integer for fermions and the spin-s representation has dimension $2s+1$ and $J^2 = s(s+1)$. For $D = 5$, the corresponding rotation group $SO(4)$ is the product of two $SU(2)$ groups, so that irreducible representations are characterized by two spin values, (s_+, s_-). Their dimension is equal to $(2s_+ + 1)(2s_- + 1)$ and $J^2 = 2(J_+^2 + J_-^2)$ with $J_\pm^2 = s_\pm(s_\pm + 1)$. Summarizing:

$$J^2 = \begin{cases} s(s+1) & \text{for} \quad D = 4\,,\\ 2s_+(s_+ + 1) + 2s_-(s_- + 1) & \text{for} \quad D = 5\,.\end{cases} \tag{6.21}$$

The $SO(D-1)$ representations for $D > 5$ are specified by giving the eigenvalues of additional (higher-order) $SO(D-1)$ Casimir operators. A restricted class of representations will be discussed in a sequel; a more general discussion of all possible representations requires a more technical set-up and is outside the scope of these lectures.

In this section we restrict ourselves to the bosonic case, but in passing, let us already briefly indicate how some of the other (anti-)commutators of the simple anti-de Sitter superalgebra decompose ((6.5))

$$\begin{aligned}\{Q_\alpha\,, Q_\beta^\dagger\} &= H\,\delta_{\alpha\beta} - \tfrac{1}{2}iM_{ab}\,(\Gamma^a\Gamma^b\Gamma^0)_{\alpha\beta}\\ &\quad + \tfrac{1}{2}(M_a^+\,\Gamma^a\,(1+i\Gamma^0) + M_a^-\,\Gamma^a\,(1-i\Gamma^0))_{\alpha\beta}\,,\end{aligned}$$

$$\begin{aligned} [H\,, Q_\alpha] &= -\tfrac{1}{2} i (\Gamma^0\, Q)_\alpha\,, \\ [M_a^\pm\,, Q_\alpha] &= \mp\tfrac{1}{2}(\Gamma_a (1 \mp i\Gamma^0)\, Q)_\alpha\,. \end{aligned} \tag{6.22}$$

For the anti-de Sitter superalgebra, all the bosonic operators can be expressed as bilinears of the supercharges, so that in principle one could restrict oneself to fermionic operators only and employ the projections $(1 \pm i\Gamma^0)Q$ as the basic lowering and raising operators. This will be discussed later in Section 6.3.

We now turn to irreducible representations of the anti-de Sitter algebra (6.18). We start with the observation that the energy operator can be diagonalized so that we can label the states according to their eigenvalue E. Because application of $M_a^\pm$ leads to the states with higher and lower eigenvalues E, we expect the representation to cover an infinite range of eigenvalues, all separated by integers. For a unitary representation the $M_a^+ M_a^-$ term in (6.19) is positive, which implies that the Casimir operator is bounded by

$$\mathcal{C}_2 \leq -\tfrac{1}{4}(D-1)^2 + \Big[J^2 + \big(E - \tfrac{1}{2}(D-1)\big)^2\Big]_{\rm minimal}\,, \tag{6.23}$$

where the subscript indicates that one must choose the minimal value that $J^2 + (E - \frac{1}{2}(D-1))^2$ takes in the representation. Among other things, this number will depend on whether the eigenvalues E take integer or half-integer values.

Continuous representations cover the whole range of eigenvalues E extending from $-\infty$ to ∞. However, when there is a state with some eigenvalue E_0 that is annihilated by all the M_a^-, then only states with eigenvalues $E > E_0$ will appear in the representation. This is therefore not a continuous representation but a so-called lowest-weight representation. The ground state of this representation (which itself transforms as an irreducible representation of the rotation group and may thus be degenerate) is denoted by $|E_0, J\rangle$ and satisfies

$$M_a^-\, |E_0, J\rangle = 0\,. \tag{6.24}$$

The unitarity upper bound (6.23) on $\mathcal{C}_2$ is primarily useful for continuous representations. For unitary lowest-weight representations one can derive various lower bounds, as we shall see below. Substituting the condition (6.24) in the expression (6.18) applied to the ground state $|E_0, J\rangle$, we derive at once the eigenvalue of the quadratic Casimir operator associated with this representation in terms of E_0 and J^2,

$$\mathcal{C}_2 = E_0(E_0 - D + 1) + J^2\,. \tag{6.25}$$

Since $\mathcal{C}_2$ is a Casimir operator, this value holds for any state belonging to the corresponding irreducible representation. For real values of E_0 the Casimir

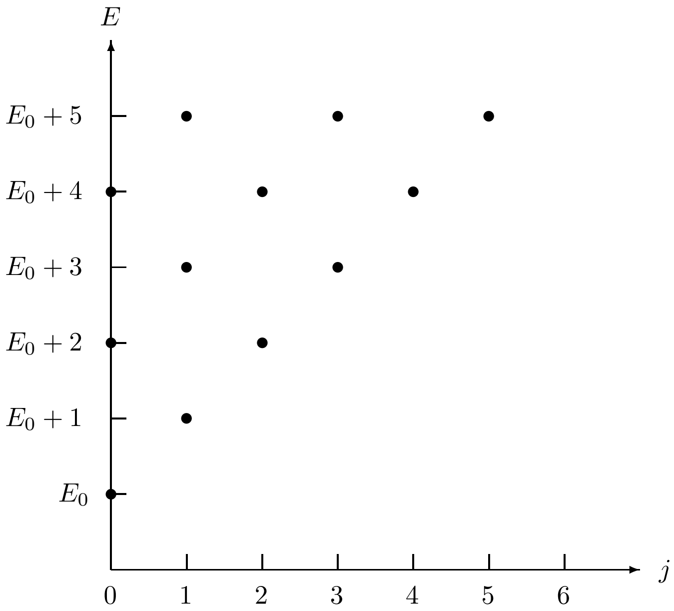


Fig. 1. States of the spinless representation in terms of the energy eigenvalues E and the angular momentum j. Each point corresponds to the spherical harmonics of S^{D-1}: traceless, symmetric tensors $Y^{a_1\cdots a_l}$ of rank $l = j$.

operator is bounded by

$$\mathcal{C}_2 \geq J^2 - \tfrac{1}{4}(D-1)^2\,. \tag{6.26}$$

As we already discussed at the end of the previous section, for scalar fields ($J^2 = 0$) this is just the Breitenlohner-Freedman bound [82]. Additional restrictions based on unitarity will be derived shortly. They generally lead to a lower bound for E_0 and thus to a corresponding lower bound for $\mathcal{C}_2$. Unless this bound supersedes (6.26) there can exist a degeneracy in the sense that there are two possible, permissible values for E_0 with the same value for $\mathcal{C}_2$. These two values correspond to two different solutions of the field equations subject to different boundary conditions at spatial infinity.

In what follows we restrict ourselves to lowest-weight representations, because those have a natural interpretation in the limit of large anti-de Sitter radius in terms of Poincaré representations. Alternatively we can construct highest-weight representations, but those will be similar and need not to be discussed separately.

Table 17. Two generic $SO(D-1)$ representations. One is the symmetric traceless tensor representation (corresponding to the spherical harmonics on S^{D-2}) denoted by l-rank tensors $Y^{A_1\cdots A_l}$, and the representation spanned by mixed tensors $Y^{B;A_1\cdots A_l}$ of rank $l+1$ (which is not independent for $D=4$). We list the corresponding eigenvalues of the quadratic Casimir operator J^2, for general D. For $D=4$ these representations are characterized by an integer spin s. For $D=5$ there are two such numbers, $s_\pm$, as we explained in the text.

representation	$Y^{A_1\cdots A_l}$	$Y^{B;A_1\cdots A_l}$
D	$l(l+D-3)$	$(l+D-4)(l+1)$
$D=4$	$s=l$	$s=l$
$D=5$	$s_\pm=\frac{1}{2}l$	$s_\pm=s_\mp+1=\frac{1}{2}(l+1)$

The full lowest-weight representation can now be constructed by acting with the raising operators on the ground state $\vert E_0,J\rangle$. To be precise, all states of energy $E=E_0+n$ are constructed by an n-fold product of creation operators M_a^+ In this way one obtains states of higher eigenvalues E with higher spin. The simplest case is the one where the vacuum has no spin ($J=0$). For given eigenvalue E, the states decompose into the state of the highest spin generated by the traceless symmetric product of $E-E_0$ operators M_a^+ and a number of lower-spin descendants. These states are all shown in Figure 1.

In the following we consider a number of representations of $SO(D-1)$ that exist for any dimension. For the bosons we consider the spherical harmonics, spanned by l-rank traceless, symmetric tensors $Y^{a_1\cdots a_l}$. Multiplying such tensors with the vector representation gives rise to two of these representations with rank $l\pm 1$, and a "mixed" representation, spanned by mixed tensors $Y^{b;a_1\cdots a_l}$. Table 17 lists the value of J^2 for these representations, for general D and for the specific cases of $D=4,5$. In a similar Table 18 we list the value of J^2 for the irreducible symmetric tensor-spinors, denoted by $Y^{\alpha;a_1\cdots a_l}$. They are symmetric l-rank tensor spinors that vanish upon contraction by a gamma matrix and appear when taking products of spherical harmonics with a simple spinor.

Armed with this information it is straightforward to find the decompositions of the spinor representation of the anti-de Sitter algebra. One simply takes the direct product of the spinless representation with a spin-$\frac{1}{2}$ state. That implies that every point with spin j in Figure 1 generates two points with spin $j\pm\frac{1}{2}$, with the exception of points associated with $j=0$, which will simply move to $j=\frac{1}{2}$. The result of this is shown in Figure 2.

Table 18. The eigenvalues of the quadratic $SO(D-1)$ Casimir operator J^2 for the symmetric tensor-spinor representation spanned by tensors $Y^{\alpha;a_1\cdots a_l}$ for general dimension D and for the specific cases $D=4,5$.

representation	$Y^{\alpha;a_1\cdots a_l}$
J^2	$l(l+D-2)+\frac{1}{8}(D-1)(D-2)$
$D=4$	$s=l+\frac{1}{2}$
$D=5$	$s_\pm = s_\mp - \frac{1}{2} = \frac{1}{2}l$

However, the spinless and the spinor representations that we have constructed so far are not necessarily irreducible. To see this consider the excited state that has the same spin content as the ground state, but with an energy equal to E_0+2 or E_0+1, for the scalar and spinor representations, respectively, and compare their value for the Casimir operator with that of the corresponding ground state. In this way we find for the scalar

$$E_0(E_0-D+1) = (E_0+2)(E_0-D+3) + \Big| M_a^- \,|E_0+2, \text{spinless}\rangle \Big|^2 . \quad (6.27)$$

This leads to

$$2E_0+3-D = \tfrac{1}{2}\Big| M_a^- \,|E_0+2, \text{spinless}\rangle \Big|^2 , \qquad (6.28)$$

so that unitary of the representation requires the inequality,

$$E_0 \geq \tfrac{1}{2}(D-3) . \qquad (6.29)$$

For $E_0 = \frac{1}{2}(D-3)$ we have the so-called [illegible] representation[29], where we have only one state for each given spherical harmonic. The Casimir eigenvalue for this representation equals

$$\mathcal{C}_2(\text{spinless singleton}) = -\tfrac{1}{4}(D+1)(D-3) . \qquad (6.30)$$

The excited state then constitutes the ground state for a separate irreducible spinless representation, but now with $E_0 = \frac{1}{2}(D+1)$, which, not surprisingly, has the same value for $\mathcal{C}_2$.

[29] The singleton representation was first found by Dirac [86] in 4-dimensional anti-de Sitter space and was known as a "remarkable representation". In the context of the oscillator method, which we will refer to later, singletons in anti-de Sitter spaces of dimension $D \neq 4$ are called "doubletons" [87]. In these lectures we will only use the name singleton to denote these remarkable representations.

For the spinor representation one finds a similar result,

$$\begin{aligned} E_0(E_0 - D + 1) + J^2 &= (E_0 + 1)(E_0 - D + 2) + J^2 \\ &\quad + \Big| M_a^- |E_0 + 1, \text{spinor}\rangle \Big|^2 . \end{aligned} \tag{6.31}$$

As the value for J^2 are the same for the ground state and the excited state one readily derives

$$2E_0 - D + 3 = \Big| M_a^- |E_0 + 1, \text{spinor}\rangle \Big|^2 , \tag{6.32}$$

so that one obtains the unitarity bound

$$E_0 \geq \tfrac{1}{2}(D-2) . \tag{6.33}$$

For $E_0 = \frac{1}{2}(D-2)$ we have the spinor singleton representation, which again consists of just one state for every value of the total spin. For the spinor representation the value of the Casimir operator equals

$$\mathcal{C}_2(\text{spinor singleton}) = -\tfrac{1}{8}(D+1)(D-2) . \tag{6.34}$$

Note that in $D = 4$, both singleton representations have the same eigenvalue of the Casimir operator.

The existence of the singletons was first noted by Dirac [86]. These representations are characterized by the fact that they do not exist in the Poincaré limit. To see this, note that Poincaré representations correspond to plane waves which are decomposable into an infinite number of spherical harmonics, irrespective of the size of the spatial momentum (related to the energy eigenvalue). That means that, for given spin, one is dealing with an infinite, continuous tower of modes, which is just what one obtains in the limit of vanishing energy increments for the generic spectrum shown in,

, Figure 1. In contradistinction, the singleton spectrum is different as the states have a single energy eigenvalue for any given value of the spin, as is obvious in Figure 3. Consequently, wave functions that constitute singleton representations do not depend on the radius of the anti-de Sitter spacetime and can be regarded as living on the boundary.

To obtain the spin-1 representation one can take the direct product of the spinless multiplet with a spin-1 state. Now the situation is more complicated, however, as the resulting multiplet contains states of spin lower than that of the ground state. In principle, each point in Figure 1 now generates three points, associated with two spherical harmonics, associated with rank-$j^{\pm}1$ tensors as well as mixed tensors of rank $j+1$ (so that $l = j$). An exception are the spinless points, which simply move to $j = 1$. The result of taking the product is depicted in Figure 4. This procedure can be

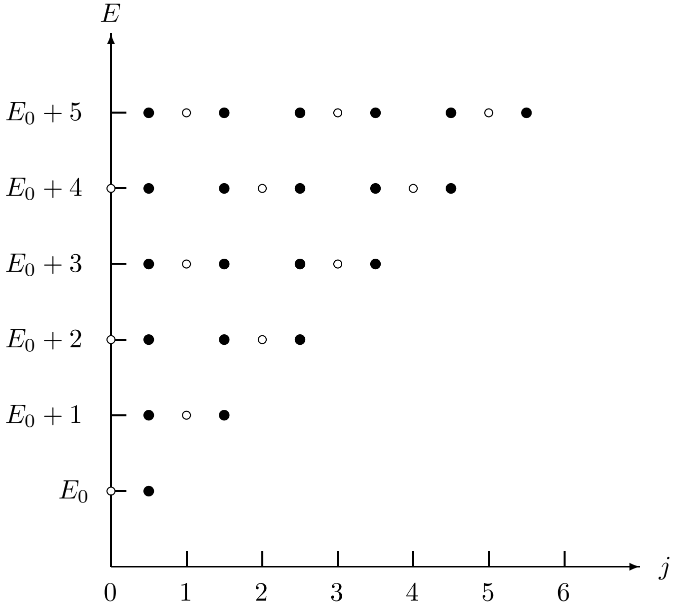


Fig. 2. States of the spinor representation in terms of the energy eigenvalues E and the angular momentum; the half-integer values for $j = l + \frac{1}{2}$ denote that we are dealing with a symmetric tensor-spinor of rank l. The small circles denote the original spinless multiplet from which the spinor multiplet has been constructed by a direct product with a spinor.

extended directly to ground states that transform as a spherical harmonic $Y^{a_1\cdots a_l}$.

Along the same lines as before, we investigate whether this representation can become reducible for special values of the ground state energy. We compare the value of the Casimir operator for the first excited states with minimal spin to the value for the ground state specified in (6.25). Hence we consider the states with $E = E_0 + 1$ and $j = l - 1$, assuming that the ground state has $l \geq 1$. In that case we find

$$\begin{aligned} \mathcal{C}_2 &= (E_0+1)(E_0-D+2)+(l-1)(l+D-4)-\Big|M_a^-|E_0+1,l-1\rangle\Big|^2 \\ &= E_0(E_0-D+1)+l(l+D-3)\,, \end{aligned} \tag{6.35}$$

so that

$$E_0 - l - D + 3 = \tfrac{1}{2}\Big|M_a^-|E_0+1,l-1\rangle\Big|^2 . \tag{6.36}$$

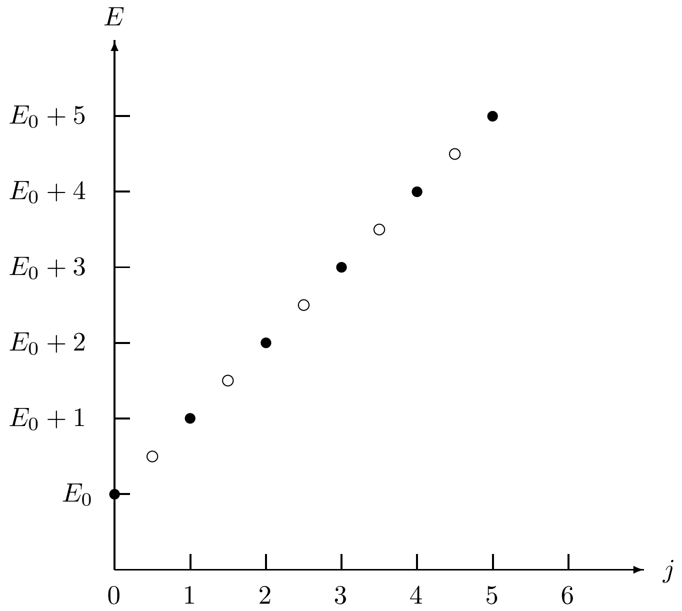


Fig. 3. The spin-0 and spin-$\frac{1}{2}$ singleton representations. The solid dots indicate the states of the spin-0 singleton, the circles the states of the spin-$\frac{1}{2}$ singleton. It is obvious that singletons contain much less degrees of freedom than a generic local field. The value of E_0, which denotes the spin-0 ground state energy, is equal to $E_0 = \frac{1}{2}(D-3)$. The spin-$\frac{1}{2}$ singleton ground state has an energy which is one half unit higher, as is explained in the text.

Therefore we establish the unitarity bound

$$E_0 \geq l + D - 3 \qquad (l \geq 1). \tag{6.37}$$

When $E_0 = l + D - 3$, however, the state $|E_0 + 1, l - 1\rangle$ is itself the ground state of an irreducible multiplet, which decouples from the original multiplet together with its corresponding excited states. This can be interpreted as the result of a gauge symmetry. Because these representations have a smooth Poincaré limit they are not singletons and can therefore be regarded as [illegible] representations. Hence massless representations with spin $l \geq 1$ are characterized by

$$E_0 = l + D - 3\,. \qquad (l \geq 1) \tag{6.38}$$

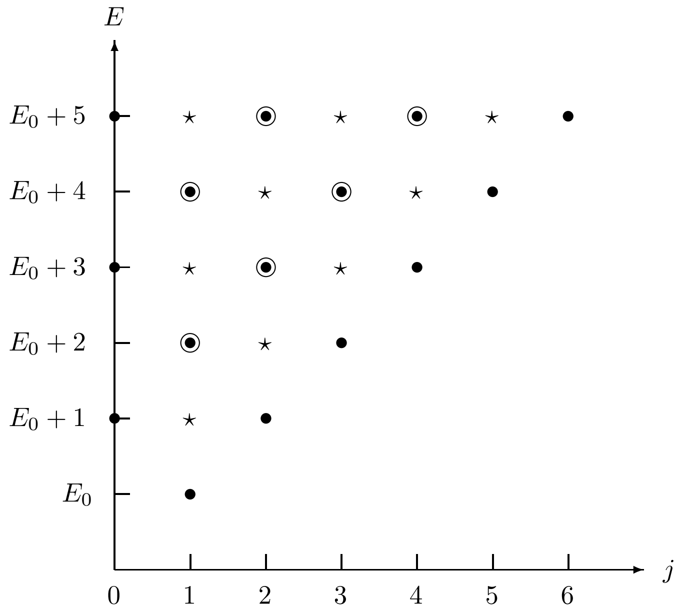


Fig. 4. States of the spin-1 representation in terms of the energy eigenvalues E and the angular momentum j. Observe that there are now points with double occupancy, indicated by the circle superimposed on the dots and states transforming as mixed tensors (with $l = j$) denoted by a $\star$. The double-occupancy points exhibit the structure of a spin-0 multiplet with ground state energy $E_0 + 1$. This multiplet becomes reducible and can be dropped when $E_0 = D - 2$, as is explained in the text. The remaining points then constitute a massless spin-1 multiplet, shown in Figure 5.

For these particular values the quadratic Casimir operator acquires a minimal value equal to

$$\mathcal{C}_2(\text{massless}) = 2(l-1)(l+D-3)\,. \qquad (l \geq 1) \tag{6.39}$$

We recall that this result is only derived for $l \geq 1$. For certain other cases, the identification of massless representation is somewhat ambiguous, as we already discussed. We return to this issue at the end of this section.

The above arguments can be easily extended to other grounds states, but this requires further knowledge of the various representations of the rotation group, at least for general dimension. This is outside the scope of these lectures. However, in $D = 4, 5$ dimensions this information is readily available. For a spin-s ground state in 4 spacetime dimensions we

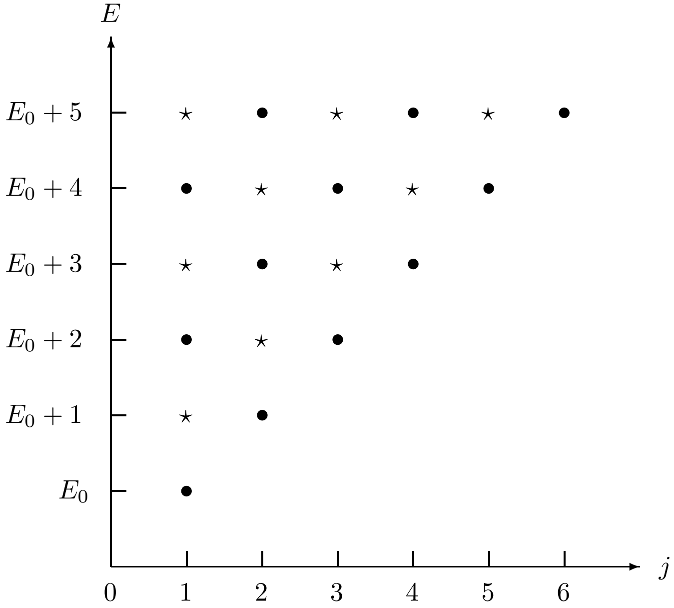


Fig. 5. States of the massless $s = 1$ representation in terms of the energy eigenvalues E and the angular momentum j. Now E_0 is no longer arbitrary but it is fixed to $E_0 = D - 2$.

immediately derive the unitarity bound (for $s > \frac{1}{2}$),

$$E_0 \geq s + 1\,, \tag{6.40}$$

by following the same procedure as leading to (6.37). When the bound is saturated we obtain a massless representation. The equation corresponding to (6.39) becomes

$$\mathcal{C}_2 = 2(s^2 - 1). \tag{6.41}$$

It turns out that this result applies to all spin-s representations, even to $s = 0, \frac{1}{2}$ conformal fields, for which we cannot use this derivation. This is a special property for $D = 4$ dimensions.

The case of $D = 5$ requires extra attention, because here the rotation group factorizes into two $SU(2)$ groups. We briefly summarize some results. First let us assume the the groundstate has spin (s_+, s_-) with $s_\pm \geq \frac{1}{2}$. In that case we find that the ground state energy satisfies the unitarity bound,

$$E_0 \geq s_+ + s_- + 2\,. \qquad (s_\pm \geq \tfrac{1}{2}) \tag{6.42}$$

This bound is saturated for massless states, for which the $E = E_0 + 1$ states with spin $(s_+ - \frac{1}{2}, s_- - \frac{1}{2})$ decouples. The corresponding value for the Casimir operator is equal to

$$\mathcal{C}_2 = (s_+ + s_-)^2 + 2s_+(s_+ + 1) + 2s_-(s_- + 1) - 4\,. \tag{6.43}$$

For $s_\pm = \frac{1}{2}l$ these values are in agreement with earlier result.

What remains to be considered are the ground states with spin $(0, s)$. Here we find

$$E_0 \geq 1 + s\,. \tag{6.44}$$

When the bound is saturated we have again a singleton representation. The corresponding values for the Casimir operator are

$$\mathcal{C}_2(\text{singleton}) = 3(s^2 - 1)\,. \tag{6.45}$$

The singleton representations for $s = 0, \frac{1}{2}$ were already found earlier. Note that for $D = 5$ there are thus infinitely many singleton representations, unlike in 4 dimensions, with a large variety of spin values. This is generically the case for arbitrary dimensions $D \neq 4$ and is thus related to the fact that the rotation group is of higher rank.

From the above it is clear that we are dealing with the phenomenon of multiplet shortening for specific values of the energy and spin of the representation, just as in the earlier discussions on BPS multiplets in previous sections. This phenomenon can be understood from the fact that the $[M_a^+, M_b^-]$ commutator acquires zero or negative eigenvalues for certain values of E_0 and J^2. When viewed in this way, the shortening of the representation is qualitatively similar to the shortening of BPS multiplets based on the anticommutator of the supercharges. Our discussion of the shortening of anti-de Sitter supermultiplets in Section 6.3 will support this point of view. The same phenomenon of multiplet shortening is well known and relevant in conformal field theory in $1 + 1$ dimensions.

The purpose of this section was to elucidate the various principles that underlie the anti-de Sitter representations and their relation with the field theory description. Here we are not striving for completeness. There are in fact alternative and often more systematic techniques for constructing the lowest-weight representations. A powerful method to construct the unitary irreducible representations of the anti-de Sitter algebra, is known as the oscillator method [88], which is applicable in any number of spacetime dimensions and which can also be used for supersymmetric extensions of the anti-de Sitter algebra. There is an extensive literature on this; for a recent elementary introduction to this method we refer to [80].

We close this section with a number of comments regarding “massless” representations and their field-theoretic description. As we demonstrated above, certain representations can, for a specific value of E_0, decouple into different irreducible representations. This phenomenon takes place when some unitarity bound is saturated. In that case one has representations that contain fewer degrees of freedom. When these "shortened" representations have a smooth Poincaré limit, they are called massless; when they do not, they are called singletons. For the case of spin-0 or spin-$\frac{1}{2}$ representations, for example, the spectrum of states is qualitatively independent of the value for E_0, as long as E_0 does not saturate the unitarity bound and a singleton representation decouples. Therefore the concept of mass remains ambiguous. We have already discussed this in Section 6.1, where we emphasized that the absence of mass terms in the field equations is also not a relevant criterion for masslessness. In Table 19 we have collected a number of examples of spin-0 and spin-$\frac{1}{2}$ representations with the criteria according to which they can be regarded as massless. One of them is tied to the fact that the corresponding field equation is conformally invariant, as we discussed at the end of Section 6.1. Another one follows from the fact that we are dealing with a gauge field. Here the example is an antisymmetric rank-$(D-2)$ gauge field, which is on-shell equivalent to a scalar.

We also invoke a criterion introduced by Günaydin (see [89] and the discussion in [90]), according to which every representation should be regarded as massless that appears in the tensor product of two singleton representations. For instance, it is easy to verify that the product of two spinless singletons leads to an infinite series of higher spin representations that are all massless according to (6.38). However, it also contains the $l=0$ representation with $E_0 = D-3$, to which (6.38) does not apply so that the interpretation as a massless representation is less obvious. It is interesting to consider this criterion for masslessness in $D=5$ dimensions. The tensor product of the singleton representations with spin $(0, s_-)$ and $(s_+, 0)$ leads to a ground state with spin (s_+, s_-) and $E_0 = 2 + s_+ + s_-$, which are obviously massless in view of (6.44) and (6.42). Taking the product of two singleton representations, one with spin $(s_1, 0)$ and another one with spin $(s_2, 0)$ leads to ground states with spin $(s, 0)$ and energy $E_0 = 2 + s + n$, where n is an arbitrary positive integer. Hence these representations should be regarded as massless.

This interpretation can be tested as follows. In maximal gauged supergravity in 5 dimensions with gauge group $SO(6)$, one of these representations appears as part of the "massless" supergravity multiplet. This anti-de Sitter representation is described by a (complex) tensor field, whose field equation takes the form,

$$\mathrm{e}^{-1}\varepsilon^{\mu\nu\rho\sigma\lambda}\, D_\rho B_{\sigma\lambda} + 2im\, B^{\mu\nu} = 0\,, \tag{6.46}$$

where $m = \pm g$. From this equation one can show that $B_{\mu\nu}$ satisfies (6.10) so that $\mathcal{C}_2 = 1$ ((6.13)). On shell the equation (6.46) projects out the

Table 19. Some unitary anti-de Sitter representations of spin 0 and $\frac{1}{2}$ which are massless according to various criteria, and the corresponding values for E_0 and $\mathcal{C}_2$.

spin	E_0	$\mathcal{C}_2$	type
0	$\frac{1}{2}D-1$	$-\frac{1}{4}D(D-2)$	conformal scalar
0	$\frac{1}{2}D$	$-\frac{1}{4}D(D-2)$	conformal scalar
0	$D-3$	$-2(D-3)$	$\in$ singleton $\times$ singleton
0	$D-1$	0	$(D-2)$-rank gauge field
$\frac{1}{2}$	$\frac{1}{2}D-\frac{1}{2}$	$-\frac{1}{8}D(D-1)$	conformal spinor
$\frac{1}{2}$	$D-\frac{5}{2}$	$\frac{1}{8}(D^2-15D+32)$	$\in$ singleton $\times$ singleton

degrees of freedom corresponding to spin $(1,0)$ or $(0,1)$, depending on the sign of m. From this one derives that $E_0=3$ (a second solution with $E_0=1$ violates the unitarity bound (6.44)).

6.3 *The superalgebras* $OSp(N|4)$

In this section we return to the anti-de Sitter superalgebras. We start from the (anti-)commutation relations already presented in (6.18) and (6.22). For definiteness we discuss the case of 4 spacetime dimensions with a Majorana supercharge Q. This allows us to make contact with the material discussed in Section 6.1. These anti-de Sitter supermultiplets were first discussed in [82, 84, 85, 91]. In most of the section we discuss simple supersymmetry (, $N=1$), but at the end we turn to more general N. We choose conventions where the 4×4 gamma matrices are given by

$$\Gamma^0=\begin{pmatrix} -i\mathbf{1} & 0 \\ 0 & i\mathbf{1}\end{pmatrix}\,,\qquad \Gamma^a=\begin{pmatrix} 0 & -i\sigma^a \\ i\sigma^a & 0\end{pmatrix}\,,\quad a=1,2,3\,, \tag{6.47}$$

and write the Majorana spinor Q in the form

$$Q=\begin{pmatrix} q_\alpha \\ \varepsilon_{\alpha\beta}\,q^\beta\end{pmatrix}\,, \tag{6.48}$$

where $q^\alpha\equiv q_\alpha^\dagger$, the indices $\alpha,\beta,\ldots$ are two-component spinor indices and the σ^a are the Pauli spin matrices. We substitute these definitions into (6.22) and obtain

$$[H\,,q_\alpha]\;\;=\;\;-\tfrac{1}{2}q_\alpha\,,$$

$$
\begin{aligned}
[H\,, q^\alpha] &= \tfrac{1}{2} q^\alpha\,, \\
\{q_\alpha\,, q^\beta\} &= (H\,\mathbf{1} + \vec{J}\cdot\vec{\sigma})_\alpha{}^\beta\,, \\
\{q_\alpha\,, q_\beta\} &= M_a^-\,(\sigma^a\sigma^2)_{\alpha\beta}\,, \\
\{q^\alpha\,, q^\beta\} &= M_a^+\,(\sigma^2\sigma^a)^{\alpha\beta}\,,
\end{aligned} \tag{6.49}
$$

where we have defined the (hermitean) angular momentum operators $J_a = -\frac{1}{2}\,i\,\varepsilon_{abc}M^{bc}$. We see that the operators q_α and q^α are lowering and raising operators, respectively. They change the energy of a state by half a unit. Observe that the relative sign between H and $\vec{J}\cdot\sigma$ in the third (anti)commutator is not arbitrary but fixed by the closure of the algebra.

In analogy to the bosonic case, we study unitary irreducible representations of the $OSp(1|4)$ superalgebra. We assume that there exists a lowest-weight state $|E_0, s\rangle$, characterized by the fact that it is annihilated by the lowering operators q_α,

$$
q_\alpha|E_0, s\rangle = 0\,. \tag{6.50}
$$

In principle we can now choose a ground state and build the whole representation upon it by applying products of raising operators q^α. However, we only have to study the [illegible] products of the q^α, because the symmetric ones just yield products of the operators M_a^+ by virtue of (6.49). Products of the M_a^+ simply lead to the higher-energy states in the anti-de Sitter representations of given spin that we considered in Section 6.2. By restricting ourselves to the antisymmetrized products of the q^α we thus restrict ourselves to the ground states upon which the full anti-de Sitter representations are build. These ground states are $|E_0, s\rangle$, $q^\alpha|E_0, s\rangle$ and $q^{[\alpha}q^{\beta]}|E_0, s\rangle$. Let us briefly discuss these representations for different s.

The $s = 0$ case is special since it contains less anti-de Sitter representations than the generic case. It includes the spinless states $|E_0, 0\rangle$ and $q^{[\alpha}q^{\beta]}|E_0, 0\rangle$ with ground-state energies E_0 and $E_0 + 1$, respectively. There is one spin-$\frac{1}{2}$ pair of ground states $q^\alpha|E_0, 0\rangle$, with energy $E_0 + \frac{1}{2}$. As we will see below, these states can be described by the scalar field A, the pseudoscalar field B and the spinor field ψ of the scalar chiral supermultiplet, that we studied in Section 6.1. Obviously, the bounds for E_0 that we derived in the previous sections should be respected, so that $E_0 > \frac{1}{2}$. For $E_0 = \frac{1}{2}$ the multiplet degenerates and decomposes into a super-singleton, consisting of a spin-0 and a spin-$\frac{1}{2}$ singleton, and another spinless supermultiplet with $E_0 = \frac{3}{2}$.

For $s \geq \frac{1}{2}$ we are in the generic situation. We obtain the ground states $|E_0, s\rangle$ and $q^{[\alpha}q^{\beta]}|E_0, s\rangle$ which have both spin s and which have energies E_0 and $E_0 + 1$, respectively. There are two more (degenerate) ground states, $q^\alpha|E_0, s\rangle$, both with energy $E_0 + \frac{1}{2}$, which decompose into the ground states with spin $j = s - \frac{1}{2}$ and $j = s + \frac{1}{2}$.

As in the purely bosonic case of Section 6.2, there can be situations in which states decouple so that we are dealing with multiplet shortening associated with gauge invariance in the corresponding field theory. The corresponding multiplets are then again called massless. We now discuss this in a general way analogous to the way in which one discusses BPS multiplets in flat space. Namely, we consider the matrix elements of the operator $q_\alpha\, q^\beta$ between the $(2s+1)$-degenerate ground states $|E_0, s\rangle$

$$\begin{aligned}\langle E_0, s|\, q_\alpha q^\beta\, |E_0, s\rangle &= \langle E_0, s|\{q_\alpha\,, q^\beta\}|E_0, s\rangle \\ &= \langle E_0, s|(E_0\,\mathbf{1} + \vec{J}\cdot\vec{\sigma})_\alpha{}^\beta|E_0, s\rangle\cdot \end{aligned}\tag{6.51}$$

This expression constitutes an hermitean matrix in both the quantum numbers of the degenerate groundstate and in the indices α and β, so that it is $(4s+2)$-by-$(4s+2)$. Because we assume that the representation is unitary, this matrix must be positive definite, as one can verify by inserting a complete set of intermediate states between the operators q_α and q^β in the matrix element on the left-hand side. Obviously, the right-hand side is manifestly hermitean as well, but in order to be positive definite the eigenvalue E_0 of H must be big enough to compensate for possible negative eigenvalues of $\vec{J}\cdot\vec{\sigma}$, where the latter is again regarded as a $(4s+2)$-by-$(4s+2)$ matrix. To determine its eigenvalues, we note that $\vec{J}\cdot\vec{\sigma}$ satisfies the following identity

$$(\vec{J}\cdot\vec{\sigma})^2 + (\vec{J}\cdot\vec{\sigma}) = s(s+1)\mathbf{1}\,, \tag{6.52}$$

as follows by straightforward calculation. This shows that $\vec{J}\cdot\vec{\sigma}$ has only two (degenerate) eigenvalues (assuming $s\neq 0$, so that the above equation is not trivially satisfied), namely s and $-(s+1)$. Hence in order for (6.51) to be positive definite, E_0 must satisfy the inequality

$$E_0 \geq s+1\,, \quad \text{for } s\geq \tfrac{1}{2}\cdot \tag{6.53}$$

If the bound is saturated, i.e. if $E_0 = s+1$, the expression on the right-hand side of (6.51) has zero eigenvalues so that there are zero-norm states in the multiplet which decouple. In that case we must be dealing with a massless multiplet. This is the bound (6.40), whose applicability is extended to spin-$\frac{1}{2}$. The ground state with $s=\frac{1}{2}$ and $E_0=\frac{3}{2}$ leads to the massless vector supermultiplet in 4 spacetime dimensions.

As we already mentioned one can also use the oscillator method to construct the irreducible representations. There is an extended literature on this. The reader may consult, for instance [87, 92].

Armed with these results we return to the masslike terms of Section 6.1 for the chiral supermultiplet. The ground-state energy for anti-de Sitter multiplets corresponding to the scalar field A, the pseudo-scalar field B and the Majorana spinor field ψ, are equal to E_0, E_0+1 and $E_0+\frac{1}{2}$, respectively. The Casimir operator therefore takes the values

$$\begin{aligned}\mathcal{C}_2(A) &= E_0(E_0-3)\,,\\ \mathcal{C}_2(B) &= (E_0+1)(E_0-2)\,,\\ \mathcal{C}_2(\psi) &= (E_0+\tfrac{1}{2})(E_0-\tfrac{5}{2})+\tfrac{3}{4}\cdot\end{aligned}\tag{6.54}$$

For massless anti-de Sitter multiplets, we know that the quadratic Casimir operator is given by (6.41), so we present the value for $\mathcal{C}_2-2(s^2-1)$ for the three multiplets,

$$\begin{aligned}\mathcal{C}_2(A)+2 &= (E_0-1)(E_0-2)\,,\\ \mathcal{C}_2(B)+2 &= E_0(E_0-1)\,,\\ \mathcal{C}_2(\psi)+\tfrac{3}{2} &= (E_0-1)^2\,.\end{aligned}\tag{6.55}$$

The terms on the right-hand side are not present for massless fields and we should therefore identify them somehow with the common mass parameter m of the supermultiplet. Comparison with the field equations (6.6) shows (for $g=1$) that we obtain the correct contributions provided we make the identification $E_0=m+1$. Observe that we could have made a slightly different identification here; the above result remains the same under the interchange of A and B combined with a change of sign in m (the latter is accompanied by a chiral redefinition of ψ).

When $E_0=2$ there exists, in principle, an alternative field representation for describing this supermultiplet. The spinless representation with $E_0=2$ can be described by a scalar field, the spin-$\frac{1}{2}$ representation with $E_0=\frac{5}{2}$ by a spinor field, and the second spinless representation with $E_0=3$ by a rank-2 field. The Lagrangian for the tensor supermultiplet is not conformally invariant in 4 dimensions, and this could account for the unusual ground state energy for the spinor representation. We have not constructed this supermultiplet in anti-de Sitter space; in view of the fact that it contains a tensor gauge field, it should be regarded as massless.

From Kaluza-Klein compactifications of supergravity one can deduce that there should also exist shortened supermultiplets. The reason is that the underlying supergravity multiplet in higher dimensions is shortened because it is massless. When compactifying to an anti-de Sitter ground state with supersymmetry the massless supermultiplets remain shortened by the same mechanism, but also the infinite tower of massive Kaluza-Klein states should comprise shortened supermultiplets. For toroidal compactifications the massive Kaluza-Klein states belong to BPS multiplets whose central charges are the momenta associated with the compactified dimensions. For nontrivial compactifications that correspond to supersymmetric anti-de Sitter ground states, the massive Kaluza-Klein states must be shortened according to the mechanism exhibited in this section. The singleton multiplets decouple from the Kaluza-Klein spectrum. Therefore it follows that there must exist shortened massive representations of the extended supersymmetric anti-de Sitter algebra.

To exhibit this we generalize the previous analysis to the N-extended superalgebra, denoted by $OSp(N,4)$. As it turns out, the analysis is rather similar. The supercharges now carry an extra $SO(N)$ index and are denoted by $q_{\alpha i}$ and $q^{\alpha i}$, with $q^{\alpha i} = (q_{\alpha i})^\dagger$ with $i = 1, \ldots, N$. The most relevant change to the (anti)commutators (6.49) is in the third one, which reads

$$\{q_{\alpha i}\,, q^{\beta j}\} = \delta_i^j\, \delta_\alpha^\beta\, H + \delta_i^j\, \vec{J}\cdot \vec{\sigma}_\alpha{}^\beta + \delta_\alpha^\beta\, \vec{T}\cdot \vec{\Sigma}_i{}^j\,, \tag{6.56}$$

where $\vec{T}$ are the hermitean $\frac{1}{2}N(N-1)$ generators of $SO(N)$ which act on the supercharges in the fundamental representation, generated by the hermitean matrices $\vec{\Sigma}$. The last two anticommutators are given by

$$\begin{aligned} \{q_{\alpha i}\,, q_{\beta j}\} &= M_a^-\, (\sigma^a \sigma^2)_{\alpha\beta}\, \delta_{ij}\,, \\ \{q^{\alpha i}\,, q^{\beta j}\} &= M_a^+\, (\sigma^2 \sigma^a)^{\alpha\beta}\, \delta^{ij}\,. \end{aligned} \tag{6.57}$$

The construction of lowest-weight representations proceeds in the same way as before. One starts with a ground state of energy E_0 which has a certain spin and transforms according to a representation of $SO(N)$ which is annihilated by the $q_{\alpha i}$. Denoting the $SO(N)$ representation by t (which can be expressed in terms of the eigenvalues of the Casimir operators or Dynkin labels), we have

$$q_{\alpha i}|E_0, s, t\rangle = 0\,. \tag{6.58}$$

Excited states are generated by application of the $q^{\alpha i}$, which are mutually anticommuting, with exception of the combination that leads to the operators M_a^+ which will generate the full anti-de Sitter representations. Hence the generic N-extended representations decompose into ordinary anti-de Sitter representation whose ground states have energy $E_0 + \frac{1}{2}n$ and which can be written as

$$q^{[\alpha_1 i_1}\, \cdots\, q^{\alpha_n i_n]}\, |E_0, s, t\rangle\cdot \tag{6.59}$$

Here the antisymmetrization applies to the combined (αi) labels. As before the unitarity limits follow from the separate limits on the anti-de Sitter representations and from the right-hand side of the anticommutator (6.56), which decomposes into three terms, namely the Hamiltonian, the rotation generators and the R-symmetry generators, taken in the space of ground state configurations ((6.51)). We have already determined the possible eigenvalues of $\vec{J}\cdot\vec{\sigma}$ which are equal to s or $-(s+1)$. In a similar way one can determine the eigenvalues for $\vec{T}\cdot\vec{\Sigma}$ by noting that it satisfies a polynomial matrix equation such as (6.52) with coefficients determined by the Casimir operators. For instance, for $N = 3$ we derive,

$$-(\vec{T}\cdot\vec{\Sigma})^3 + 2(\vec{T}\cdot\vec{\Sigma})^2 + (t^2+t-1)(\vec{T}\cdot\vec{\Sigma}) = t(t+1)\,\mathbf{1}\,, \tag{6.60}$$

where $\vec{T}^2 = t(t+1)\,\mathbf{1}$. This equation shows that the eigenvalues of $\vec{T}\cdot\vec{\Sigma}$ take the values $-t$, 1 or $t+1$, unless $t = 0$. Combining these results we find that the right-hand side of (6.56) in the space of degenerate ground state configurations has the following six eigenvalues: E_0+s-t, E_0+s+1, $E_0+s+t+1$, $E_0-s-t-1$, E_0-s or E_0+t+1. All these eigenvalues must be positive, so that in the generic case where s and t are nonvanishing, we derive the unitarity bound, $E_0 \geq 1+s+t$. Incorporating also the possibility that s or t vanishes, the combined result takes the following form,

$$\begin{array}{lll} E_0 \geq 1+s+t & \text{for} & s\geq \frac{1}{2}\,,\ t\geq\frac{1}{2}\,,\\ E_0 \geq 1+s & \text{for} & s\geq\frac{1}{2}\,,\ t=0\,,\\ E_0\geq t & \text{for} & s=0\,,\ t\geq\frac{1}{2}\,. \end{array} \tag{6.61}$$

Whenever one of these bounds is saturated, certain anti-de Sitter representations must decouple. The ground states with $s = 0$ and $E_0 = t$ define massive shortened representations of the type that appear in Kaluza-Klein compactifications [84]. In the Poincaré limit these representations become all massless.

Obviously these techniques can be extended to other cases, either by changing the number of supersymmetries or by changing the spacetime dimension. There is an extended literature to which we refer the reader for applications and further details.

Before closing the section we want to return to the remarkable singleton representations. Long before the formulation of the AdS/CFT correspondence it was realized that supersingleton representations could be described by conformal supersymmetric field theories on a boundary. Two prominent examples were noted (see, [87, 92]), namely the singleton representations in $D = 5$ and 7 anti-de Sitter space, which correspond to $N = 4, D = 4$ supersymmetric gauge theories and the chiral $(2,0)$ tensor multiplet in $D = 6$ dimensions. The singletons decouple from the Kaluza-Klein spectrum, precisely because they are related to boundary degrees of freedom. Group-theoretically they are of interest because their products lead to the massless and massive representations that one encountered in the Kaluza-Klein context. Another theme addresses the connection between singletons and higher-spin theories. Here the issue is whether the singletons play only a group-theoretic role or whether they have also a more dynamical significance. We refrain from speculating about these questions and just refer to some recent papers [93–95]. In [94] the reader may also find a summary of some useful results about singletons as well as an extensive list of references.

In the next section we will move to a discussion of superconformal symmetries, which are based on the same anti-de Sitter algebra. We draw the attention of the reader to the fact that in Section 7, D will always denote the spacetime dimension of the superconformal theory. The corresponding superalgebra is then the anti-de Sitter superalgebra, but in spacetime dimension $D+1$.

7 Superconformal symmetry

Invariances of the metric are known as isometries. Continuous isometries are generated by so-called Killing vectors, satisfying

$$D_\mu \xi_\nu + D_\nu \xi_\mu = 0\,. \tag{7.1}$$

The maximal number of linearly independent Killing vectors is equal to $\frac{1}{2}D(D+1)$. A space that has the maximal number of isometries is called maximally symmetric. A weaker condition than (7.1) is,

$$D_\mu \xi_\nu + D_\nu \xi_\mu = \frac{2}{D}\, g_{\mu\nu} D_\rho \xi^\rho\,. \tag{7.2}$$

Solutions to this equation are called [illegible] Killing vectors. Note that the above equation is the traceless part of (7.1). The conformal Killing vectors that are not isometries are thus characterized by a nonvanishing $\xi = D_\mu \xi^\mu$. For general dimension $D > 2$ there are at most $\frac{1}{2}(D+1)(D+2)$ conformal Killing vectors. For $D = 2$ there can be infinitely many conformal Killing vectors. These result can be derived as follows. First one shows that

$$D_\mu D_\nu \xi_\rho = R_{\nu\rho\mu}{}^\sigma\, \xi_\sigma - \frac{1}{D}\Big[g_{\mu\nu}\, D_\rho \xi - g_{\rho\mu}\, D_\nu \xi - g_{\rho\nu}\, D_\mu \xi \Big]\,. \tag{7.3}$$

For Killing vectors (which satisfy $\xi = 0$) this result implies that the second derivatives of Killing vectors are determined by the vector and its first derivatives. When expanding about a certain point on the manifold, the Killing vector is thus fully determined by its value at that point and the values of its first derivatives (which are antisymmetric in view of (7.1)). Altogether there are thus $\frac{1}{2}D(D+1)$ initial conditions to be fixed and they parametrize the number of independent Killing vectors. For conformal Killing vectors, where $\xi \neq 0$ one then proves that $(D-2)D_\mu D_\nu \xi$ and $D^\mu D_\mu \xi$ are determined in terms of lower derivatives. This suffices to derive the maximal number of conformal Killing vectors quoted above for $D > 2$. Both ordinary and conformal Killing vectors generate a group.

In what follows we choose a Minkowski signature for the D-dimensional space, a restriction that is mainly relevant when considering supersymmetry. Flat Minkowski spacetime has the maximal number of conformal Killing vectors, which decompose as follows,

$$\xi^\mu = \begin{cases} \xi_P^\mu & \text{spacetime translations } (P) \\ \epsilon^\mu{}_\nu \, x^\nu & \text{Lorentz transformations } (M) \\ \Lambda_D \, x^\mu & \text{scale transformations } (D) \\ (2\, x^\mu x^\nu - x^2 \, \eta^{\mu\nu})\Lambda_{K\nu} & \text{conformal boosts } (K). \end{cases} \tag{7.4}$$

Here ξ_P^μ, $\epsilon^{\mu\nu} = -\epsilon^{\nu\mu}$, Λ_D and Λ_K^μ are constant parameters. Obviously $\xi = D(\Lambda_D + x_\mu \Lambda_K^\mu)$. The above conformal Killing vectors generate the group $SO(D,2)$. This is the same group as the anti-de Sitter group in $D+1$ dimensions. The case of $D=2$ is special because in that case the above transformations generate a semisimple group, $SO(2,2) \cong (SL(2,R)\times SL(2,R))/Z_2$. This follows directly by writing out the infinitesimal transformations (7.4) for the linear combinations $x \pm t$,

$$\delta(x\pm t) = (\xi_P^x \pm \xi_P^t) + (\Lambda_D \mp \epsilon^{xt})(x\pm t) + \tfrac{1}{2}(\Lambda_K^x \mp \Lambda_K^t)(x\pm t)^2 \,. \tag{7.5}$$

However, for $D=2$ there are infinitely many conformal Killing vectors, corresponding to two copies of the Virasoro algebra. The corresponding diffeomorphisms can be characterized in terms of two independent functions $f_\pm$ and take the form,

$$\delta x = f_+(x+t) + f_-(x-t)\,, \qquad \delta t = f_+(x+t) - f_-(x-t)\,. \tag{7.6}$$

The fact that, for $D \geq 3$ the anti-de Sitter and the conformal group coincide for dimensions $D+1$ and D, respectively, can be clarified by extending the D-dimensional spacetime parametrized by coordinates x^μ with an extra (noncompact) coordinate y, assuming the line element,

$$\mathrm{d}s^2 = \frac{g_{\mu\nu}\, \mathrm{d}x^\mu \mathrm{d}x^\nu + \mathrm{d}y^2}{y^2}\,, \tag{7.7}$$

so that the right-hand side of (7.2), which is responsible for the lack of invariance of the line element of the original D-dimensional space, can be cancelled by a scale transformation of extra coordinate y. It is straightforward to derive the nonvanishing Christoffel symbols for this extended space

$$\{{}_\mu{}^y{}_\nu\} = y^{-1}\, g_{\mu\nu}\,, \quad \{{}_\mu{}^\nu{}_y\} = -y^{-1}\,\delta_\mu^\nu\,, \quad \{{}_y{}^y{}_y\} = -y^{-1}\,, \tag{7.8}$$

where $\{{}_\mu{}^\rho{}_\nu\}$ remains the same for both spaces and all other components vanish. The corresponding expressions for the curvature components are

$$R_{\mu\nu\rho}{}^\sigma \;=\; R^D_{\mu\nu\rho}{}^\sigma + 2y^{-2}\, g_{\rho[\mu}\delta^\sigma_{\nu]}\,,$$

$$\begin{aligned} R_{\mu y \rho}{}^{y} &= y^{-2}\, g_{\mu\rho}\,, \\ R_{y\nu y}{}^{\sigma} &= y^{-2}\delta_\nu^\sigma\,. \end{aligned} \tag{7.9}$$

With these results one easily verifies that the curvature tensor of the $(D+1)$-dimensional extension of a flat D-dimensional Minkowski space is that of an anti-de Sitter spacetime with unit anti-de Sitter radius ($g = 1$ in (3.14)). This was the reason why we adopted a positive signature in the line element (7.7) for the coordinate y.

Subsequently one can show that the D-dimensional conformal Killing vectors satisfying $D_\mu D_\nu \xi = 0$ can be extended to Killing vectors of the $(D+1)$-dimensional space,

$$\xi^\mu(x,y) = \xi^\mu(x) - \frac{y^2}{2D}\,\partial^\mu \xi(x)\,, \qquad \xi^y(x,y) = \frac{y}{D}\,\xi(x)\,. \tag{7.10}$$

The condition $D_\mu D_\nu \xi = 0$ holds for the conformal Killing vectors (7.4). For $D = 2$ these vectors generate a finite subgroup of the infinite-dimensional conformal group, and only this group can be extended to isometries of the $(D+1)$-dimensional space. Nevertheless, near the boundary [77] of the space ($y \approx 0$), the conformal Killing vectors generate asymptotic symmetries. Such a phenomenon was first analyzed in [96].

This setting is relevant for the adS/CFT correspondence and there exists an extensive literature on this (see, [28, 97–101], and also the lectures presented at this school). Also the relation between the D'Alembertians of the extended and of the original D-dimensional spacetime is relevant in this context. Straightforward calculation yields,

$$\Box_{D+1} = y^2\,\Box^D + (y\,\partial_y)^2 - D\,y\,\partial_y\,. \tag{7.11}$$

Near the boundary where y is small, the fields can be approximated by $y^\Delta\,\phi(x)$. We may compare this to solutions of the Klein-Gordon equation in the anti-de Sitter space, for which we know that the D'Alembertian equals the quadratic Casimir operator $\mathcal{C}_2$. In terms of the ground state energy E_0 of the anti-de Sitter representation, we have $\mathcal{C}_2 = E_0(E_0 - D)$ (observe that we must replace D by $D+1$ in (6.25)), which shows that we have the identification $\Delta = E_0$ or $\Delta = D - E_0$. This identification is somewhat remarkable in view of the fact that E_0 is the energy eigenvalue associated with the $SO(2)$ generator of the anti-de Sitter algebra and not with the noncompact scale transformation of y, which associated with the $SO(1,1)$ eigenvalue. This identification of the generators is discussed in more detail in the next section.

7.1 The superconformal algebra

From the relation between the conformal and the anti-de Sitter algebra one can determine the superextension of the conformal algebra generated by the above conformal Killing vectors. In comparison to the anti-de Sitter algebra and superalgebra ((6.4) and (6.5)) we make a different decomposition than the one that led to (6.18) and (6.22). We start from a D-dimensional spacetime of coordinates carrying indices $a = 0, 1, \ldots, D-1$, which we extend with extra index values, so that $A = -, 0, 1, \ldots, D-1, D$. For the bosonic generators which generate the group $SO(D,2)$ we have

$$\begin{array}{rcl} M_{D-} & \longrightarrow & D\,, \\ M_{ab} & \longrightarrow & M_{ab}\,, \\ M_{Da} & \longrightarrow & \tfrac{1}{2}(P_a - K_a)\,, \\ M_{-a} & \longrightarrow & \tfrac{1}{2}(P_a + K_a). \end{array} \tag{7.12}$$

Here we distinguish the generator D of the dilatations, $\frac{1}{2}D(D-1)$ generators M_{ab} of the Lorentz transformations, D generators P_a of the translations, and D generators K_a of the conformal boosts.

The algebra associated with $SO(D,2)$ was given in (6.4) and corresponds to the following commutation relations

$$\begin{array}{rclrcl} [D, P_a] &=& -P_a\,, & [D, K_a] &=& K_a\,, \\ [M_{ab}, P_c] &=& -2\,\eta_{c[a}\,P_{b]}\,, & [M_{ab}, K_c] &=& -2\,\eta_{c[a}\,K_{b]}\,, \\ [M_{ab}, M_{cd}] &=& 4\,\eta_{[a[c}\,M_{d]b]}\,, & [P_a, P_b] &=& [K_a, K_b] = 0\,, \\ [D, M_{ab}] &=& 0\,, & [K_a, P_b] &=& 2(M_{ab} + \eta_{ab}\,D)\,. \end{array} \tag{7.13}$$

To obtain the superextension (for $D \leq 6$) one must first extend the spinor representation associated with the D-dimensional spacetime to incorporate two extra gamma matrices Γ_D and Γ_-. According to the discussion in Section 2.5 (see, in particular, Table 9) this requires a doubling of the spinor charges,

$$Q \to \mathcal{Q} = \begin{pmatrix} S_\alpha \\ Q_\alpha \end{pmatrix}\,, \qquad \bar{Q} \to \bar{\mathcal{Q}} = (\bar{Q}_\alpha, \bar{S}_\alpha)\,, \tag{7.14}$$

and we define an extended set of gamma matrices Γ_A by,

$$\Gamma_a = \begin{pmatrix} \Gamma^a & 0 \\ 0 & -\Gamma^a \end{pmatrix} \qquad \Gamma_D = \begin{pmatrix} 0 & \mathbf{1} \\ \mathbf{1} & 0 \end{pmatrix} \qquad \Gamma_- = \begin{pmatrix} 0 & \mathbf{1} \\ -\mathbf{1} & 0 \end{pmatrix}\,. \tag{7.15}$$

The new charges S_α generate so-called supersymmetry transformations [27]. The decomposition of the conjugate spinor is somewhat subtle, to make contact with the Majorana condition employed for the anti-de Sitter algebra.

The anticommutation relation for the spinor charges follows from (6.5) and can be written as

$$
\begin{aligned}
\{\mathcal{Q}, \bar{\mathcal{Q}}\} &= \begin{pmatrix} \{S, \bar{Q}\} & \{S, \bar{S}\} \\ \{Q, \bar{Q}\} & \{Q, \bar{S}\} \end{pmatrix} \\
&= \begin{pmatrix} -\frac{1}{2}\Gamma^{ab} M_{ab} - D & -\Gamma^a K_a \\ -\Gamma^a P_a & -\frac{1}{2}\Gamma_{ab} M_{ab} + D \end{pmatrix}, \qquad (7.16)
\end{aligned}
$$

or

$$
\begin{aligned}
\{Q_\alpha, \bar{Q}_\beta\} &= -\Gamma^a_{\alpha\beta}\, P_a\,, \\
\{S_\alpha, \bar{S}_\beta\} &= -\Gamma^a_{\alpha\beta}\, K_a\,, \\
\{Q_\alpha, \bar{S}_\beta\} &= -\tfrac{1}{2}\Gamma^{ab}_{\alpha\beta}\, M_{ab} + \eta_{\alpha\beta}\, D. \qquad (7.17)
\end{aligned}
$$

The nonvanishing commutators of the spinor charges with the bosonic generators read

$$
\begin{aligned}
[M_{ab}, \bar{Q}_\alpha] &= \tfrac{1}{2}(\bar{Q}\Gamma_{ab})_\alpha\,, & [M_{ab}, \bar{Q}_\alpha] &= \tfrac{1}{2}(\bar{Q}\Gamma_{ab})_\alpha\,, \\
[D, \bar{Q}_\alpha] &= -\tfrac{1}{2}\bar{Q}_a\,, & [D, \bar{S}_\alpha] &= \tfrac{1}{2}\bar{S}_\alpha\,, \qquad (7.18) \\
[K_a, \bar{Q}_\alpha] &= -(\bar{S}\Gamma_a)_\alpha\,, & [P_a, \bar{S}_\alpha] &= -(\bar{Q}\Gamma_a)_\alpha\,.
\end{aligned}
$$

Here we are assuming the same gamma matrix conventions as in the beginning of Section 3. From the results quoted in the previous section, we know that, up to $D = 6$, the bosonic subalgebra will be the sum of the conformal algebra and the R-symmetry algebra. The R-symmetry can be identified from Table 9 and the corresponding generators will appear on the righ-hand side of the $\{Q, S\}$ anticommutator; the other (anti)commutation relations listed above remain unchanged. In addition, commutators with the R-symmetry generators must be specified, but those follow from the R-symmetry assignments of the supercharges. The above (anti)commutators satisfy the Jacobi identities that are at most quadratic in the fermionic generators. The validity of the remaining Jacobi identities, which are cubic in the fermionic generators, requires in general the presence of the R-symmetry charges. The results given so far suffice to discuss the most salient features of the superconformal algebra and henceforth we will be ignoring the contributions of the R-symmetry generators. Note also that the numbers of bosonic and fermion generators do not match; this mismatch will in general remain when including the R-symmetry generators.

As before, the matrix on the right-hand side of (7.16) may have zero eigenvalues, leading to shortened supermultiplets. Those multiplets are in one-to-one correspondence with the anti-de Sitter supermultiplets. Its eigenvalues are subject to certain positivity requirements in order that the algebra is realized in a positive-definite Hilbert space.

The abstract algebra can be connected to the spacetime transformations (7.4) in flat spacetime introduced at the beginning of this section. To see this we derive how the conformal transformations act on generic fields. In principle, this is an application of the theory of homogeneous spaces discussed in Section 4 and we will demonstrate this for the bosonic transformations [102]; a supersymmetric extension can be given in superspace. Let us assume that the action of these spacetime transformations denoted by g takes the following form on a generic multicomponent field ϕ,

$$\phi(x)\longrightarrow\phi_g(x) = S(g,x)\,\phi(g^{-1}x)\,, \tag{7.19}$$

where S is some matrix acting on the components of ϕ. Observe that there exists a subgroup of the conformal group that leaves a point in spacetime invariant and choose, by a suitable translation, this point equal to $x^a = 0$. From (7.4) it then follows that the corresponding stability group of this point is generated by the generators M of the Lorentz group, the generator D of the scale transformations and the generators K of the conformal boosts. Hence we conclude that the matrices $S(g,0)$ must form a representation of this subgroup, whose generators are denoted by the matrices $\hat{M}_{ab}$, $\hat{D}$ and $\hat{K}_a$. Generic fields are thus assigned to representations of this subgroup.

On the other hand, we want the translation operators to act exclusively on the coordinates x^a, so that (the generators have been taken antihermitean),

$$P_a\,\phi(x) = \frac{\partial}{\partial x^a}\,\phi(x)\,, \tag{7.20}$$

so that $\phi(x)$ can be written as $\exp(x^a P_a)\phi(0)$. Combining (7.19) with (7.20), it follows that

$$S(g,x) = \exp(-y^a P_a)\,S(g,0)\,\exp(x^b P_b)\,, \quad \text{with } y = g\,x\,. \tag{7.21}$$

Writing this out for infinitesimal transformations with $y^a = x^a + \xi^a$, and $O(g,0) \approx \mathbf{1} + \frac{1}{2}\epsilon^{ab}\hat{M}_{ab} + \Lambda_D\hat{D} + \Lambda_K^a\hat{K}$, we derive that conformal transformations act infinitesimally on ϕ according to

$$\begin{aligned}\delta\phi(x) \;=\;& -\xi^a\partial_a\phi(x) \\ &+\Big[\Big(\tfrac{1}{2}\epsilon^{ab} + 2\Lambda_K^{[a}x^{b]}\Big)\hat{M}_{ab} + \Big(\Lambda_D - 2\Lambda_K^a x_a\Big)\hat{D} + \Lambda_K^a\hat{K}_a\Big]\phi(x)\,,\end{aligned} \tag{7.22}$$

where ξ^a denotes the conformal Killing vectors parametrized in (7.4).

The procedure applied above is just a simple example of the construction of induced representations on a G/H coset manifold. Indeed, we are describing flat space as a coset manifold, where the conformal group plays the role of the isometry group G and the stability group plays the role of the isotropy group H. The coset representative equals $\exp(x^a P_a)$, from which it follows ((4.16)) that the vielbein is constant and diagonal and the connections associated with the stability group are zero. Hence the metric is invariant under the conformal transformations, as established earlier, while the vielbein is invariant after including the compensating transformations represented by the second line of (7.22). Explicit evaluation then shows that the invariance of the flat vielbein requires the compensating tangent-space transformations,

$$\delta e_\mu{}^a = \epsilon^{ab}\, e_{\mu,b} - \Lambda_D\, e_\mu{}^a\,, \tag{7.23}$$

with parameters specified by (7.22). Note that the special conformal boosts do not act on the tangent space index of the vielbein.

In the next two sections we will discuss how one can deviate from flat space in the context of the conformal group. There are two approaches here which lead to related results. One is to start from a gauge theory of the conformal group. This conformal group has nothing to do with spacetime transformations and the resulting theory is described in some unspecified spacetime. Then one imposes a constraint on certain curvatures. This is similar to what we described in Section 3, where we imposed a constraint on the torsion tensor ((3.6)), so that the spin connection becomes a dependent field and the Riemann tensor becomes proportional to the curvature of the spin connection field. This approach amounts to imposing the maximal number of conventional constraints. The second approach starts from the coupling to superconformal matter and the corresponding superconformal currents.

7.2 Superconformal gauge theory and supergravity

In principle it is straightforward to set up a gauge theory associated with the superconformal algebra. We start by associating a gauge field to every generator,

$$\begin{array}{lcccccc} \text{generators:} & P & M & D & K & Q & S \\ \text{gauge fields:} & e_\mu^{\ a} & \omega_\mu^{ab} & b_\mu & f_\mu^{\ a} & \psi_\mu & \phi_\mu \\ \text{parameters:} & \xi_P^a & \epsilon^{ab} & \Lambda_D & \Lambda_K^a & \epsilon & \eta. \end{array} \tag{7.24}$$

Up to normalization factors, the transformation rules for the gauge fields, which we specify below, follow directly from the structure constants of the superconformal algebra,

$$\begin{aligned} \delta e_\mu{}^a &= \mathcal{D}_\mu \xi_P^a - \Lambda_D e_\mu^{\ a} + \tfrac{1}{2}\bar\epsilon\Gamma^a\psi_\mu\,, \\ \delta\omega_\mu^{ab} &= \mathcal{D}_\mu\epsilon^{ab} + \Lambda_K^{[a} e_\mu^{\ b]} - \xi_P^{[a} f_\mu^{\ b]} - \tfrac{1}{4}\bar\epsilon\Gamma^{ab}\phi_\mu + \tfrac{1}{4}\bar\psi_\mu\Gamma^{ab}\eta\,, \end{aligned}$$

$$
\begin{aligned}
\delta b_\mu &= \mathcal{D}_\mu\Lambda_D + \tfrac{1}{2}\Lambda_{Ka}e_\mu{}^a - \tfrac{1}{2}\xi_{Pa}f_\mu{}^a + \tfrac{1}{4}\bar\epsilon\phi_\mu - \tfrac{1}{4}\bar\psi_\mu\eta\,,\\
\delta f_\mu{}^a &= \mathcal{D}_\mu\Lambda_K^a + \Lambda_D e_\mu{}^a + \tfrac{1}{2}\bar\eta\Gamma^a\phi_\mu\,,\\
\delta\psi_\mu &= \mathcal{D}_\mu\epsilon - \tfrac{1}{2}\Lambda_D\psi_\mu - \tfrac{1}{2}e_\mu{}^a\Gamma_a\eta + \tfrac{1}{2}\xi_P^a\Gamma_a\phi_\mu\,,\\
\delta\phi_\mu &= \mathcal{D}_\mu\eta + \tfrac{1}{2}\Lambda_D\phi_\mu - \tfrac{1}{2}f_\mu{}^a\Gamma_a\epsilon + \tfrac{1}{2}\Lambda_K^a\Gamma_a\psi_\mu\,.
\end{aligned}
\qquad (7.25)
$$

Here we use derivatives that are covariantized with respect to dilatations and Lorentz transformations, ,

$$
\begin{aligned}
\mathcal{D}_\mu\xi_P^a &= \partial_\mu\xi_P^a + b_\mu\xi_P^a - \omega_\mu{}^{ab}\xi_{Pb}\,,\\
\mathcal{D}_\mu\Lambda_K^a &= \partial_\mu\Lambda_K^a - b_\mu\Lambda_K^a - \omega_\mu{}^{ab}\Lambda_{Kb}\,,\\
\mathcal{D}_\mu\Lambda_D &= \partial_\mu\Lambda_D\,,\\
\mathcal{D}_\mu\epsilon &= (\partial_\mu + \tfrac{1}{2}b_\mu - \tfrac{1}{4}\omega_\mu^{ab}\Gamma_{ab})\epsilon\,,\\
\mathcal{D}_\mu\eta &= (\partial_\mu - \tfrac{1}{2}b_\mu - \tfrac{1}{4}\omega_\mu^{ab}\Gamma_{ab})\eta\,.
\end{aligned}
\qquad (7.26)
$$

Again we suppressed the gauge fields for the R-symmetry generators.

The above transformation rules close under commutation, up to the commutators of two supersymmetry transformations acting on the fermionic gauge fields. In that case, one needs Fierz reorderings to establish the closure of the algebra, which depend sensitively on the dimension and on the presence of additional generators (for $D=4$, see, for example [27]). As an example we list some of the commutation relations that can be obtained from (7.25)

$$
\begin{aligned}
[\delta_P(\xi_P),\delta_K(\Lambda_K)] &= \delta_D(\tfrac{1}{2}\Lambda_K^a\xi_P^b\,\eta_{ab}) + \delta_M(\Lambda_K^{[a}\xi_P^{b]})\,,\\
\{\delta_Q(\epsilon_1),\delta_Q(\epsilon_2)\} &= \delta_P(\tfrac{1}{2}\bar\epsilon_2\Gamma^a\epsilon_1)\,,\\
\{\delta_S(\eta_1),\delta_S(\eta_2)\} &= \delta_K(\tfrac{1}{2}\bar\eta_2\Gamma^a\eta_1)\,,\\
\{\delta_Q(\epsilon),\delta_S(\eta)\} &= \delta_M(\tfrac{1}{4}\bar\epsilon\Gamma^{ab}\eta) + \delta_D(-\tfrac{1}{4}\bar\epsilon\eta)\,,\\
[\delta_Q(\epsilon),\delta_K(\Lambda_K)] &= \delta_S(\tfrac{1}{2}\Lambda_K^a\Gamma_a\epsilon)\,,\\
[\delta_S(\eta),\delta_P(\Lambda_P)] &= \delta_Q(\tfrac{1}{2}\xi_P^a\Gamma_a\eta)\,.
\end{aligned}
\qquad (7.27)
$$

For completeness we also present the corresponding curvature tensors of the superconformal gauge theory,

$$
\begin{aligned}
R_{\mu\nu}^a(P) &= 2\,\mathcal{D}_{[\mu}e_{\nu]}^a - \tfrac{1}{2}\bar\psi_{[\mu}\Gamma^a\psi_{\nu]}\,,\\
R_{\mu\nu}^{ab}(M) &= 2\,\partial_{[\mu}\,\omega_{\nu]}^{ab} - 2\,\omega_{[\mu}^{ac}\,\omega_{\nu]\,c}{}^b - 2\,f_{[\mu}^{[a}\,e_{\nu]}^{b]} + \tfrac{1}{2}\bar\psi_{[\mu}\Gamma^{ab}\phi_{\nu]}\,,\\
R_{\mu\nu}(D) &= 2\,\mathcal{D}_{[\mu}b_{\nu]} - f_{[\mu}^a\,e_{\nu]\,a} - \tfrac{1}{2}\bar\psi_{[\mu}\phi_{\nu]}\,,\\
R_{\mu\nu}^a(K) &= 2\,\mathcal{D}_{[\mu}f_{\nu]}^a - \tfrac{1}{2}\bar\phi_{[\mu}\Gamma^a\phi_{\nu]}\,,
\end{aligned}
$$

$$
\begin{aligned}
R_{\mu\nu}(Q) &= 2\,\mathcal{D}_{[\mu}\psi_{\nu]} - e_{[\mu}{}^{a}\,\Gamma_a\phi_{\nu]}\,,\\
R_{\mu\nu}(S) &= 2\,\mathcal{D}_{[\mu}\phi_{\nu]} - f_{[\mu}{}^{a}\,\Gamma_a\psi_{\nu]}\,.
\end{aligned}
\tag{7.28}
$$

These curvature tensors transform covariantly and their transformation rules follow from the structure constants of the superconformal algebra. They also satisfy a number of Bianchi identities which are straightforward to write down. As an example and for future reference we list the first three identities,

$$
\begin{aligned}
\mathcal{D}_{[\mu}R^{a}_{\rho]}(P) + R^{ab}_{[\mu\nu}(M)\,e_{\rho]\,b} - R_{[\mu\nu}(D)\,e_{\rho]}^{\;a} - \tfrac{1}{2}\bar\psi_{[\rho}\Gamma^a R_{\mu\nu]}(Q) &= 0\,,\\
\mathcal{D}_{[\mu}R^{ab}_{\nu\rho]}(M) + R^{[a}_{[\mu\nu}(K)\,e_{\rho]}^{\;b]} + R^{[a}_{[\mu\nu}(P)\,f_{\rho]}^{\;b]} &\\
+\tfrac{1}{4}\bar\phi_{[\rho}\Gamma^{ab}R_{\mu\nu]}(Q) + \tfrac{1}{4}\bar\psi_{[\rho}\Gamma^{ab}R_{\mu\nu]}(S) &= 0\,,\\
\mathcal{D}_{[\mu}R_{\nu\rho]}(D) + \tfrac{1}{2}R^{a}_{[\mu\nu}(K)\,e_{\rho]\,a} - R^{a}_{[\mu\nu}(P)\,f_{\rho]\,a} &\\
+\tfrac{1}{4}\bar\phi_{[\rho}R_{\mu\nu]}(Q) - \tfrac{1}{4}\bar\psi_{[\rho}R_{\mu\nu]}(S) &= 0\,.
\end{aligned}
\tag{7.29}
$$

At this stage, the superconformal algebra is not related to symmetries of spacetime. Of course, the gauge fields independently transform as vectors under general coordinate transformations but these transformations have no intrinsic relation with the gauge transformations. This is the reason why, at this stage, there is no need for the bosonic and fermionic degrees of freedom to match, as one would expect for a conventional supersymmetric theory.

There is a procedure to introducing a nontrivial entangling between the spacetime diffeomorphisms and the (internal) symmetries associated with the superconformal gauge algebra, based on curvature constraints. Here one regards the P gauge field $e_\mu{}^a$ as a nonsingular vielbein field, whose inverse will be denoted by $e_a{}^\mu$. This interpretation is in line with the interpretation presented in the previous section, where flat space was viewed as a coset space. In that case, the curvature $R(P)$ has the interpretation of a torsion tensor, and one can impose a constraint $R(P)=0$, so that the M gauge field ω_μ^{ab} becomes a dependent field, just as in (3.6). The effect of this constraint is also that the P gauge transformations are effectively replaced by general-coordinate transformations. To see this, let us rewrite a P-transformation on $e_\mu{}^a$, making use of the fact that there exists an inverse vielbein $e_a{}^\mu$,

$$
\delta e_\mu{}^a = \mathcal{D}_\mu\xi^a_P = \partial_\mu\xi^\nu\,e_\nu{}^a - \xi^\nu\,D_\nu e_\mu{}^a + \xi^\nu R^a_{\mu\nu}(P)\,,
\tag{7.30}
$$

where $\xi^\mu = \xi^a_P\,e_a{}^\mu$. Hence, when imposing the torsion constraint $R(P)=0$, a P-transformation takes the form of a (covariant) general coordinate transformation. This is completely in line with the field transformations (7.22), where the P-transformations were also exclusively represented by coordinate changes, except that we are now dealing with arbitrary diffeomorphisms.

A constraint such as $R(P) = 0$ is called a [illegible] constraint, because it algebraically expresses some of the gauge fields in terms of the others. Of course, by doing so, the transformation rules of the dependent fields are determined and they may acquire extra terms beyond the original ones presented in (7.25). Because $R(P) = 0$ is consistent with spacetime diffeomorphisms, and the bosonic conformal transformations, the field the field $\omega_\mu{}^{ab}$ will still transform under these symmetries according to (7.25). This is also the case for S-supersymmetry, but not for Q-supersymmetry, because the constraint $R(P) = 0$ is inconsistent with Q-supersymmetry. Indeed, under Q-supersymmetry, the field $\omega_\mu{}^{ab}$ acquires an extra term beyond what was presented in (7.25), which is proportional to $R(Q)$. We will not elaborate on the systematics of this procedure but concentrate on a number of noteworthy features. One of them is that there are potentially more conventional constraints. Inspection of (7.28) shows that constraints on $R(M)$, $R(D)$ and $R(Q)$ can be conventional and may lead to additional dependent gauge fields $f_\mu{}^a$ and ϕ_μ associated with special conformal boosts and special supersymmetry transformations. A maximal set of conventional constraints that achieves just that, takes the form

$$\begin{aligned} R^a_{\mu\nu}(P) &= 0\,, \\ e_b{}^{b} R^{ab}_{\mu\nu}(M) &= 0\,, \\ \Gamma^\mu R_{\mu\nu}(Q) &= 0\,, \end{aligned} \tag{7.31}$$

where, for reasons of covariance, one should include possible modifications of the curvatures due to the changes in the transformation laws of the dependent fields. Other than that, the precise form of the constraints is not so important, because constraints that differ by the addition of other covariant terms result in the addition of covariant terms to the dependent gauge fields, which can easily be eliminated by a field redefinition. Note that $R_{\mu\nu}(D)$ is not independent as a result of the first Bianchi identity on $R^a_{\mu\nu}(P)$ given in (7.29) and should not be constrained.

At this point we are left with the vielbein field $e_\mu{}^a$, the gauge field b_μ associated with the scale transformations, and the gravitino field ψ_μ associated with Q-supersymmetry. All other gauge fields have become dependent. The gauge transformations remain with the exception of the P transformations; we have diffeomorphisms, local Lorentz transformations (M), local scale transformations (D), local conformal boosts (K), Q-supersymmetry and S-supersymmetry. Note that b_μ is the only field that transforms non-trivially on special conformal boosts and therefore acts as a compensator which induces all the K-transformations for the dependent fields. Because the constraints are consistent with all the bosonic transformations, those will not change and will still describe a closed algebra. The superalgebra will, however, not close, as one can verify by comparing the numbers of bosonic and fermionic degrees of freedom. In order to have a consistent superconformal theory one must add additional fields (for a review, see [103]). A practical way to do this makes use of the superconformal multiplet of currents [22], which we will discuss in the next section. This construction is limited to theories with $Q = 16$ supercharges and leads to consistent conformal supergravity theories [22, 104, 105].

We close this section with a comment regarding the number of degrees of freedom described by the above gauge fields. The independent bosonic fields, $e_\mu{}^a$ and b_μ, comprise $D^2 + D$ degrees of freedom, which are subject to the $\frac{1}{2}D^2 - \frac{3}{2}D - 1$ independent, bosonic, gauge invariances of the conformal group. This leaves us with $\frac{1}{2}D(D-1) - 1$ degrees of freedom, corresponding to the independent components of a symmetric, traceless, rank-2 tensor in $D - 1$ dimensions, which constitutes an irreducible representation of the Poincaré algebra. This representation is the minimal representation that is required for an off-shell description of gravitons in D spacetime dimensions. A similar off-shell counting argument applies to the fermions, which comprise $(D-2)n_\mathrm{s}$ degrees of freedom after subtracting the gauge degrees of freedom associate with Q- and S-supersymmetry. Here n_s denotes the spinor dimension. Hence, the conformal framework is set up to reduce the field representation to the smallest possible one that describes the leading spin without putting the fields on shell. The fact that the fields can exist off the mass shell, implies that they must constitute massive representations of the Lorentz group. Similarly, the supermultiplet of fields on which conformal supergravity is based, comprise the smallest supermultiplet whose highest spin coincides with the graviton spin.

7.3 Matter fields and currents

In the previous section we described how to set up a consistent gauge theory for conformal supergravity. This theory has an obvious rigid limit, where all the gauge fields are equal to zero, with the exception of the vielbein which is equal to the flat vielbein, $e_\mu{}^a = \delta_\mu^a$. This is the background we considered in Section 7.1. In this background we may have (matter) theories that are superconformally invariant under rigid transformations, described by (7.22). Suppose that we couple such a rigidly superconformal matter theory in first order to the gauge fields of conformal supergravity. Hence we write,

$$\begin{aligned}\mathcal{L} \;=\;& \mathcal{L}_\mathrm{matter} \\ &+ h_\mu^a\,\theta_a{}^\mu + \tfrac{1}{2}\omega_\mu^{ab}\,S^\mu_{ab} + b_\mu\,T^\mu + f_\mu{}^a\,U_a{}^\mu + \bar\psi_\mu J^\mu + \bar\phi_\mu J^\mu_S\,, \end{aligned} \tag{7.32}$$

where h_μ^a denotes the deviation of the vielbein from its flat space value, , $e_\mu{}^a \approx \delta_\mu^a + h_\mu{}^a$. The first term denotes the matter Lagrangian in flat space. The current $\theta_a{}^\mu$ is the energy-momentum tensor. In linearized approximation the above Lagrangian is invariant under superconformal transformations. To examine the consequences of this we need the leading (inhomogeneous) terms in the transformations of the gauge fields ((7.25)),

$$
\begin{array}{rlll}
\text{translations:} & \delta h_\mu^{\ a} = \partial_\mu \xi_P^a \,, & & \\
\text{Lorentz:} & \delta\omega_\mu^{ab} = \partial_\mu \epsilon^{ab} \,, & \delta h_\mu^{\ a} = \epsilon^{ab}\,\delta_{\mu\, b} \,, & \\
\text{dilatations:} & \delta b_\mu = \partial_\mu \Lambda_D \,, & \delta h_\mu^{\ a} = -\delta_\mu^a\,\Lambda_D \,, & \\
\text{conformal boosts:} & \delta f_\mu^{\ a} = \partial_\mu \Lambda_K^a \,, & \delta\omega_\mu^{ab} = \Lambda_K^{[a}\,\delta_\mu^{b]} \,, & \delta b_\mu = \tfrac{1}{2}\Lambda_{K\mu} \,, \\
Q\text{-supersymmetry:} & \delta\psi_\mu = \partial_\mu \epsilon \,, & & \\
S\text{-supersymmetry:} & \delta\phi_\mu = \partial_\mu \eta \,, & \delta\psi_\mu = -\tfrac{1}{2}\Gamma_\mu \eta . &
\end{array}
\tag{7.33}
$$

The variations of the action corresponding to (7.32) under the superconformal transformations, ignoring variations that are proportional to the superconformal gauge fields and assuming that the matter fields satisfy their equations of motion, must vanish. One can verify that this leads to a number of conservation equations for the currents,

$$
\begin{array}{rclrcl}
\partial_\mu \theta_a{}^\mu &=& 0\,, & \partial_\mu U_a{}^\mu - \tfrac{1}{2} S^\mu_{a\mu} - \tfrac{1}{2} T_a &=& 0\,, \\
\partial_\mu S^\mu_{ab} - 2\,\theta_{[ab]} &=& 0\,, & \partial_\mu J^\mu &=& 0\,, \\
\partial_\mu T^\mu + \theta_\mu{}^\mu &=& 0\,, & \partial_\mu J_S^\mu + \tfrac{1}{2}\Gamma_\mu J^\mu &=& 0\,,
\end{array}
\tag{7.34}
$$

where we used the flat vielbein to convert world into tangent space indices and ; for instance, we employed the notation $\theta_{ab} = \theta_a{}^\mu\, e_{\mu\, b}$ and $\theta_\mu{}^\mu = \theta_a{}^\mu\, e_\mu{}^a$. Obviously, not all currents are conserved, but we can define a set of conserved currents by allowing an explicit dependence on the coordinates,

$$
\begin{aligned}
\partial_\mu \theta_a{}^\mu &= 0\,, \\
\partial_\mu \left(S^\mu_{ab} - 2\,\theta_{[a}{}^\mu x_{b]} \right) &= 0\,, \\
\partial_\mu \left(T^\mu + \theta_a{}^\mu x^a \right) &= 0\,, \\
\partial_\mu \left(U_a{}^\mu - \tfrac{1}{2} S^\mu_{ab} x^b - \tfrac{1}{2} T^\mu x^a - \tfrac{1}{2}\theta_b{}^\mu \left(x_a x^b - \tfrac{1}{2} x^2 \delta_a^b \right) \right) &= 0\,, \\
\partial_\mu J^\mu &= 0\,, \\
\partial_\mu \left(J_S^\mu + \tfrac{1}{2}\Gamma_\nu J^\mu x^\nu \right) &= 0\,.
\end{aligned}
\tag{7.35}
$$

In this result one recognizes the various components in (7.22) and in (7.4). For S-supersymmetry one can understand the expression of the current by noting that the following combination of a constant S transformation with a spacetime dependent Q-transformation with $\epsilon = \frac{1}{2} x^\mu \Gamma_\mu \eta$ leaves the gravitino field ψ_μ invariant. Observe that the terms involving the energy-momentum tensor take the form $\theta_a{}^\mu\, \xi^a$, where ξ^a are the conformal Killing vectors defined in (7.4).

So far we have assumed that the gauge fields in (7.32) are independent. However, we have argued in the previous section that it is possible to choose the gauge fields associated with the generators M, K and S, to depend on the other fields. At the linearized level, the fields $\omega_\mu{}^{ab}$, $f_\mu{}^a$ and ϕ_μ can then be written as linear combinations of curls of the independent gauge fields. After a partial integration, the currents $\theta_a{}^\mu$, T^μ and J^μ are modified by improvement terms: terms of the form $\partial_\nu A^{[\nu\mu]}$, which can be included into the currents without affecting their divergence. Hence, the currents S^μ_{ab}, $U_a{}^\mu$ and J^μ_S no longer appear explicitly but are absorbed in the remaining currents as improvement terms. We do not have to work out their explicit form, because we can simply repeat the analysis leading to (7.34), suppressing S^μ_{ab}, $U_a{}^\mu$ and J^μ_S. We then obtain the following conditions for the currents,

$$\begin{aligned} \partial^\mu \theta^{\rm imp}_{\mu\nu} = \theta^{\rm imp}_{[\mu\nu]} = \theta^{\rm imp}{}_\mu{}^\mu = 0\,, \\ \partial^\mu J^{\rm imp}_\mu = \Gamma^\mu J^{\rm imp}_\mu = 0\,. \end{aligned} \tag{7.36}$$

Observe that these equations reduce the currents to irreducible representations of the Poincaré group, in accord with the earlier counting arguments given for the gauge fields.

To illustrate the construction of the currents, let us consider a nonlinear sigma model in flat spacetime with Lagrangian,

$$\mathcal{L} = \tfrac{1}{2} g_{AB}\, \partial_\mu\phi^A\, \partial^\mu\phi^B\,. \tag{7.37}$$

Its energy-momentum operator can be derived by standard methods and is equal to

$$\theta_{\mu\nu} = \tfrac{1}{2} g_{AB} \left(\partial_\mu\phi^A\, \partial_\nu\phi^B - \tfrac{1}{2}\eta_{\mu\nu}\, \partial_\rho\phi^A \partial^\rho\phi^B\right)\,. \tag{7.38}$$

It is conserved by virtue of the field equations; moreover it is symmetric, but not traceless. It is, however, possible to introduce an improvement term,

$$\begin{aligned} \theta^{\rm imp}_{\mu\nu} \quad &= \quad \tfrac{1}{2} g_{AB}(\partial_\mu\phi^A\, \partial_\nu\phi^B - \tfrac{1}{2}\eta_{\mu\nu}\, \partial_\rho\phi^A \partial^\rho\phi^B) \\ &\quad + \frac{D-2}{4(D-1)}\,\left(\eta_{\mu\nu}\,\partial^2 - \partial_\mu\partial_\nu\right)\chi(\phi)\,. \end{aligned} \tag{7.39}$$

When $\chi(\phi)$ satisfies

$$D_A \partial_B \chi(\phi) = g_{AB}\,, \tag{7.40}$$

the improved energy-momentum tensor is conserved, symmetric and traceless (again, upon using the field equations). This implies that, $\chi_A = \partial_A\chi$ is a homothetic vector[30]. From this result it follows that locally in the target

[30] A homothetic vector satisfies $D_A\chi_B + D_A\chi_B = 2g_{AB}$ Here we are dealing with an *exact* homothety, for which $D_A\chi_B = D_B\chi_A$, and which can be solved by a potential χ.

space, χ can be written as

$$\chi = \tfrac{1}{2} g^{AB}\, \chi_A\, \chi_B\,, \tag{7.41}$$

up to an integration constant. Spaces that have such a homothety are cones. To see this, we decompose the target-space coordinates ϕ^A into ϕ and remaining coordinates φ^a, where ϕ is defined by

$$\chi^A \frac{\partial}{\partial \phi^A} = \frac{\partial}{\partial \phi}\cdot \tag{7.42}$$

It then follows that $\chi(\phi, \varphi) = \exp[2\phi]\, \hat{\chi}(\varphi)$, where $\hat{\chi}$ is an undetermined function of the coordinates φ^a. In terms of these new coordinates we have $\chi^A = (1, 0, \cdots, 0)$ and $g_{A\phi} = \chi_A = (2\chi, g_{a\phi})$. From this result one proves directly that the metric takes the form

$$(\mathrm{d}s)^2 = \frac{(\mathrm{d}\chi)^2}{2\chi} + \chi h_{ab}(\varphi)\, \mathrm{d}\varphi^a \mathrm{d}\varphi^b\,, \tag{7.43}$$

where the ϕ-independence of h_{ab} can be deduced directly from (7.40). This result shows that the target space is a cone over a base manifold $\mathcal{M}_B$ parametrized in terms of the coordinates φ^a with metric h_{ab} [106]. In the supersymmetic context it is important to note that, when the cone is a Kähler or hyperkähler space, the cone must also be invariant under $U(1)$ or $SU(2)$. These features play an important role when extending to the supersymmetric case. In that case $U(1)$ or $SU(2)$ must be associated with the R-symmetry of the superconformal algebra.

Coupling the improved energy-momentum tensor (7.39) to gravity must lead to a conformally invariant theory of the nonlinear sigma model and gravity. The relevant Lagrangian reads,

$$e^{-1}\, \mathcal{L} = \tfrac{1}{2} g_{AB}\, \partial_\mu \phi^A\, \partial^\mu \phi^B - \frac{D-2}{4(D-1)}\, \chi(\phi)\, R\,. \tag{7.44}$$

Indeed, this Lagrangian is invariant under local scale transformations characterized by the functions $\Lambda_D(x)$,

$$\delta_D \phi^A = w \Lambda_D \chi^A\,, \qquad \delta_D g_{\mu\nu} = -2\Lambda_D\, g_{\mu\nu} \tag{7.45}$$

where w is the Weyl weight of the scalar fields which is equal to $w = \frac{1}{2}(D-2)$. The transformation of $g_{\mu\nu}$ is in accord with the vielbein scale transformation written down in Section 7.2. We should also point out that the coupling with the Ricci scalar can be understood in the context of the results of the previous section. Using the gauge fields of the conformal group, the Lagrangian reads,

$$e^{-1}\, \mathcal{L} = \tfrac{1}{2} g_{AB}\, g^{\mu\nu} (\partial_\mu \phi^A - w\, b_\mu\, \chi^A)\, (\partial_\nu \phi^B - w\, b_\nu\, \chi^B) - \tfrac{1}{2} w\, f_\mu{}^\mu\, \chi\,. \tag{7.46}$$

As one can easily verify from the transformation rules (7.25), this Lagrangian is invariant under local dilatations, conformal boosts and spacetime diffeomorphisms. Upon using the second constraint (7.31) for the gauge field ${f_\mu}^a$ associated with the conformal boosts and setting $b_\mu = 0$ as a gauge condition for the conformal boosts, the Lagrangian becomes equal to (7.44), which is still invariant under local dilatations. This example thus demonstrates the relation between improvement terms in the currents and constraints on the gauge fields.

It is possible to also employ a gauge condition for the dilatations. An obvious one amounts to putting χ equal to a constant χ_0, with the dimension of $[\mathrm{mass}]^{D-2}$,

$$\chi = \chi_0 \,. \tag{7.47}$$

Substituting the metric (7.43) the Lagrangian then acquires the form,

$$e^{-1}\,\mathcal{L} \propto \tfrac{1}{2} h_{ab}\,\partial_\mu\varphi^a\,\partial^\mu\varphi^b - \frac{D-2}{4(D-1)}\,R\,. \tag{7.48}$$

This Lagrangian describes a nonlinear sigma model with the base manifold $\mathcal{M}_B$ of the cone as a target space, coupled to (nonconformal) gravity. The constant χ_0 appears as an overall constant and is inversely proportional to Newton's constant in D spacetime dimensions. Observe that in order to obtain positive kinetic terms, the metric h_{ab} should be negative definite and χ_0 must be positive.

The above example forms an important ingredient in the so-called superconformal multiplet calculus that has been used extensively in the construction of nonmaximal supergravity couplings. There is an extensive literature on this. For an introduction to the 4-dimensional $N = 1$ multiplet calculus, see, [81], for 4-dimensional $N = 2$ vector multiplets and hypermultiplets, we refer to [67, 107].

References

[1] P. van Nieuwenhuizen, *Phys. Rep.* **68** (1981) 189; B. de Wit and D.Z. Freedman, *Supergravity – The basics and beyond*, in *Supersymmetry*, NATO ASI *B* **125** 135, edited by K. Dietz, R. Flume, G. von Gehlen and V. Rittenberg (Plenum, 1985); S. Ferrara, *Supersymmetry*, Vols. 1 & 2 (North-Holland/World Scientific, 1989); A. Salam and E. Sezgin, *Supergravities in diverse dimensions*, Vols. 1 & 2 (North-Holland/World Scientific, 1989).

[2] M. Green, J. Schwarz and E. Witten, *Superstring theory*, Vols. 1 & 2 (Cambridge Univ. Press, 1987); D. Lüst and S. Theisen, *Lectures on string theory* (Springer, 1989); J. Polchinski, *String Theory*, Vols. I & II (Cambridge Univ. Press, 1998); E. Kiritsis, *Leuven Notes Math. Theor. Phys.* **9** (1997) [`hep-th/9709062`]; *AIP Conf. Proc.* **419** (1998) 265 [`hep-th/9708130`].
For a collection of reprints, see, for example, J.H. Schwarz, *Superstrings. The first 15-years of superstring theory*, Vols. 1 & 2 (World Scientific, 1985); M. Dine, *String*

Theory in four dimensions (North-Holland, 1988); A.N. Schellekens, *Superstring construction* (North-Holland, 1989).

[3] W. Nahm, *Nucl. Phys. B* **135** (1978) 149.

[4] S. Ferrara, C.A. Savoy and B. Zumino, *Phys. Lett. B* **100** (1981) 393; S. Ferrara and C.A. Savoy, *Representations of extended supersymmetry on one- and two-particle states*, in *Supergravity '81*, edited by S. Ferrara and J.G. Taylor (Cambridge Univ. Press, 1982), p. 47; B. de Wit, *Maximal supergravity*, in *Frontiers in Particle Physics '83*, edited by Dj. Šijački, N. Bilić, B. Dragovič and D. Popovič (World Scientific, 1984), p. 152.

[5] J. Strathdee, *Int. J. Mod. Phys. A* **2** (1987) 273.

[6] B. de Wit and J. Louis, *Supersymmetry and dualities in various dimensions*, in *Strings, Branes and Dualities*, NATO ASI **C520**, edited by L. Baulieu, P. Di Francesco, M. Douglas, V. Kazakov, M. Picco and P. Windey (Kluwer, 1999), p. 33 [hep-th/9801132].

[7] A. Van Proeyen, *Tools for supersymmetry* [hep-th/9910030].

[8] B. de Wit, A. Tollstén and H. Nicolai, *Nucl. Phys. B* **392** (1993) 3 [hep-th/9208074].

[9] E. Cremmer, *Supergravities in 5 dimensions*, in *Superspace & Supergravity*, edited by S.W. Hawking and M. Roček (Cambridge Univ. Press, 1981), p. 267.

[10] C.M. Hull, *JHEP* **0006** (2000) 019 [hep-th/0004086]; *Nucl. Phys. B* **583** (2000) 237 [hep-th/0004195].

[11] E. Cremmer, B. Julia and J. Scherk, *Phys. Lett. B* **76** (1978) 409.

[12] R. Slansky, *Phys. Rep.* **79** (1981) 1.

[13] L. Brink, J. Scherk and J.H. Schwarz, *Nucl. Phys. B* **121** (1977) 77; F. Gliozzi, J. Scherk and D. Olive, *Nucl. Phys. B* **122** (1977) 253.

[14] M.B. Green and J.H. Schwarz, *Phys. Lett. B* **122** (1983) 143; J.H. Schwarz and P.C. West, *Phys. Lett. B* **126** (1983) 301; J.H. Schwarz, *Nucl. Phys. B* **226** (1983) 269; P. Howe and P.C. West, *Nucl. Phys. B* **238** (1984) 181.

[15] M. Huq and M.A. Namazie, *Class. Quantum Grav.* **2** (1985) 293; F. Giani and M. Pernici, *Phys. Rev. D* **30** (1984) 325; I.C.G. Campbell and P.C. West, *Nucl. Phys. B* **243** (1984) 112.

[16] M. Dine, P. Huet and N. Seiberg, *Nucl. Phys. B* **322** (1989) 301.

[17] J. Dai, R.G. Leigh and J. Polchinski, *Mod. Phys. Lett. A* **4** (1989) 2073.

[18] E. Bergshoeff, C. Hull and T. Ortin, *Nucl. Phys. B* **451** (1995) 547 [hep-th/9504081].

[19] M. Abou-Zeid, B. de Wit, H. Nicolai and D. Lüst, *Phys. Lett. B* **466** (1999) 144 [hep-th/9908169].

[20] P.K. Townsend, *Phys. Lett. B* **139** (1984) 283.

[21] R. D'Auria, S. Ferrara and C. Kounnas, $N = (4,2)$ *Chiral supergravity in six dimensions and solvable Lie algebras* [hep-th/9711048].

[22] E.A. Bergshoeff, M. de Roo and B. de Wit, *Nucl. Phys. B* **182** (1981) 173.

[23] P.K. Townsend, *Four lectures on M-theory*, lectures given at the 1996 ICTP Summer School in High Energy Physics and Cosmology, Trieste [hep-th/9612121]; *M-theory from its superalgebra*, in *Strings, Branes and Dualities*, NATO ASI **C520**, edited by L. Baulieu, P. Di Francesco, M. Douglas, V. Kazakov, M. Picco and P. Windey (Kluwer, 1999), p. 141 [hep-th/9712004].

[24] N.A. Obers and B. Pioline, *Phys. Rep.* **318** (1999) 113 [hep-th/9809039].

[25] S. Ferrara, *Nucl. Phys. B (Proc. Suppl.)* **55** (1997) 145; L. Andrianopoli, R. D'Auria and S. Ferrara, *U-Duality and central charges in various dimensions revisited* [hep-th/9612105].

[26] R. Coquereaux, *Phys. Lett. B* **115** (1982) 389, and references quoted therein.

[27] S. Ferrara, M. Kaku, P.K. Townsend and P. van Nieuwenhuizen, *Nucl. Phys. B* **129** (1977) 125.

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

[29] M. Günaydin, L.J. Romans and N.P. Warner, *Nucl. Phys. B* **272** (1986) 598.

[30] B. de Wit, *Introduction to supergravity*, in *Supersymmetry and Supergravity '84*, edited by B. de Wit, P. Fayet and P. van Nieuwenhuizen (World Scientific, 1984), p. 49.

[31] P.K. Townsend, *Phys. Rev. D* **15** (1977) 2802.

[32] S. Deser and B. Zumino, *Phys. Rev. Lett.* **38** (1977) 1433.

[33] S. Ferrara, *Phys. Lett. B* **69** (1977) 481.

[34] B. de Wit and A. Zwartkruis, *Class. Quantum Grav.* **4** (1987) L59.

[35] K. Bautier, S. Deser, M. Henneaux and D. Seminara, *Phys. Lett. B* **406** (1997) 49.

[36] C. Fronsdal, *Phys. Rev. D* **18** (1978) 3264; J. Fang and C. Fronsdal, *Phys. Rev. D* **18** (1978) 3630; J. Schwinger, *Particles, sources and fields* (Addison Wesley, 1970); T. Curtright, *Phys. Lett. B* **85** (1979) 219.

[37] C. Aragone and S. Deser, *Spin 2 matter-graviton coupling problems*, in *Supergravity*, edited by P. van Nieuwenhuizen and D.Z. Freedman (North-Holland, 1979).

[38] C. Aragone and S. Deser, *Phys. Lett. B* **86** (1979) 161; F.A. Berends, J.W. van Holten, P. van Nieuwenhuizen and B. de Wit, *J. Phys. A* **13** (1980) 1643; B. de Wit and D.Z. Freedman, *Phys. Rev. D* **21** (1980) 358.

[39] M.A. Vasiliev, *Int. J. Mod. Phys. D* **5** (1996) 763 [`hep-th/9611024`].

[40] B. de Wit, *Int. J. Mod. Phys. A* **16** (2001) 1002, Proc. *Strings 2000*, edited by M.J. Duff, J.T. Liu and J. Lu [`hep-th/0010292`].

[41] B. de Wit and H. Nicolai, *Class. Quantum Grav.* **18** (2001) 3095 [`hep-th/0011239`].

[42] P.K. Townsend and P. van Nieuwenhuizen, *Phys. Lett. B* **67** (1977) 439; A.H. Chamseddine and P.C. West, *Nucl. Phys. B* **129** (1977) 39.

[43] D.N. Page, *Phys. Rev. D* **28** (1983) 2976.

[44] M.J. Duff and K.S. Stelle, *Phys. Lett. B* **253** (1991) 113.

[45] R. Güven, *Phys. Lett. B* **276** (1992) 49.

[46] For a review see, for example, M.J. Duff, R.R. Khuri and J.X. Lü, *Phys. Rep.* **259** (1995) 213 [`hep-th/9412184`]; K.S. Stelle, *Lectures on supergravity p-branes*, in *Trieste 1996, High energy physics and cosmology*, p. 287 [`hep-th/9701088`].

[47] B. Biran, F. Englert, B. de Wit and H. Nicolai, *Phys. Lett. B* **124** (1983) 45.

[48] For a reprint collection and a review of Kaluza-Klein theory, see, for instance, M.J. Duff, B.E.W. Nilsson and C.N. Pope, *Phys. Rep.* **130** (1986) 1; T Appelquist, A. Chodos and P.G.O. Freund, *Modern Kaluza-Klein Theories* (Addison-Wesley, 1987).

[49] B. de Wit, F. Vanderseypen and A. Van Proeyen, *Nucl. Phys. B* **400** (1993) 463 [`hep-th/9210068`].

[50] E. Cremmer, *N=8 Supergravity*, in *Unification of fundamental particle interactions*, edited by S. Ferrara, J. Ellis and P. van Nieuwenhuizen (Plenum, 1980), p. 137; B. Julia, *"Group disintegrations"*, in *Superspace & Supergravity*, edited by S.W. Hawking and M. Roček (Cambridge University Press, 1981), p. 331.

[51] B. de Wit, *Nucl. Phys. B (Proc. Suppl.)* **101** (2001) 154, Proc. *Thirty Years of Supersymmetry*, edited by K.A. Olive, S. Rudaz and M. Shifman [`hep-th/0103086`].

[52] Y. Tanii, *Phys. Lett. B* **145** (1984) 368.

[53] E. Cremmer and B. Julia, *Phys. Lett. B* **80** (1978) 48.

[54] C.M. Hull and P.K. Townsend, *Nucl. Phys. B* **438** (1995) 109 [`hep-th/9410167`].

[55] E. Witten, *Nucl. Phys. B* **443** (1995) 85 [`hep-th/9503124`].

[56] P.K. Townsend, *Phys. Lett. B* **350** (1995) 184 [`hep-th/9501068`]; *Phys. Lett. B* **373** (1996) 68 [`hep-th/9512062`].

[57] E.A. Bergshoeff, E. Sezgin and P.K. Townsend, *Phys. Lett. B* **189** (1987) 75; *Ann. Phys.* **185** (1988) 330.

[58] J.H. Schwarz, *Phys. Lett. B* **367** (1996) 97 [`hep-th/9510086`]; P.S. Aspinwall, *Nucl. Phys. B (Proc. Suppl.)* **46** (1996) 30 [`hep-th/9508154`].

[59] A.H. Chamseddine, *Phys. Rev. D* **24** (1981) 3065; E. Bergshoeff, M. de Roo, B. de Wit and P. van Nieuwenhuizen, *Nucl. Phys. B* **195** (1982) 97; E. Bergshoeff, M. de Roo and B. de Wit, *Nucl. Phys. B* **217** (1983) 143; G. Chapline and N.S. Manton, *Phys. Lett. B* **120** (1983) 105.

[60] J. Polchinski and E. Witten, *Nucl. Phys. B* **460** (1996) 525 [`hep-th/9510169`]; I. Antoniadis, H. Partouche and T.R. Taylor, *NATO ASI* **C520**, edited by L. Baulieu, P. Di Francesco, M. Douglas, V. Kazakov, M. Picco and P. Windey (Kluwer, 1999) p. 179 [`hep-th/9706211`].

[61] R. Gilmore, *Lie groups, Lie algebras, and some of their applications* (Wiley Interscience, 1974).

[62] L. Castellani, R. D'Auria and P. Fré, *Supergravity and superstrings I* (World Scientific, 1991).

[63] B. de Wit and H. Nicolai, *Phys. Lett. B* **108** (1982) 285; *Nucl. Phys. B* **208** (1982) 323.

[64] B. de Wit, *Nucl. Phys. B* **158** (1979) 189.

[65] M.K. Gaillard and B. Zumino, *Nucl. Phys. B* **193** (1981) 221.

[66] S. Cecotti, S. Ferrara and L. Girardello, *Int. J. Mod. Phys. A* **4** (1989) 2475.

[67] B. de Wit and A. Van Proeyen, *Nucl. Phys. B* **245** (1984) 89.

[68] P.K. Townsend, K. Pilch and P. van Nieuwenhuizen, *Phys. Lett. B* **136** (1984) 38.

[69] H. Nicolai and H. Samtleben, *Phys. Rev. Lett.* **86** (2001) 1686 [`hep-th/0010076`].

[70] M. Pernici, K. Pilch and P. van Nieuwenhuizen, *Phys. Lett. B* **143** (1984) 103.

[71] C.M. Hull, *Phys. Rev. D* **30** (1984) 760; *Phys. Lett. B* **142** (1984) 39; *Phys. Lett. B* **148** (1984) 297; *Class. Quantum Grav.* **2** (1985) 343; *New gauged $N = 8$, $D = 4$ supergravities* [`hep-th/0204156`].

[72] L.Andreanopoli, F. Cordaro, P. Fré and L. Gualtieri, *Class. Qunatum Grav.* **18** (2001) 395 [`hep-th/0009048`].

[73] L. Andreanopoli, R. D'Auria, S. Ferrara and M.A. Lledó, *JHEP* **0207** (2002) 010 [`hep-th/0203206`]; *Duality and spontaneously broken supergravity in flat backgrounds* [`hep-th/0204145`].

[74] H. Nicolai and H. Samtleben, *JHEP* **0104** (2001) 022 [`hep-th/0103032`].

[75] B. de Wit, H. Samtleben and M. Trigiante (to appear).

[76] J.-W. van Holten and A. Van Proeyen, *J. Phys. A* **15** (1982) 3763.

[77] S.W. Hawking and G.F.R. Ellis, *The large scale structure of space-time* (Cambridge Univ. Press, 1973).

[78] S.J. Avis, C.J. Isham and D. Storey, *Phys. Rev. D* **18** (1978) 3565.

[79] K. Pilch and A.N. Schellekens, *J. Math. Phys.* **25** (1984) 3455.

[80] B. de Wit and I. Herger, *Lect. Notes Phys.* **541** (2000) 79, also in *Polanica 1999, Towards Quantum Gravity*, p. 79 [`hep-th/9908005`].

[81] B. de Wit, *Multiplet calculus*, in *Supersymmetry and Supergravity '82*, edited by S. Ferrara, J.G. Taylor and P. Van Nieuwenhuizen (World Scientific, 1983).

[82] P. Breitenlohner and D.Z. Freedman, *Ann. Phys.* **144** (1982) 249.

[83] C. Fronsdal, *Rev. Mod. Phys.* **37** (1965) 221; *Phys. Rev. D* **10** (1974) 589; C. Fronsdal and R.B. Haugen, *Phys. Rev. D* **12** (1975) 3810; C. Fronsdal, *Phys. Rev. D* **12** (1975) 3819.

[84] D.Z. Freedman and H. Nicolai, *Nucl. Phys. B* **237** (1984) 342.

[85] H. Nicolai, *Representations of supersymmetry in anti-de Sitter space*, in *Supersymmetry and Supergravity '84*, edited by B. de Wit, P. Fayet, P. van Nieuwenhuizen (World Scientific, 1984).

[86] P.A.M. Dirac, *J. Math. Phys.* **4** (1963) 901.

[87] M. Günaydin, P. van Nieuwenhuizen and N.P. Warner, *Nucl. Phys. B* **255** (1985) 63.

[88] M. Günaydin and C. Saclioglu, *Commun. Math. Phys.* **91** (1982) 159; M. Günaydin, *Oscillator like unitary representations of noncompact groups and supergroups and extended supergravity theories*, in Int. Colloq. on Group Theoretical Methods in Physics (Istanbul, 1982), edited by M. Serdaroglu and E. Inönü, *Lecture Notes Phys.* **180** (Springer, 1983).

[89] M. Günaydin, *Singleton and doubleton supermultiplets of space-time supergroups and infinite spin superalgebras*, in *Supermembranes and physics of* 2+1 *dimensions*, edited by M.J. Duff, C.N. Pope and E. Sezgin (World Scientific, 1990), p. 442.

[90] M. Günaydin and D. Minic, *Nucl. Phys. B* **523** (1998) 145 [hep-th/9802047]; M. Günaydin D. Minic and M. Zagerman, *Nucl. Phys. B* **534** (1998) 96 [hep-th/9806042].

[91] W. Heidenreich, *Phys. Lett. B* **110** (1982) 461.

[92] M. Günaydin and N. Marcus, *Class. Quantum Grav.* **2** (1985) L11; M. Günaydin and N.P. Warner, *Nucl. Phys. B* **272** (1986) 99.

[93] S. Ferrara and C. Fronsdal, *Class. Quantum Grav.* **15** (1998) 2153 [hep-th/9712239].

[94] E. Sezgin and P. Sundell, *Massless higher spins and holography* [hep-th/0205131].

[95] A. Segal, *Conformal higher spin theory* [hep-th/0207212].

[96] J.D. Brown and M. Henneaux, *Commun. Math. Phys.* **104** (1986) 207.

[97] S.S. Gubser, I.R. Klebanov and A.M. Polyakov, *Phys. Lett. B* **428** (1998) 105 [hep-th/9802109].

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

[99] V. Balasubramanian, P. Kraus and A. Lawrence, *Phys. Rev. D* **59** (1999) 046003 [hep-th/9805171].

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

[101] S. de Haro, K. Skenderis and S.N. Solodukhin, *Commun. Math. Phys.* **217** (2001) 595 [hep-th/0002230].

[102] G. Mack and A. Salam, *Ann. Phys.* **53** (1969) 174.

[103] B. de Wit, *Conformal invariance in extended supergravity*, in *Supergravity '81*, edited by S. Ferrara and J.G. Taylor (Cambridge Univ. Press, 1982).

[104] M. Kaku, P.K. Townsend and P. van Nieuwenhuizen, *Phys. Rev. D* **17** (1978) 3179; P.K. Townsend and P. van Nieuwenhuizen, *Phys. Rev. D* **19** (1979) 3166.

[105] B. de Wit and J.-W. van Holten, *Nucl. Phys. B* **155** (1979) 530; B. de Wit, J.-W. van Holten and A. Van Proeyen, *Nucl. Phys. B* **167** (1980) 186.

[106] G.W. Gibbons and P. Rychenkova, *Phys. Lett. B* **443** (1998) 138 [hep-th/9809158].

[107] K. Galicki, *Class. Quantum Grav.* **9** (1992) 27; B. de Wit, B. Kleijn and S. Vandoren, *Rigid $N = 2$ superconformal hypermultiplets*, in *Dubna 1979, Supersymmetries and Quantum Symmetries*, p. 37 [hep-th/9808160]; *Nucl. Phys. B* **568** (2000) 475 [hep-th/9909228]; B. de Wit, M. Roček and S. Vandoren, *JHEP* **0102** (2001) 039 [hep-th/0101161].

LECTURE 2

SUPERSYMMETRIC GAUGE THEORIES

E. RABINOVICI

Racah Institute, Hebrew University,
Jerusalem, Israel

Contents

SUPERSYMMETRIC GAUGE THEORIES

D.S. Berman[1] and E. Rabinovici[1]

Abstract

We introduce simple and more advanced concepts that have played a key role in the development of supersymmetric systems. This is done by first describing various supersymmetric quantum mechanics models. Topics covered include the basic construction of supersymmetric field theories, the phase structure of supersymmetric systems with and without gauge particles, superconformal theories and infrared duality in both field theory and string theory. A discussion of the relation of conformal symmetry to a vanishing vacuum energy (cosmological constant) is included.

1 Introduction

At least three phases of gauge theory are manifest in nature. The weak interactions are in the Higgs phase. The electromagnetic interaction is in the Coulomb phase and the colour interactions are in the confining phase. None of these phases exhibits supersymmetry explicitly. What is then the motivation to introduce and study supersymmetry? Here is a list of reasons that have motivated people over the years.

- In their seminal paper Golfand and Likhtman [1] introduced supersymmetry in order to constrain the possible forms of interaction, little did they know that the constrained interactions [2] will give rise to a multitude of vacua;
- The local version of supersymmetry contains automatically gravity [3,4], the force in nature that has yet to be tamed by theoretical physics;

This work is supported by BSF- the American-Israeli Bi-National Science foundation, the Israel academy of sciences and humanities-Centers of excellence program, the German-Israeli Bi-National science foundation, the european RTN network HPRN-CT-2000-00122 and DOE grant DE-FG02-90ER40560.

[1]Racah Institute, Hebrew University, Jerusalem, Israel.

- The weak interaction scale and the Plank scale are widely separated for theoreticians this could be a problem, called the Hierarchy problem. This problem is much softer in supersymmetric theories [5–7];

- In various supersymmetric models the gauge couplings unify sensibly [8–10];

- Strong coupling problems are generically intractable in field theory. In the presence of supersymmetry constraints the problems can be solved in interesting cases. In particular the idea of monopole condensation as the responsible mechanism for confinement becomes tractable. Such models serve as analytical laboratories for physical ideas;

- Perturbative string theory seams to need supersymmetry in order to be stable. In fact, in order for string theories to make sense the number of Bosonic degrees of freedom can differ from those of the Fermionic degrees of freedom by at most those of a $d = 2$ field theory [11].

The construction and analysis of supersymmetric systems requires the use of many different concepts and ideas in field theory. In section two we will introduce several of these ideas by using a simple quantum mechanical context. The ideas include: supersymmetry and its spontaneous breaking; index theorems as a tool to answer physical questions; the impact of conformal symmetries; and treating theories with no ground states.

Section three will contain a review of the methods usedto construct supersymmetric field theories and supersymmetric gauge theories. There are many excellent reviews on the subject, we refer the reader to them for a more complete exposition [3, 4, 12–18].

In section four, we review the intricate phase structure of classical and quantum supersymmetric gauge theories. It may occasionally seem that one sees only the trees and not the forest however note that many of the trees are central problems in field theory for years. They include the understanding of confinement, chiral symmetry breaking and the emergence of massless fermions to mention just a few. the material in this section is elaborated in similar or greater detail in [12, 15]. This section includes also a discussion of conformal field theories and some properties in dimension greater than two including four. In particular, the vacuum energy of these theories is discussed in relation to the cosmological constant problem.

In section five we describe the phenomena of infra-red duality in supersymmetric gauge theories. This duality is described from both a field theoretical and string theoretical point of view. In the process we will discuss connections between string theory in the presence of branes and gauge theories.

Many topics in the study of supersymmetric theories which are as worthy have not been covered for lack of time or because they were covered by other lectures. Examples of such topics are: large N gauge theories [19]; non-supersymmetric deformations of the models described [20]; supersymmetric matrix models [21]; supersymmetry on the lattice [22] as well as many others.

2 Supersymmetric quantum mechanics

The ideas in this section were introduced in [24, 25]. First we will examine the quantum mechanical realization of the supersymmetry algebra so as to introduce various ideas that will later carry over to field theory. The questions we wish to examine are: what is supersymmetry; how is SUSY broken spontaneously; and what are nonrenormalisation theorems. Along the way we will introduce some useful tools such as the Witten index. Now we begin with a one variable realization of $N = 1$ SUSY. Later we will present a two variable realization where we can introduce the notion of a flat direction.

Quantum mechanics is a one dimensional field theory. The Hamiltonian is the only member of the Poincare group in that case. Thus, the basic anti-commutation relations that define the supersymmetric algebra are:

$$\{Q_i, Q_j^+\} = 2H\delta_{ij} \qquad\qquad i, j = 1..N \tag{2.1}$$

H is the Hamiltonian and Q, Q^+ are called the supercharges. N denotes the number of supersymmetries.

A rather general $N = 1$ realization with n bosonic and n Fermionic degrees of freedom is given by:

$$Q = \sum_{\alpha=1}^{n} \psi_\alpha^+ \left(-p_\alpha + i\frac{\partial W}{\partial x_\alpha}\right), \tag{2.2}$$

$W(x, .., x_n)$ is a general function of the n bosonic variables. The Hamiltonian is then given by:

$$H = \frac{1}{2}\{Q, Q^+\} = \frac{1}{2}\left(\left(\sum_\alpha p_\alpha^2 + \frac{\partial W}{\partial x_\alpha}^2\right) 1_{L\times L} - \sum_{\alpha\beta} B_{\alpha\beta}\frac{\partial^2 W}{\partial x_\alpha \partial x_\beta}\right) \tag{2.3}$$

and B is:

$$B_{\alpha\beta} = \frac{1}{2}[\psi_\alpha^+, \psi_\beta] \tag{2.4}$$

and the ψ variables obey the following anticommutation relations.

$$\{\psi^+_\alpha, \psi_\beta\} = \delta_{\alpha\beta}\ ,\ \ \{\psi_\alpha, \psi_\beta\} = 0\ ,\ \ \{\psi^+_\alpha, \psi^+_\beta\} = 0. \tag{2.5}$$

The dimension of the Fermionic Hilbert space is $L = 2^n$. Before moving on to $n = 1$ realization of this super algebra, we first will recall some basic facts about the Bosonic harmonic oscillator.

$$H = \frac{p_q^2}{2m} + \frac{1}{2}m\omega^2 q^2 \tag{2.6}$$

define The energy scale is extracted by defining dimensionless variables x and p_x.

$$x = \left(\frac{m\omega}{\hbar}\right)^{\frac{1}{2}} q\ , \qquad p_x = (m\omega\hbar)^{-\frac{1}{2}} p_q. \tag{2.7}$$

This gives the following commutation relations:

$$-i\hbar = [p_q, q] = \hbar[p_x, x]\ , \tag{2.8}$$

the Hamiltonian is now given by:

$$H = \hbar\omega\frac{1}{2}(p_x^2 + x^2) \tag{2.9}$$

$\hbar\omega$ is the energy scale in the problem. It is useful to define creation and annihilation operators, a^+, a by:

$$x = \frac{1}{\sqrt{2}}(a + a^+)\ ,\ \ p = \frac{i}{\sqrt{2}}(a^+ - a) \tag{2.10}$$

and so we obtain the commutation relation:

$$[a, a^+] = 1. \tag{2.11}$$

The number operator, N is constructed out of the creation and annihilation operators:

$$N = a^+ a. \tag{2.12}$$

The spectrum of this operator is given by the nonnegative integers. The Hamiltonian in these variables then becomes:

$$H = \hbar\omega\left(N + \frac{1}{2}\right) \tag{2.13}$$

and the energy spacings are given by the scale $\hbar\omega$.

The energy of the ground state differs from the classical minimum energy by $\frac{1}{2}\hbar\omega$. This is essentially due the uncertainty relation.

The eigenstates, $|n\rangle$ can be recast in terms of the variable x. The ground state, $|0\rangle$ is obtained by solving the equation:

$$a|0\rangle = 0. \tag{2.14}$$

Note that this equation is a first order as opposed to the second order equation that one would have to solve if one attempted to directly solve the Schroedinger equation. This is possible only for the ground state. Using equations (2.10, 2.14):

$$(x + ip)\phi(x) = 0. \tag{2.15}$$

That is:

$$\left(x + \frac{\mathrm{d}}{\mathrm{d}x}\right)\phi(x) = 0 \tag{2.16}$$

yielding the ground state wave function:

$$\phi(x) = c \exp\left(-\frac{x^2}{2}\right) \cdot \tag{2.17}$$

The state $|n\rangle$ is given by:

$$|n\rangle = \frac{(a^+)^n}{\sqrt{n!}}|0\rangle , \tag{2.18}$$

which can now be expressed in x-space. This completes the solution of the Bosonic harmonic oscillator.

Also note that one could have written H as:

$$H = \hbar\omega\left(aa^+ - \frac{1}{2}\right) \cdot \tag{2.19}$$

So why not obtain a ground state energy of $-\frac{1}{2}\hbar\omega$ by solving

$$a^+|\mathrm{GS}\rangle = 0. \tag{2.20}$$

This first order equation always has a solution. It is given by:

$$\phi(x) = c \exp\left(\frac{1}{2}x^2\right) . \tag{2.21}$$

To find a quantum state it is not enough that it be a zero energy solution of the Schrödinger equation but it must also be normalizable (by normalizable we include [illegible] normalisability).

The above solution is not even plane wave normalizable and hence it is not a quantum state.

Physically, how do we motivate this restriction? After all, the universe may well be finite. In such a circumstance the wave function will be exponentially confined to the edge of the universe and thus irrelevant for bulk physics. There may be situations where one will want to study the physics on the boundary. In such cases the "non-normalizable" states should be kept and may play an important role. From now on we will only accept plane wave normalizable states and nothing "worse".

Also recall that, the simple harmonic oscillator is a useful approximation for small excitations above the minimum of a generic potential.

We will now note some basic facts concerning the validity of a perturbative expansion. Consider the Hamiltonian,

$$H = \frac{p_q^2}{2m} + \frac{1}{2} g q^n. \tag{2.22}$$

One may wonder whether if one can make a perturbative expansion in small or large g or small or large m. To answer this one needs to find out if one can remove the g, m dependence the same type of rescaling used for the harmonic oscillator (2.7). Is it possible to define a new set of dimensionless canonical variables p_x, x that preserve the commutation relations, such that:

$$[p_q, q] = [p_x, x]\hbar \tag{2.23}$$

and

$$H = h(m, g)\frac{1}{2}(p_x^2 + x^n). \tag{2.24}$$

We will make the following ansatz:

$$q = f(m, g)x\ , \qquad p_q = \frac{1}{f(m, g)} p_x\ , \tag{2.25}$$

giving

$$2H = \frac{p_x^2}{mf^2(m, g)} + gf(m, g)^n x^n\ , \tag{2.26}$$

and so one may choose

$$gf(m, g) = \left(\frac{1}{mf(m, g)^2}\right)^{\frac{1}{n+2}}. \tag{2.27}$$

The Hamiltonian becomes:

$$H = g^{1-\frac{n}{n+2}} m^{-\frac{n}{n+2}} \frac{1}{2}(p_q^2 + q^n). \tag{2.28}$$

The role of g and m is just to determine the overall energy scale. They may not serve as perturbation parameters. (This does not apply to the special case of $n = -2$, which we will return to later.) Thus, Hamiltonians of the form $p^2 + x^n$ can't be analyzed perturbatively. Further analysis shows that for Hamiltonians of the form $p^2 + x^n + x^m$ with $m < n$, perturbation theory is valid.

After reviewing the Bosonic harmonic oscillator, we review the harmonic Fermionic Oscillator. The commutation relations become anticommutation relations.

$$\{a_{\rm F}, a_{\rm F}^+\} = 1\ ,\ \ \{a_{\rm F}, a_{\rm F}\} = \{a_{\rm f}^+, a_{\rm F}^+\} = 0. \tag{2.29}$$

The Hamiltonian (that does not have a classical analogue) is taken to be,

$$H = \hbar\omega_{\rm F}(a_{\rm F}^+ a_{\rm F}\ +\ \text{const.})\ , \tag{2.30}$$

where $\omega_{\rm F}$ is the Fermionic oscillator frequency and,

$$[a_{\rm F}^+, H] = -a_{\rm F}^+\ , \qquad [a_{\rm F}, H] = a_{\rm F}. \tag{2.31}$$

This demonstrates that $a_{\rm F}^+$, $a_{\rm F}$ are creation and annihilation operators. The spectrum of $N_{\rm F}$ is 0, 1. This is essentially a manifestation of the Pauli exclusion principle. The states are given by

$$|0\rangle\ ,\ \ a_{\rm F}^+|0\rangle. \tag{2.32}$$

This algebra can be realized using the Pauli matrices $\{\sigma^i\}$ as follows. Define,

$$\sigma_- = \frac{1}{2}(\sigma_1 - i\sigma_2)\ ,\ \ \sigma_+ = \frac{1}{2}(\sigma_1 + i\sigma_2)\cdot \tag{2.33}$$

We then identify,

$$a_{\rm F} = \sigma_-\ ,\ \ a_{\rm F}^+ = \sigma_+. \tag{2.34}$$

Hence the Hamiltonian becomes:

$$H = \frac{1}{2}\hbar\omega_{\rm F}(\sigma_3 + 1) + c. \tag{2.35}$$

For the Bosonic case, we have

$$H_{\rm B} = \frac{1}{2}\hbar\omega_{\rm B}\{a_{\rm B}, a_{\rm B}^+\} = \hbar\omega_{\rm B}\left(N_{\rm B} + \frac{1}{2}\right)\cdot \tag{2.36}$$

This shifts the energy levels by $+\frac{1}{2}\hbar\omega$. By analogy, we choose $H_{\rm F}$ to be,

$$H_{\rm F} = \frac{1}{2}\hbar\omega_{\rm F}[a_{\rm F}^+, a_{\rm F}] = \hbar\omega_{\rm F}\left(N_{\rm F} - \frac{1}{2}\right)\cdot \tag{2.37}$$

With this choice, the vacuum energy may cancel between the Bosons and Fermions. We see that the total Hamiltonian becomes:

$$H = \hbar\omega_{\mathrm{B}}\left(N_{\mathrm{B}} + \frac{1}{2}\right) + \hbar\omega_{\mathrm{F}}\left(N_{\mathrm{F}} - \frac{1}{2}\right) \cdot \tag{2.38}$$

The Fermionic and Bosonic number operators both commute with the total Hamiltonian.

Consider the case when $\omega = \omega_{\mathrm{F}} = \omega_{\mathrm{B}}$, here the vacuum energy precisely vanishes.

$$E_n = \hbar\omega(n_{\mathrm{B}} + n_{\mathrm{F}}). \tag{2.39}$$

There is now a symmetry **except** for the $E = 0$ state.

Define

$$Q = \frac{1}{\sqrt{2}}(\sigma_1 p + \sigma_2 x) \tag{2.40}$$

$$H = \frac{1}{2}\{Q, Q\} = Q^2 \tag{2.41}$$

Q commutes with H.

$$\begin{aligned} Q^2 &= \frac{1}{2}\left(\sigma_1^2 p^2 + \sigma_2^2 x^2 + \sigma_1\sigma_2 px + \sigma_2\sigma_1 xp\right) && (2.42)\\ &= \frac{1}{2}\left(p^2 + x^2\right) 1_{2\times 2} + i\sigma_3(px - xp) && (2.43)\\ &= \frac{1}{2}\left(p^2 + x^2\right) 1_{2\times 2} + \frac{1}{2}\sigma_3 && (2.44)\end{aligned}$$

hence,

$$H = \frac{1}{2}\omega\left(\left(p^2 + x^2\right) 1_{2\times 2} + \sigma_3\right). \tag{2.45}$$

Q anticommutes with σ_3 and commutes with H. We will label a state by its energy E and its Fermion number, $N_{\mathrm{F}} = 0, 1,$ $|E, 0\rangle$ denotes a state with energy E and $N_{\mathrm{F}} = 0$. Thus,

$$\begin{aligned} H(Q|E,0\rangle) &= E(Q|E,0\rangle) && (2.46)\\ N_{\mathrm{F}}(Q|E,0\rangle) &= 1(Q|E,0\rangle) && (2.47)\\ Q(Q|E,0\rangle) &= E|E,0\rangle. && (2.48)\end{aligned}$$

This implies

$$Q(|E,0\rangle) = \sqrt{E}|E,1\rangle \tag{2.49}$$
$$Q(|E,1\rangle) = \sqrt{E}|E,0\rangle \tag{2.50}$$

which is valid only for $E \neq 0$.

For any state $|E, N_{\mathrm{F}}\rangle$ with $E \neq 0$ there is a state with equal energy and different N_{F} obtained by the application of Q.

The ground state $|0,0\rangle$ however is annihilated by Q and so is not necessarily paired.

Note, that there is a Q_2 such that also $H = Q_2^2$.

$$Q_2 = \frac{1}{\sqrt{2}}(\sigma_2 p - \sigma_1 x) = \frac{1}{\sqrt{2}}\sigma_2(p + i\sigma_3 x) \tag{2.51}$$

and, $\{Q_1, Q_2\} = 0$.

$$Q_1 = i\sigma_3 Q_2. \tag{2.52}$$

We can then define:

$$Q_{\pm} = \frac{1}{\sqrt{2}}(Q_1 \pm iQ_2) \tag{2.53}$$

$$H = \frac{1}{2}\{Q_+, Q_-\}\ , \qquad \{Q_+, Q_+\} = 0\ , \ \ \{Q_-, Q_-\} = 0. \tag{2.54}$$

This was a "free" theory, (a simple harmonic oscillator). It can be generalized to more complicated cases. We introduce new supercharges that depend on a potential $W(x)$ as follows:

$$Q_1 = \frac{1}{\sqrt{2}}(\sigma_1 p + \sigma_2 W'(x)) \tag{2.55}$$
$$Q_2 = \frac{1}{\sqrt{2}}(\sigma_2 p - \sigma_1 W'(x)). \tag{2.56}$$

In the previous case, with the simple harmonic oscillator, $W(x) = \frac{1}{2}x^2$. The generalized Hamiltonian is now:

$$H = Q_1^2 + Q_2^2 = \frac{1}{2}(p^2 + W'(x)^2)1 + \frac{1}{2}W''(x)\sigma_3 \tag{2.57}$$
$$[H, Q_i] = 0\ , \ \ [H, \sigma_3] = 0\ , \ \ \{Q_i, \sigma_3\} = 0. \tag{2.58}$$

Thus $E \geq 0$ and for $E > 0$ the spectrum is paired.

There is an analogue in field theory. The energy gap in the Bosonic sector ($N_{\rm F}=0$) matches the energy gap in the Fermionic ($N_{\rm F}=1$) sector. In field theory the energy gap between he first excited state and the ground state is the particle mass, thus the mass of a free Boson equals that of a free Fermion.

Consider the $E=0$ case.

In general one may have any number of zero energy states in each $N_{\rm F}$ sector. For $W(x)=\frac{1}{2}x^2$, $n(E=0,N_{\rm F}=0)=1$ and $n(E=0,N_{\rm F}=1)=0$ where n denotes the number of states of given E and $N_{\rm F}$. The full spectrum can't be solved for general potential $W(x)$. It is necessary to solve two 2nd order equations:

$$\begin{pmatrix} p^2+(W')^2+W'' & 0 \\ 0 & p^2+(W')^2-W'' \end{pmatrix}\begin{pmatrix}\phi_1(x)\\ \phi_0(x)\end{pmatrix}=2E\begin{pmatrix}\phi_1(x)\\ \phi_0(x)\end{pmatrix}. \tag{2.59}$$

The zero energy solutions on the other hand, can be found (if they exist) for any $W(x)$ by using the following:

$$H\phi=E\phi\,,\quad H=Q^2. \tag{2.60}$$

So

$$H\phi=0\Leftrightarrow Q\phi=0. \tag{2.61}$$

Proof:

$$0=\,\langle 0|H|0\rangle\,=\,\langle 0|QQ|0\rangle\,=(||Q|0\rangle||)\Rightarrow Q|0\rangle\,=0. \tag{2.62}$$

Where we have used that Q is Hermitian. One can now solve the first order equation to find the zero energy states,

$$Q\phi=0 \tag{2.63}$$

$$\frac{1}{\sqrt{2}}(\sigma_1 p+\sigma_2 W'(x))\phi=0. \tag{2.64}$$

This leads to two independent first order differential equations:

$$\left(-\frac{\mathrm{d}}{\mathrm{d}x}+W'(x)\right)\phi_1(x)\quad=\quad 0 \tag{2.65}$$

$$\left(\frac{\mathrm{d}}{\mathrm{d}x}+W'(x)\right)\phi_0(x)\quad=\quad 0. \tag{2.66}$$

These can be integrated to give the following solutions:

$$\phi_1(x)\quad=\quad\phi_1(0)\exp(-W(0))\exp(W(x)) \tag{2.67}$$

$$\phi_0(x)\quad=\quad\phi_0(0)\exp(W(0))\exp(-W(x)). \tag{2.68}$$

A zero energy solution always exists. Is it a physically acceptable solution? If $W(x) \to \infty$ as $|x| \to \infty$ then $\phi_0(x)$ is a normalizable solution and one must set $\phi_1(x) = 0$. This corresponds to a Bosonic ground state. If on the other hand $W(x) \to -\infty$ as $|x| \to \infty$ then $\phi_1(x)$ is a normalizable solution and one must set ϕ_0 to zero and obtain a Fermionic ground state. For $W(x) \to$ const. as $|x| \to \infty$ then both solutions are possible. There is at most one unpaired solution at $E = 0$. For $W(x) = \sum_{n=0}^{N} a_n x^n$ there is an $E = 0$ solution if N is even. There is no $E = 0$ solution if N is odd. Only the leading asymptotic behavior matters. The fact that $n_{E=0}$ depends on such global information is very useful. It will appear again in a variety of contexts.

2.1 Symmetry and symmetry breaking

One can break a symmetry explicitly, as it occurs in the Zeeman effect where one introduces a magnetic field in the action that does not possess all the symmetries of the original action. The symmetry may be broken spontaneously. This occurs when the ground state does not posses the symmetry of the action. Spontaneous symmetry breaking plays an important role in the standard model and in a variety of statistical mechanical systems.

Assume S is a symmetry of the Hamiltonian, $[H, S] = 0$. Let $|G.S.\rangle$ denote the energy ground state of the system. If the ground state is not invariant under S, that is, if $S|G.S\rangle \neq |G.S.\rangle$ then it is said that the symmetry is spontaneously broken. When the symmetry is continuous, it has generators, g and the symmetry is spontaneously broken if $g|G.S.\rangle \neq 0$. In the case of a continuous symmetry, such a symmetry breaking may result in massless particles (called Goldstone Bosons).

Returning to our model, Q is the generator of a continuous symmetry (supersymmetry). The relevant question is whether $Q|G.S.\rangle = 0$, if $Q|0\rangle = 0$ then there is a supersymmetric ground state and spontaneous supersymmetry breaking does not occur.

$Q|G.S.\rangle \neq 0$ iff $E_{G.S.} \neq 0$ and $Q|G.S.\rangle = 0$ iff $E_{G.S.} = 0$. For the previously considered potential, $W(x) = \sum_i^N a_i x^i$, one sees that if N is even then there is no spontaneous symmetry breaking since there are zero energy normalizable solutions and conversely if N is odd then there is spontaneous symmetry breaking.

The consequences on the spectrum are as follows. When there is no spontaneous supersymmetry breaking Fermions and Bosons have the same mass. If supersymmetry is spontaneously broken then the ground state will not have zero energy. Previously it was shown that any state with non-zero energy must be paired with at least one other state (its supersymmetric partner). This implies there will be at least two ground states with identical

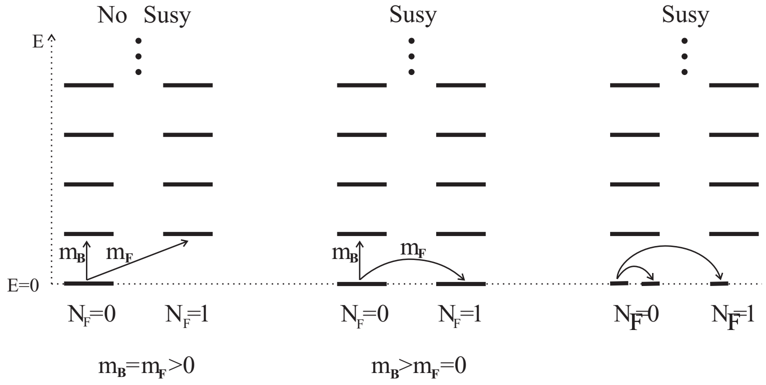


Fig. 1. Spectra.

energies and one must make a choice for the vacuum. For any choice of ground state, there is a zero energy gap between that ground state and the other possible ground state with Fermion number different by one. For example, taking the ground state to be the state with Fermion number zero implies that there is a Fermionic state with the same energy as the ground state thus the Fermion mass will now be zero while the Boson will remain with mass Δ. After spontaneous symmetry breaking the Boson and Fermion masses are no longer equal. The zero mass Fermion is called the Goldstino. It arises from the breaking of a Fermionic symmetry just as the Goldstone Boson arises from breaking a Bosonic symmetry. The Goldstone Boson and the Goldstino have special low energy couplings.

2.2 A nonrenormalisation theorem

Consider now the system from a perturbative perspective. The Hamiltonian is

$$H = \frac{1}{2}\left(-\hbar^2 \frac{\partial^2}{\partial x^2} + W'(x)^2\right) 1 + \frac{1}{2}\sigma_3 W''(x). \tag{2.69}$$

First we will do a classical analysis and then we will examine the $\hbar$ corrections to the classical result. The classical limit is taken by $\hbar \to 0$.

$$H \to \frac{1}{2}(W'(x))^2 \geq 0. \tag{2.70}$$

For N even $W'(x)$ has at least one zero where $W(x_0) = 0$. Classically therefore the ground state has zero energy. This the same as the exact result calculated previously by solving the Schroedinger equation! So how do classical results become exact? First let us note that in the Bosonic case, this does not work. For example, the ground state of the simple harmonic oscillator has an energy $\frac{1}{2}\hbar\omega$ above the classical ground state energy. More generally for the Bosonic Hamiltonian:

$$H_{\mathrm{B}} = \frac{1}{2}(p^2 + V(x)), \tag{2.71}$$

when x_0 is a minimum of $V(x)$ so that $E_{\mathrm{cl}} = \frac{1}{2}V(x_0)$ we have a perturbed energy given by:

$$E_{\mathrm{pert}} = E_{\mathrm{cl}} + \frac{1}{2}\hbar V''(x_0). \tag{2.72}$$

In a theory with Fermions though, due to the presence of a Yukawa coupling, there is another source for $\hbar$ correction.

This is analogous to how the zero point energies of a free oscillator cancel between the Bosons and Fermions.

Calculating the perturbative energy of the ground state for a supersymmetric theory we have:

$$E_{\mathrm{pert}}^{\mathrm{SUSY}} = \frac{1}{2}((W'(x_0))^2 + \hbar W''(x_0) - \hbar W''(x_0)). \tag{2.73}$$

Thus the Bosonic correction is canceled by the Fermionic correction and the classical result is exact.

The result is true to all orders in perturbation theory. This generalizes under some circumstances to supersymmetric field theories and is known by the name "nonrenormalisation theorem" [26]. As classical results are easier to obtain, this can be made into a very powerful tool.

Consider the case $W(x) = x^3$, N is odd and the supersymmetry is broken; there is no ground state with $E = 0$. However, $V_{\mathrm{cl}} = \frac{1}{2}(3x^2)^2$ and $E_{\mathrm{cl}}^{G.S.} = 0$. In this case the classical result is not exact. There are nonperturbative effects that provide corrections.

In order to see how nonperturbative effects become relevant we will actually consider the following:

$$W(x) = \frac{1}{3}x^3 + ax \qquad a < 0. \tag{2.74}$$

This also has $E = 0$ classical solutions and the leading term in $W(x)$ is odd thus there are no exact quantum $E = 0$ states; the potential is:

$$|W(x)'|^2 = (x^2 + a)^2. \tag{2.75}$$

This potential clearly has a vacuum degeneracy. We will label the vacuum states $|+\rangle, |-\rangle$. Perturbatively, the energies of these two states are equal.

The energy eigenstates are:

$$|S\rangle = \frac{1}{\sqrt{2}}(|+\rangle + |-\rangle) \tag{2.76}$$

$$|A\rangle = \frac{}{\sqrt{2}}(|+\rangle - |-\rangle) \tag{2.77}$$

$E(|A\rangle) - E(|S\rangle) > 0$ and the ground state is the symmetric state; its energy gap is:

$$\Delta E = a \ \exp\left(-\frac{c}{\lambda}\right) \tag{2.78}$$

a, c are numerical coefficients. This is a tunneling phenomenon (essentially a nonperturbative effect) that is associated with the presence of instantons.

Note that for $a > 0$ there are no zero energy solutions already classically.

2.3 A two variable realization and flat potentials

Consider a two variable realization of supersymmetry. The supercharge is given by:

$$Q = \sum_{\alpha=1}^{2} \psi_\alpha^+ \left(-p_\alpha + i\frac{\partial W}{\partial x_\alpha}\right) \tag{2.79}$$

where $W(x, y)$ is a general function of the Bosonic variables $x,\ y$. One can realize ψ_1 and ψ_2 by the following matrices:

$$\psi_1 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{2.80}$$

$$\psi_2 = \begin{pmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{2.81}$$

The Hamiltonian has the following structure:

$$H = \begin{pmatrix} H_0 & 0 & 0 & 0 \\ 0 & H_{11} & H_{12} & 0 \\ 0 & H_{21} & H_{22} & 0 \\ 0 & 0 & 0 & H_2 \end{pmatrix}. \tag{2.82}$$

The possible states are:

$$
\begin{aligned}
(1,1) &= a^{+}_{\mathrm{F}_1} a^{+}_{\mathrm{F}_2}|00\rangle && (2.83)\\
(0,1);(1,0) &= a^{+}_{\mathrm{F}_1}|00\rangle, a^{+}_{\mathrm{F}_2}|00\rangle && (2.84)\\
(0,0) &= |00\rangle. && (2.85)
\end{aligned}
$$

With,

$$
\begin{aligned}
H_0 &= \frac{1}{2}[-\Delta + (\nabla W)^2 - \Delta W] && (2.86)\\
H_2 &= \frac{1}{2}[-\Delta + (\nabla W)^2 + \Delta W] && (2.87)
\end{aligned}
$$

$$
H_1 = \frac{1}{2}\left((-\Delta + (\nabla W)^2 - \Delta W)1 + 2\begin{pmatrix} \partial_{11}W & \partial_{12}W \\ \partial_{21}W & \partial_{22}W \end{pmatrix}\right)\cdot \qquad (2.88)
$$

For any W the $n = 0,\ 2$ sectors can be solved exactly for $E = 0$ just as before.

Consider the potential $W = x(y^2 + c)$ then $V_{\mathrm{cl}} = \frac{1}{2}[(y^2 + c)^2 2 + 4x^2y^2]$. For $c > 0$, $V_{\mathrm{cl}} = \frac{1}{2}c^2 > 0$ which leads to classical SUSY breaking [27].

Note, $V_{\mathrm{cl}} = \frac{1}{2}c^2$ at $y = 0$. This is a "flat direction", the potential is the same for all values of x. Flat directions imply the presence of many vacua. Such flat directions appear in abundance in supersymmetric models and lead in some contexts to supersymmetry breaking. Flat directions in purely Bosonic models are lifted by quantum fluctuations. In supersymmetric systems the Bosonic and Fermionic fluctuations cancel and the flat directions remain.

Classically the ground state is non zero which implies that the supersymmetry is broken. What about quantum mechanical effects, can supersymmetry actually be restored? (There are cases where symmetries have been known to be restored quantum mechanically.)

For $n = 0,\ 2$ sectors, the answer is no; neither ground state is normalisable. What about $n = 1$ sector? To find these states we need to solve the following pseudo-analytic equations:

$$
\begin{aligned}
\partial_1\left(e^{W}\phi_{1,0}\right) &= \partial_2\left(e^{W}\phi_{0,1}\right) && (2.89)\\
\partial_1\left(e^{-W}\phi_{0,1}\right) &= -\partial_2\left(e^{-W}\phi_{1,0}\right)\cdot && (2.90)
\end{aligned}
$$

Unfortunately, these cannot be solved in general. We will now show that from more general arguments that supersymmetry cannot be restored by quantum effects.

$$E^W = \frac{\langle\phi|H|\phi\rangle}{\langle\phi|\phi\rangle}\,. \tag{2.91}$$

Consider perturbing the potential,

$$W^\lambda = W^0 + \lambda x^i n^i \tag{2.92}$$

where $n^i n^i = 1$ and $W^0 = x(y^2 + c)$ is the unperturbed potential.

$$E^W(\lambda) = E^W(0) + \frac{1}{2}\lambda^2 + \lambda \int \mathrm{d}^2x(n\cdot\nabla W)\phi^2 \tag{2.93}$$

where ϕ denotes the unperturbed solution. (There is no λ dependence from the Yukawa term.) $\partial_x W = y^2 + x > 0$, $\phi^2 \geq 0$ and $\int \mathrm{d}^2x\phi^2 = 1$, thus,

$$\int \mathrm{d}^2x(n\cdot\nabla W)\phi^2 > 0. \tag{2.94}$$

So we may write:

$$E^W(\lambda) = E^W(0) + \frac{1}{2}\lambda^2 + \lambda r^2 \tag{2.95}$$

where r is some finite real quantity.

$E^{W(\lambda)} \geq 0$; if we now assume $E^{W(0)} = 0$ then we obtain a contradiction because by taking λ to be small and negative we obtain $E^{W(\lambda)} < 0$.

This argument is rather general and also works for the 3 variable potential

$$W = ayz + bx(y^2 + c) \tag{2.96}$$

(this is the potential used to break supersymmetry spontaneously in scalar field theories.)

There is a short more elegant argument for non-restoration of supersymmetry that is based on an index theorem.

The "Witten index", [25] is defined by,

$$\Delta = \mathrm{Tr}(-1)^F = \sum_E \left(n^{F=\mathrm{even}}(E) - n^{F=\mathrm{odd}}(E)\right). \tag{2.97}$$

The trace indicates a sum over all states in the Hilbert space, F is the fermion number of the state. $n^{F=\mathrm{even}}(E)$, $n^{F=\mathrm{odd}}(E)$ denote the number of solutions with Fermi number even/odd with energy E. Since the Bosons and fermions are paired at all energies greater than zero then

$$\Delta = n^{F=\text{even}}(E=0) - n^{F=\text{odd}}(E=0). \tag{2.98}$$

If Δ is calculated for some potential W it will not change under perturbations of W (only E_n will change). In particular, if $\Delta \neq 0$ for some W then there will be no spontaneous symmetry breaking under any perturbations in W.

On the other hand if $\Delta = 0$ then we do not know whether SUSY may be broken or not since either

$$n^{F=\text{even}} = n^{F=\text{odd}} \neq 0 \tag{2.99}$$

or

$$n^{F=\text{even}} = n^{F=\text{odd}} = 0. \tag{2.100}$$

Thus for $\Delta = 0$ one needs more information.

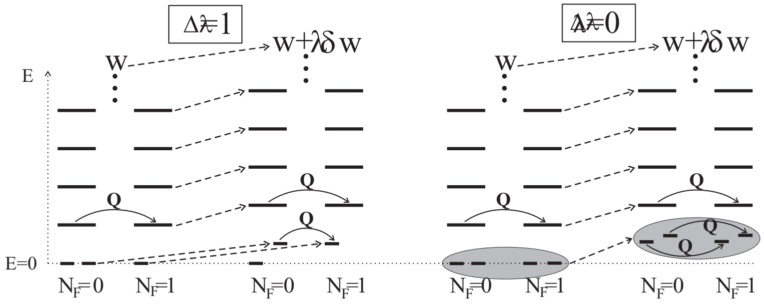


Fig. 2. Perturbing the spectrum; Δ is unchanged.

Returning to the case described above. We have calculated the classical Witten Index to be:

$$\Delta_{\text{cl}} = 0. \tag{2.101}$$

One considers turning on quantum corrections as a perturbation. Since the index is invariant under perturbations, one obtains:

$$0 = \Delta_{\text{cl}} = \Delta_{\text{quantum}} = n^q_{N_{\text{F}}=0}(E=0) + n^q_{N_{\text{F}}=2}(E=0) - n^q_{N_{\text{F}}=1}(E=0). \tag{2.102}$$

We have already shown that

$$n^q_{N_{\text{F}}=0}(E=0) = n^q_{N_{\text{F}}=2}(E=0) = 0 \tag{2.103}$$

which in turn implies

$$n_{N_{\mathrm{F}}=1}(E=0)=0\;, \tag{2.104}$$

thus, there is no supersymmetry restoration.

2.4 Geometric meaning of the Witten index

Consider a supersymmetry sigma model with a D dimensional target space whose metric is g_{ij} [25],

$$S=\int \mathrm{d}t g_{ij}(x)\frac{\mathrm{d}x^i}{\mathrm{d}t}\frac{\mathrm{d}x^i}{\mathrm{d}t}+i\bar{\psi}\gamma^0 D_{\mathrm{t}}\psi_i+\frac{1}{4}R_{ijkl}\psi^i\psi^k\psi^j\psi^l\;, \tag{2.105}$$

D_{t} denotes a covariant derivative, R_{ijkl} is the curvature.

$$\{\psi_i,\psi_j\}=\{\psi_i^*,\psi_j^*\}=0\;,\quad \{\psi_i,\psi_j^*\}=g_{ij}\;, \tag{2.106}$$

$$Q=i\sum_i^D=\phi_i^* p_i\;,\qquad Q^*=-i\sum_i^D=\phi_i^* p_i \tag{2.107}$$

and

$$p_i=-D_{x^i}. \tag{2.108}$$

The Hilbert space may be graded according to the Fermion occupation number:

$$\begin{array}{ccccccc}
\phi_0 & = & |0,..,0\rangle & \mathrm{Dim} & = & 1 \\
\phi_1 & = & \{|0,..,1,..,0\rangle\} & \mathrm{Dim} & = & D \\
\phi_2 & = & \{|0,..,1,..,1,..,0\rangle\} & \mathrm{Dim} & = & \frac{1}{2}D(D-1) \\
. & & . & & & \\
. & & . & & & \\
\phi_{\mathrm{D}} & = & |1,1,..,1\rangle & \mathrm{Dim} & = & 1.
\end{array} \tag{2.109}$$

Generically the Dimension of a state with p-Fermions is given by:

$$\mathrm{Dim}(p,D)=\frac{D!}{(D-p)!p!}\cdot \tag{2.110}$$

This is identical to the dimension of p-forms in D dimensions. Q acts by adding a Fermion hence is a map

$$Q\colon \phi_p\rightarrow\phi_{p+1}. \tag{2.111}$$

From simple Fermi statistics we see that, Q is nilpotent. Hence, we have an isomorphism between p-forms with an exterior derivative, d and the space of states with Fermion occupation number p and supercharge Q.

As with any nilpotent operator one can examine its cohomology defined by:

$$H^p = \frac{\{\mathrm{d}\phi_p = 0\}}{\{\phi_p = d\phi_{p-1}\}} \,. \tag{2.112}$$

The dimension of the cohomology is denoted by the Betti number $b^p = \dim H^p$. The Euler characteristic is then given by the alternating sum of the Betti numbers.

$$\chi = \sum_p (-1)^p b^p. \tag{2.113}$$

Recall that to find the $E = 0$ states, one solved the equation $Q\phi_p = 0$, hence the $E = 0$ states are in the cohomology of Q. Thus one has the remarkable correspondence,

$$\mathrm{Tr}(-1)^F = \chi. \tag{2.114}$$

The Euler characteristic is a topological invariant and is independent of geometrical perturbations of the manifold. This explains why the Witten index is stable against non-singular perturbations. The Witten index is given a very physical realization in the following example.

2.5 Landau levels and SUSY QM

Consider an electron moving in two dimensions in the presence of a perpendicular magnetic field.

$$H = \frac{1}{2}\left[(p_x + A_x)^2 1 + (p_y + A_y)^2 1 + B_z\sigma_3\right]. \tag{2.115}$$

Two Bosons x, y and one Fermion. This is a less familiar realization of supersymmetry as the number of Bosonic and Fermionic oscillators differ. It is particular to the quantum mechanical system. The supercharges are:

$$Q^1 = \frac{1}{\sqrt{2}}\left((p_x + A_x)\sigma_1 + (p_y + A_y)\sigma_2\right) \tag{2.116}$$

$$Q^2 = \frac{1}{\sqrt{2}}\left((p_x + A_x)\sigma_2 - (p_y + A_y)\sigma_1\right) \tag{2.117}$$

$$[Q^i, H] = 0\,, \quad \{Q^i, \sigma_3\} = 0. \tag{2.118}$$

In Lorentz gauge, $\partial_i A^i = 0$ we can write $A^i = \epsilon^{ij}\partial_j a$ $B_z = -\nabla^2 a$. Take $Q = Q^1$, Q is just like the Dirac operator.

To find $E = 0$ states we again use $Q\phi = 0$. Multiplying by σ_1 gives:

$$((p_x + A_x)1 + i\sigma_3(p_y + A_y)) \begin{pmatrix} \phi_1 \\ \phi_{-1} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{2.119}$$

One can solve these equations as follows:

$$\phi_1(x, y) = \exp(-a) f(x + iy)\ , \quad \phi_{-1} = \exp(a) g(x - iy) \tag{2.120}$$

where f and g are arbitrary functions. The normalisability of these solutions depends on the function a.

In a constant magnetic field, taking for example, $a = -\frac{1}{2}y^2 B$. Depending on the sign of B, either ϕ_1 or ϕ_{-1} may be normalizable. If $B > 0$, take $\phi_1 = 0$ and

$$\phi_{-1} = \exp\left(-\frac{1}{2}y^2 B\right) g(x - iy). \tag{2.121}$$

From translational invariance, a convenient choice is

$$g(x - iy) = \exp(ik(x - iy)). \tag{2.122}$$

There are an infinite number of $E = 0$ states. (This is true also for any E not just $E \neq 0$.)

Let us examine this from a topological point of view. The total magnetic flux is:

$$\Phi = \frac{1}{2\pi}\int \mathrm{d}^2 r B_z = -\frac{1}{2\pi}\int_0^{2\pi} \mathrm{d}\theta r \partial_{\mathrm{r}} a \mid_{r=\infty}. \tag{2.123}$$

$\Phi = 0$ implies $a(r = \infty) = 0$ and one has only plane wave normalizable states. Take a constant negative B field with $a = -\frac{1}{4}r^2 B$. The solution is:

$$\phi_{+1} = \frac{c}{r^{|\Phi|}} r^n \exp(in\theta). \tag{2.124}$$

For B constant $\Phi \to \infty$ and there are an infinite number of allowed states, all n. Now assume that Φ is finite because B is confined in a solenoid, for example. In such a case, normalisability requires that n is bounded by $[\Phi - 1]$. (Φ is quantized). Thus the number of $E = 0$ states is given by the total flux number,

$$n_{(E=0)} = |\Phi|. \tag{2.125}$$

The total magnetic flux is a global quantity that is independent of local fluctuations. Again we have shown how the number of zero energy states does not depend on the local details but only upon information.

2.6 Conformal quantum mechanics

Relevant material for this section may be found in [28–30]. Recall the Hamiltonian:

$$H = \frac{1}{2}(p^2 + gx^{-2}) \tag{2.126}$$

is special since g has a meaning. H is part of an algebra:

$$[H, D] = iH\ , \quad [K, D] = iK\ , \quad [H, K] = 2iD. \tag{2.127}$$

This forms an $SO(2{,}1)$ algebra where:

$$D = -\frac{1}{4}(xp + px)\ , \quad K = \frac{1}{2}x^2 \tag{2.128}$$

and H is as given above. The Casimir is given by:

$$\frac{1}{2}(HK + KH) - D^2 = \frac{g}{4} - \frac{3}{16}\cdot \tag{2.129}$$

The meaning of D and K is perhaps clearer in the Lagrangian formalism:

$$\mathcal{L} = \frac{1}{2}\left(\dot{x}^2 - \frac{g}{x^2}\right)\ , \quad S = \int \mathrm{d}t\mathcal{L}. \tag{2.130}$$

Symmetries of the action S, and not the Lagrangian $\mathcal{L}$ alone, are given by:

$$t' = \frac{at + b}{ct + d}\ , \quad x'(t') = \frac{1}{ct + d}x(t) \tag{2.131}$$

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}\ , \qquad \det A = ad - bc = 1 \tag{2.132}$$

H acts as translation

$$A_T = \begin{pmatrix} 1 & 0 \\ \delta & 1 \end{pmatrix}\ , \qquad t' = t + \delta \tag{2.133}$$

D acts as dilation

$$A_D = \begin{pmatrix} \alpha & 0 \\ 0 & \frac{1}{\alpha} \end{pmatrix}\ , \qquad t' = \alpha^2 t \tag{2.134}$$

K acts as a special conformal transformation

$$A_K = \begin{pmatrix} 1 & \delta \\ 0 & 1 \end{pmatrix}\ , \qquad t' = \frac{t}{\delta t + 1}\cdot \tag{2.135}$$

The spectrum of the Hamiltonian 2.126 is the open set $(0, \infty)$, the spectrum is therefore continuous and bounded from below. The wave functions are given by:

$$\psi_E(x) = \sqrt{x} J_{\sqrt{g+\frac{1}{4}}}(\sqrt{2E}x) \qquad E \neq 0. \tag{2.136}$$

We will now attempt to find the zero energy state. Take the ansatz $\phi(x) = x^{\alpha}$:

$$H = \left(-\frac{\mathrm{d}^2}{\mathrm{d}x^2} + \frac{g}{x^2}\right) x^{\alpha} = 0. \tag{2.137}$$

This implies

$$g = -\alpha(\alpha - 1) \tag{2.138}$$

solving this equation gives

$$\alpha = -\frac{1}{2} \pm \frac{\sqrt{1+4g}}{2} \,. \tag{2.139}$$

This gives two independent solutions and by completeness all the solutions. $\alpha_+ > 0$, does not lead to a normalizable solution since the function diverges at infinity. $\alpha_- < 0$, is not normalizable either since the function diverges at the origin (a result of the scale symmetry).

Thus, there is no normalizable $E = 0$ solution (not even plane wave normalizable)!

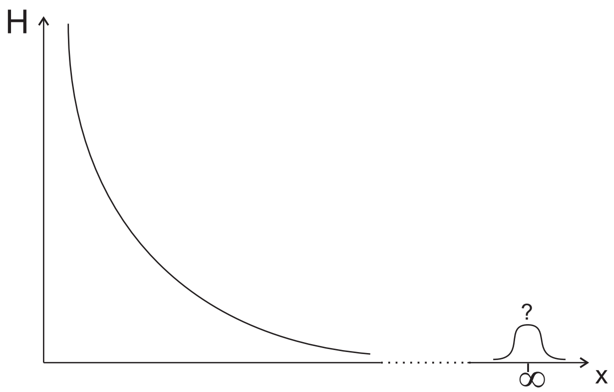


Fig. 3. The absence of a normalisable ground state for this potential.

Most of the analysis in field theory proceeds by identifying a ground state and the fluctuations around it. How do we deal with a system in the absence of a ground state?

One possibility is to accept this as a fact of life. Perhaps it is possible to view this as similar to cosmological models that also lack a ground state such those with Quintessence.

Another possibility is to define a new evolution operator that does have a ground state.

$$G = uH + vD + wK. \tag{2.140}$$

This operator has a ground state if $v^2 - 4uw < 0$. Any choice explicitly breaks Lorentz and scale invariance. Take for example,

$$G = \frac{1}{2}\left(\frac{1}{a}K + aH\right) \equiv R \tag{2.141}$$

a has dimension of length. The eigenvalues of R are

$$r_n = r_0 + n\ , \qquad r_0 = \frac{1}{2}\left(1 + \sqrt{g + \frac{1}{4}}\right) \cdot \tag{2.142}$$

One must interpret what this means physically. Surprisingly this question arises in the context of black hole physics.

Consider a particle of mass m and charge q falling into a charged black hole. The black hole is BPS meaning that its mass, M and charge, Q are related by $M = Q$. The blackhole metric is given by:

$$\mathrm{d}s^2 = -\left(1 + \frac{M}{r}\right)^{-2} \mathrm{d}t^2 + \left(1 + \frac{M}{r}\right)^2 (\mathrm{d}r^2 + r^2 \mathrm{d}\Omega^2)\ , \qquad A_{\mathrm{t}} = \frac{r}{M} \cdot \tag{2.143}$$

Now consider the near Horizon limit $r << M$, which we will take by $M \to \infty$ keeping r fixed.

$$\mathrm{d}s^2 = -\left(\frac{r}{M}\right)^2 \mathrm{d}t^2 + \left(\frac{M}{r}\right)^2 \mathrm{d}r^2 + M^2 \mathrm{d}\Omega^2. \tag{2.144}$$

This produces an $AdS_2 \times S^2$ geometry.

We also wish to keep fixed $M^2(m - q)$ as we scale M. This means we must scale $(m - q) \to 0$. That is the particle itself becomes BPS in the limit. The Hamiltonian for this in falling particle in this limit is given by our old friend:

$$H = \frac{p_{\mathrm{r}}^2}{2m} + \frac{g}{2r^2}\ , \qquad g = 8M^2(m - q) + \frac{4l(l+1)}{M} \cdot \tag{2.145}$$

For $l = 0$, we have $g > 0$ and there is no ground state. This is associated with the coordinate singularity at the Horizon. The change in evolution operator is now associated with a change of time coordinate. One for which the world line of a static particle passes through the black hole horizon instead of remaining in the exterior of the space time.

2.7 Superconformal quantum mechanics

The bosonic conformal mechanical system had no ground state. The absence of a $E = 0$ ground state in the supersymmetric context leads to the breaking of supersymmetry. This breaking has a different flavor from that which was discussed earlier. Next the supersymmetric version of conformal quantum mechanics is examined to see if supersymmetry is indeed broken. For this the superpotential is chosen to be,

$$W(x) = \frac{1}{2}g\ \log x^2 \ , \tag{2.146}$$

yielding a Hamiltonian:

$$H = \frac{1}{2}\left(\left(p^2 + \left(\frac{\mathrm{d}w}{\mathrm{d}x}\right)^2\right) 1 - \sigma_3 \frac{\mathrm{d}^2 W}{\mathrm{d}x^2}\right) \cdot \tag{2.147}$$

Representing ψ by$\frac{1}{2}\sigma_-$ and $\psi^*$by $\frac{1}{2}\sigma_+$ gives the supercharges:

$$Q = \psi^+\left(-ip + \frac{\mathrm{d}W}{\mathrm{d}x}\right)\ ,\ \ Q^+ = \psi\left(ip + \frac{\mathrm{d}w}{\mathrm{d}x}\right) \cdot \tag{2.148}$$

One now has a larger algebra, the superconformal algebra,

$$\{Q, Q^+\} \ = \ 2H\ ,\ \ \{Q, S^+\} = g - B + 2iD \tag{2.149}$$

$$\{S, S^+\} \ = \ 2K\ ,\ \ \{Q^+, S\} = g - B - 2iD. \tag{2.150}$$

A realization is:

$$B = \sigma_3\ ,\ \ S = \psi^+ x\ ,\ \ S^+ = \psi x. \tag{2.151}$$

The zero energy solutions are

$$\exp(\pm W(x)) = x^{\pm g}\ , \tag{2.152}$$

neither solution is normalizable.

H factorizes:

$$2H = \begin{pmatrix} p^2 + \frac{g(g+1)}{x^2} & 0 \\ 0 & p^2 + \frac{g(g-1)}{x^2} \end{pmatrix} \tag{2.153}$$

and we may solve for the full spectrum:

$$\psi_E(x) = x^{\frac{1}{2}} J_{\sqrt{\nu}}(x\sqrt{2E})\ , \qquad E \neq 0\ , \tag{2.154}$$

where $\nu = g(g-1) + \frac{1}{4}$ for $N_{\mathrm{F}} = 0$ and $\nu = g(g+1) + \frac{1}{4}$ for $N_{\mathrm{F}} = 1$. Again the spectrum is continuous though there is no normalizable zero energy state. Again we must interpret the absence of a normalizable ground state. One can again define a new operator with a normalizable ground state. By inspection the operator (2.141) can be used provided one makes the following identifications:

$$N_{\mathrm{F}} = 1 \qquad g_{\mathrm{B}} = g_{\mathrm{susy}}(g_{\mathrm{susy}}+1)\,, \tag{2.155}$$

$$N_{\mathrm{F}} = 0 \qquad g_{\mathrm{B}} = g_{\mathrm{susy}}(g_{\mathrm{susy}}-1). \tag{2.156}$$

Thus the spectrum differs between the $N_{\mathrm{F}} = 1$ and $N_{\mathrm{F}} = 0$ sectors and so supersymmetry would be broken. One needs to define a whole new set of operators:

$$M = Q - S \qquad M^+ = Q^+ - S^+ \tag{2.157}$$

$$N = Q^+ + S^+ \qquad N^+ = Q + S^+ \tag{2.158}$$

which produces the algebra:

$$\frac{1}{4}\{M, M^+\} = R + \frac{1}{2}B - \frac{1}{2}g \equiv T_1 \tag{2.159}$$

$$\frac{1}{4}\{N, N^+\} = R + \frac{1}{2}B + \frac{1}{2}g \equiv T_2 \tag{2.160}$$

$$\frac{1}{4}\{M, N\} = L_- \qquad \frac{1}{4}\{M^+, N^+\} = L_+ \tag{2.161}$$

$$L_\pm = -\frac{1}{2}(H - K \mp 2iD) \tag{2.162}$$

T_1, T_2, H have a doublet spectra. “Ground states” are given by:

$$T_1|0\rangle = 0\,; \quad T_2|0\rangle = 0\,; \quad H|0\rangle = 0\ . \tag{2.163}$$

A physical context arises when one considers a supersymmetric particle falling into a black hole [31,32]. This is the supersymmetric analogue of the situation already discussed.

3 Review of supersymmetric models

3.1 Kinematics

For a more detailed discussion of the material presented in this section see [3]. The possible particle content of supersymmetric (SUSY) theories is determined by the SUSY algebra, its $N = 1$ version is:

$$\{Q_\alpha, \bar{Q}_{\dot\alpha}\} = 2\sigma^\mu_{\alpha\dot\beta} P_\mu\,, \quad \{Q_\alpha, Q_\beta\} = \{\bar{Q}_{\dot\alpha}, \bar{Q}_{\dot\beta}\} = 0. \tag{3.1}$$

$$[P_\mu, Q_\alpha] = [P_\mu, \bar{Q}_{\dot\alpha}] = [P_\mu, P_\nu] = 0. \tag{3.2}$$

This algebra can be generalized to include a higher number of supersymmetries $N = 2, 4$ by:

$$\{Q^i_\alpha, \bar{Q}^j{}_{\dot{\alpha}}\} = 2\delta^{ij}\sigma^\mu_{\alpha\dot{\beta}}P_\mu + \delta_{\alpha\dot{\beta}}U_{ij} + (\gamma_5)_{\alpha\dot{\beta}}V_{ij} \ , \tag{3.3}$$

U and V are the central charges they commute with all other charges (they are antisymmetric in ij). They are associated with BPS states such as monopoles. We will discuss in this section the $d = 4$ realisations with, $\mu, \nu = 0, 1, 2, 3$ the space-time indices. In four dimensions we have two component Weyl Fermions. Those with α or β indices transform under the $(0, \frac{1}{2})$ representation of the Lorentz group; and those with dotted indices, $\dot{\alpha}$ or $\dot{\beta}$ transform under the $(\frac{1}{2}, 0)$ representation.

Consider first the massless representations of $N = 1$ supersymmetry. The simplest is the the chiral multiplet:

$$\left(-\frac{1}{2}, 0, 0, \frac{1}{2}\right) \qquad (\phi, \psi) \qquad (2, 2). \tag{3.4}$$

In the above table, first are written the helicities; then the associated component fields, ϕ denotes a complex scalar and ψ a Weyl Fermion; and finally are the number of physical degrees of freedom carried by the Bosons and Fermions. The vector multiplet has:

$$\left(-1, -\frac{1}{2}, \frac{1}{2}, 1\right) \qquad (\lambda_\alpha, A_\mu) \qquad (2, 2) \tag{3.5}$$

λ is a Weyl Fermion and A_μ is a vector field. With $N = 2$ supersymmetry, there is a massless vector multiplet:

$$\begin{pmatrix} -1 \begin{matrix} -\frac{1}{2} \\ -\frac{1}{2} \end{matrix} \begin{matrix} 0 \\ 0 \end{matrix} \begin{matrix} \frac{1}{2} \\ \frac{1}{2} \end{matrix} 1 \end{pmatrix} \qquad (\phi, \psi) + (\lambda_\alpha, A_\mu) \qquad (4, 4) \tag{3.6}$$

and a massless hypermultiplet which is given by:

$$\begin{pmatrix} \begin{matrix} -\frac{1}{2} \\ -\frac{1}{2} \end{matrix}, \begin{matrix} 0 \\ 0 \\ 0 \\ 0 \end{matrix}, \begin{matrix} \frac{1}{2} \\ \frac{1}{2} \end{matrix} \end{pmatrix} \qquad (\phi_1, \psi_1) + (\phi_2, \psi_2) \qquad (4, 4). \tag{3.7}$$

For massive multiplets, in $N = 1$, there is again the chiral multiplet which is the same as the massless multiplet but with now massive fields. The massive vector multiplet becomes:

$$\begin{pmatrix} -1 \begin{matrix} -\frac{1}{2} \\ -\frac{1}{2} \end{matrix} \begin{matrix} 0 \\ 0 \end{matrix} \begin{matrix} \frac{1}{2} \\ \frac{1}{2} \end{matrix} 1 \end{pmatrix} \qquad (h, \psi_\alpha, \lambda_\alpha, A_\mu) \qquad (4, 4). \tag{3.8}$$

Where h is a real scalar field. The massive vector multiplet has a different field content than the massless vector multiplet because a massive vector field has an additional physical degree of freedom. One sees that the massive vector multiplet is composed out of a massless chiral plus massless vector multiplet. This can occur dynamically; massive vector multiplets may appear by a supersymmetric analogue of the Higgs mechanism. With $N = 4$ supersymmetry, the massless vector multiplet is:

$$\begin{pmatrix} & & 0 & & \\ & -\frac{1}{2} & 0 & \frac{1}{2} & \\ & -\frac{1}{2} & 0 & \frac{1}{2} & \\ -1\,, & -\frac{1}{2}\,, & 0\,, & \frac{1}{2}\,, & 1 \\ & -\frac{1}{2} & 0 & \frac{1}{2} & \\ & & 0 & & \end{pmatrix} \qquad (\lambda^a, \phi^I, A_\mu) \qquad (8,8) \tag{3.9}$$

where $I = 1..6,\ a = 1..4$.

3.2 Superspace and chiral fields

Spacetime can be extended to include Grassmann spinor coordinates, $\bar\theta_{\dot\alpha}, \theta_\alpha$. Superfields are then functions of this superspace. Constructing a Lagrangian out of superfields provides a useful way to construct explicitly supersymmetric Lagrangians. Recall the integration formulas for Grassmann variables:

$$\int \mathrm{d}\theta_\alpha \theta_\alpha = \frac{\partial}{\partial\theta_\alpha} = 1\,, \quad \int \mathrm{d}\theta_\alpha = 0. \tag{3.10}$$

The following identity will be of use:

$$\int \mathrm{d}^2\theta \mathrm{d}^2\bar\theta \mathcal{L} = \int \mathrm{d}^2\theta \frac{\partial^2 \mathcal{L}}{\partial\bar\theta_1 \bar\theta_2}\,\cdot \tag{3.11}$$

The supercharges can be realized in superspace by generators of supertranslations:

$$Q_\alpha = \frac{\partial}{\partial\theta_\alpha} - i\sigma^\mu_{\alpha\dot\alpha}\bar\theta^{\dot\alpha}\partial_\mu\,, \qquad \bar Q_{\dot\alpha} = -\frac{\partial}{\partial\bar\theta_{\dot\alpha}} + i\theta^\alpha\sigma^\mu_{\alpha\dot\alpha}\partial_\mu. \tag{3.12}$$

It will also be useful to define a supercovariant derivative:

$$D_\alpha = \frac{\partial}{\partial\theta_\alpha} + i\sigma^\mu_{\alpha\dot\alpha}\bar\theta^{\dot\alpha}\partial_\mu\,, \qquad \bar D_{\dot\alpha} = -\frac{\partial}{\partial\bar\theta_{\dot\alpha}} - i\theta^\alpha\sigma^\mu_{\alpha\dot\alpha}\partial_\mu. \tag{3.13}$$

A superfield Φ is called "chiral" if:

$$\bar D_{\dot\alpha}\Phi = 0. \tag{3.14}$$

One introduces the variable,

$$y^\mu = x^\mu + i\theta\sigma^\mu\bar{\theta} \tag{3.15}$$

which produces the following expansion for a chiral field,

$$\Phi(y) = A(y) + \sqrt{2}\theta\psi(y) + \theta\theta F(y). \tag{3.16}$$

The Taylor expansion terminates because of the anticommuting property of the Grassmann coordinates. As a function of x it may be written as follows:

$$\Phi(x) = A(x) + i\theta\sigma^\mu\bar{\theta}\partial_\mu A(x) + \frac{1}{4}\theta\theta\bar{\theta}\bar{\theta}\Box A(x) + \sqrt{2}\theta\psi(x) \tag{3.17}$$

$$-\frac{i}{\sqrt{2}}\theta\theta\partial_\mu\psi(x)\sigma^\mu\bar{\theta} + \theta\theta F(x). \tag{3.18}$$

The key point is that

$$\mathcal{L} = \int \mathrm{d}^2\theta\Phi(x) \tag{3.19}$$

is a invariant under supersymmetric transformations (up to a total derivative).

After the integration some terms will disappear from the expansion of $\Phi(x)$ leaving only:

$$\Phi(x) = A(x) + \sqrt{2}\theta\psi(x) + \theta\theta F(x) \tag{3.20}$$

$A(x)$ will be associate with a complex Boson; $\psi(x)$ will be associated with a Weyl Fermion and $F(x)$ acts as an auxiliary field that contributes no physical degrees of freedom. These are called the component fields of the superfield. The product of two chiral fields also produces a chiral field. Therefore, any polynomial, $W(\Phi)$ can be used to construct a supersymmetry invariant as

$$\mathcal{L} = \int \mathrm{d}^2\theta W(\Phi) = F_{W(\Phi)} \tag{3.21}$$

is a supersymmetry invariant. This is used to provide a potential for the chiral field. The kinetic terms are described by:

$$\int \mathrm{d}^2\theta\mathrm{d}^2\bar{\theta}\bar{\Phi}_i\Phi_j = \bar{\Phi}_i\Phi_j\ |_{\theta\theta\bar{\theta}\bar{\theta}}\ . \tag{3.22}$$

After expanding and extracting the $\theta\theta\bar{\theta}\bar{\theta}$ term is (up to total derivatives):

$$F_i^* F_{\mathrm{f}} - |\partial_\mu A|^2 + \frac{i}{2}\partial_\mu\bar{\psi}\bar{\sigma}^\mu\psi. \tag{3.23}$$

One thus composes the following Lagrangian:

$$\begin{aligned} \mathcal{L} &= \bar{\Phi}_i\Phi_i \mid_{\theta\theta\bar{\theta}\bar{\theta}} + \left[\lambda_i\Phi_i + \frac{1}{2}m_{ij}\Phi_i\Phi_j + \frac{1}{3}g_{ijk}\Phi_i\Phi_j\Phi_k\right]_{\theta\theta} & (3.24)\\ &= i\partial\bar{\psi}_i\bar{\sigma}\psi_i + A_i^*\Box A_i + F_i^*F_i + \lambda_i F_i + m_{ij}\left(A_iF_j - \frac{1}{2}\psi_i\psi_j\right) \\ &\quad + g_{ijk}(A_iA_jF_k - \psi_i\psi_jA_k) + h.c. & (3.25)\end{aligned}$$

One must now eliminate the auxiliary fields F_i, F_i^*. The equation of motion for F_k^* is as follows:

$$F_k + \lambda_k^* + m_{ij}A_i^* + g_{ijk}^*A_i^*A_j^*. \tag{3.26}$$

This gives:

$$\begin{aligned} \mathcal{L} &= i\partial\bar{\psi}_i\bar{\sigma}\psi_i + A_i^*\Box A_i - \frac{1}{2}m_{ij}\psi_i\psi_j - \frac{1}{2}m_{ij}^*\psi_i^*\psi_j^* \\ &\quad - g_{ijk}\psi_i\psi_jA_k - g_{ijk}^*\bar{\psi}_i\bar{\psi}_jA_k^* - F_i^*F_i & (3.27)\end{aligned}$$

where the last term is a potential for A, A^*; these are known as the F terms, $V_{\mathrm{F}}(A^*, A)$. (Note, $V_{\mathrm{F}} \geq 0$). At the ground state this must vanish $V_{\mathrm{F}}(A^*, A) = 0$. This in turn implies that $F_i = 0$ for the ground state. Although this is a classical analysis so far, in fact it is true to all orders in perturbation theory as there exists a non renormalization theorem for the effective potential.

3.3 Kähler potentials

To describe the supersymmetric Lagrangian for scalar fields spanning a more complicated manifold it is convenient to introduce the following supersymmetry invariant:

$$\int \mathrm{d}^4\theta K(\Phi, \bar{\Phi}). \tag{3.28}$$

$K(\Phi, \bar{\Phi})$ is called the Kähler potential. One may add any function of Φ or $\bar{\Phi}$ to the integrand since these terms will vanish after integration. For the usual kinetic terms, K is taken to be given by $K = \Phi\bar{\Phi}$ which produces the $-\delta_{ij}\partial_\mu A^{*i}\partial^\mu A^j$ kinetic terms for the scalars. For the case of a sigma model with a target space whose metric is g_{ij}; this metric is related to the Kähler potential by:

$$g_{ij} = \frac{\partial^2 K}{\partial\bar{\Phi}_i\partial\Phi_j} \,\cdot \tag{3.29}$$

The above supersymmetry invariant (3.28) which previously gave the usual kinetic terms in the action, produces for general K the action of a supersymmetric sigma model, with the target space metric given by equation (3.29).

3.4 F-terms

In this section we examine the vanishing of the potential generated by the F terms. The issues we are interested in are whether supersymmetry is spontaneously broken; is there a non renormalization theorem; and are there other internal symmetries broken.

$$V_{\mathrm{F}} = 0 \Leftrightarrow F_i = 0 \qquad \forall i\ , \tag{3.30}$$

are n (complex) equations with n (complex) unknowns. Generically, they have a solution. Take the example of the one component WZ model, where

$$F_1 = -\lambda - mA + gA^2. \tag{3.31}$$

This has a solution. There is no supersymmetry breaking classically. The Witten index $\mathrm{Tr}(-1)^F = 2$. This implies the classical result is exact. Note, that for $\lambda = 0$,

$$V = A^*A|(gA - m)|^2 \tag{3.32}$$

and hence there will be a choice of vacuum: either $\langle A\rangle = \frac{m}{g}$ or $\langle A\rangle = 0$.

The claim is that the form of the effective interacting superpotential W_{eff} will be the same as the classical superpotential W_{cl}. Take W_{cl} to have the form:

$$W = \frac{1}{2}m\Phi^2 + \frac{1}{3}\lambda\Phi^3. \tag{3.33}$$

To show that the form of W remains invariant, the following ingredients are used: the holomorphic dependence of $W(\Phi, m, \lambda)$, that is W is independent of m^*, λ^*, Φ^*; and the global symmetries present in the theory [33].

3.5 Global symmetries

R-symmetry is a global $U(1)$ symmetry that does not commute with the supersymmetry. The action of the R-symmetry on a superfield Φ with R-character n as follows.

$$R\Phi(\theta, x) = \exp(2in\alpha)\Phi(\exp(-i\alpha\theta), x) \tag{3.34}$$

$$R\bar{\Phi}(\bar{\theta}, x) = \exp(-2in\alpha)\Phi(\exp(i\alpha\bar{\theta}), x). \tag{3.35}$$

Since the R-charge does not commute with the supersymmetry, the component fields of the chiral field have different R-charges. For a superfield Φ with R-character n, the R-charges of the component fields may be read off as follows:

$$R(\text{lowest component of } \Phi) = R(A) \equiv n\ , \quad R(\psi) = n - 1\ , \quad R(F) = n - 2. \tag{3.36}$$

The R-charge of the Grassmann variables is given by:

$$R(\theta_\alpha) = 1\ , \quad R(\mathrm{d}\theta_\alpha) = -1 \tag{3.37}$$

with, barred variables having opposite R charge. The kinetic term $\bar{\Phi}\Phi$ is an R invariant. ($\bar{\theta}\bar{\theta}\theta\theta$ is an invariant.) For the potential term,

$$\int \mathrm{d}^2\theta W \tag{3.38}$$

to have zero R charge requires that $R(W) = 2$. For the mass term $W = \frac{1}{2}m\Phi^2$,

$$W = m\psi\psi + m^2|A|^2\ , \tag{3.39}$$

to have vanishing R-charge requires

$$R(\Phi) = R(A) = 1\ , \quad R(\psi) = 0. \tag{3.40}$$

Adding the cubic term:

$$W_3 = \frac{\lambda}{3}\Phi^3 \tag{3.41}$$

produces

$$V = |\lambda|^2|A|^4 + \lambda A\psi\psi. \tag{3.42}$$

This term is not R-invariant with the R-charges given by (3.40). To restore R-invariance requires λ is assigned an R-charge of -1. This can be viewed as simply a book keeping device or more physically one can view the coupling as the vacuum expectation value of some field. The expectation value inherits the quantum numbers of the field. This is how one treats for example the mass parameters of Fermions in the standard model. There is also one other global $U(1)$ symmetry that commutes with the supersymmetry. All component fields are charged the same with respect to this $U(1)$ symmetry. Demanding that the terms in the action maintain this symmetry requires an assignment of $U(1)$ charges to λ, and m.

The charges are summarized in the following table:

$$\begin{array}{ccc} & U(1) & U(1)_{\mathrm{R}} \\ \Phi & 1 & 1 \\ m & -2 & 0 \\ \lambda & -3 & -1 \\ W & 0 & 2\,. \end{array} \tag{3.43}$$

These symmetries are next used to prove the nonrenormalisation theorem.

3.6 The effective potential

That the effective potential be invariant under the global $U(1)$ implies that

$$W_{\text{eff}}(\Phi, m, \lambda) = g_1(m\Phi^2, \lambda\Phi^3) \ , \tag{3.44}$$

combining this with the invariance under the $U(1)$ R-symmetry implies

$$W_{\text{eff}} = m\Phi^2 g\left(\frac{\lambda\Phi}{m}\right) \cdot \tag{3.45}$$

Consider expanding in $|\lambda| << 1$,

$$m\Phi^2 g\left(\frac{\lambda\Phi}{m}\right) = \frac{1}{2}m\Phi^2 + \frac{1}{3}\Phi^3 + \sum_{n=2}^{\infty} a_n \frac{\lambda^n\Phi^{n+2}}{m^{n-1}} \cdot \tag{3.46}$$

As $\lambda \to 0$ one must recover the classical potential. This requires that $n \geq 0$. Consider the limit $m \to 0$, for the function to be free of singularities implies that $n < 1$. There can therefore be no corrections to the form of W_{eff} [33]. In particular, if he classical superpotential has unbroken supersymmetry so does the full theory. (On the other hand, the Kahler potential is renormalized.)

3.7 Supersymmetry breaking

Let us examine now how supersymmetry may be spontaneously broken. The following anecdote may be of some pedagogical value [34]. It turns out that a short time after supersymmetry was introduced arguments were published which claimed to prove that supersymmetry cannot be broken spontaneouly. Supersymmetry resisted breaking attmepts for both theories of scalars and gauge theories. One could be surprised that the breaking was first achieved in the gauge systems. This was done by Fayet and Illiopoulos. The presence in the collaboration of a student who paid little respect to the general counter arguements made the discovery possible. Fayet went on to discover the breaking mechanism also in supersymmetric scalar theories as did O'Raighfeartaigh.

The Fayet-O'Raifeartaigh potential [35, 36] is the field theory analogue of the potential given by equation (2.96) for supersymmetric quantum mechanics. In order to break supersymmetry a minimum of three chiral fields are needed:

$$\mathcal{L}_{\text{Potential}} = \lambda\Phi_0 + m\Phi_1\Phi_2 + g\Phi_0\Phi_1\Phi_1 + h.c. \tag{3.47}$$

Minimizing the potential leads to the following equations:

$$0 = \lambda + g\Phi_1\Phi_1 \quad (3.48)$$
$$0 = m\Phi_2 + 2g\Phi_0\Phi_1 \quad (3.49)$$
$$0 = m\Phi_1. \quad (3.50)$$

These cannot be consistently solved so there cannot be a zero energy ground state and supersymmetry must be spontaneously broken. To find the ground state one must write out the full Lagrangian including the kinetic terms in component fields and then minimize. Doing so one discovers that in the ground state $A_1 = A_2 = 0$ and A_0 is arbitrary. The arbitrariness of A_0 is a flat direction in the potential, like in the quantum mechanical example (2.96). Computing the masses by examining the quadratic terms for component fields gives the following spectrum: the six real scalars have masses: $0, 0, m^2m^2, m^2 \pm 2g\lambda$; and the Fermions have masses: $0, 2m$. The zero mass Fermion is the Goldstino. We turn next to supersymmetric theories that are gauge invariant.

3.8 Supersymmetric gauge theories

A vector superfield contains spin 1 and spin $\frac{1}{2}$ component fields. It obeys a reality condition $V = \bar{V}$.

$$V = B + \theta\chi\bar{\theta}\bar{\chi} + +\theta^2 C + \bar{\theta}^2\bar{C} - \theta\sigma^\mu\bar{\theta}A_\mu \quad (3.51)$$
$$+i\theta^2\bar{\theta}(\bar{\lambda} + \frac{1}{2}\bar{\sigma}^\mu\partial_\mu\chi) - i\bar{\theta}^2\theta(\lambda - \frac{1}{2}\sigma^\mu\partial_\mu\bar{\chi}) \quad (3.52)$$
$$+\frac{1}{2}\theta^2\bar{\theta}^2(D^2 + \partial^2 B) \quad (3.53)$$

B, D, A_μ are real and C is complex. There will be a local $U(1)$ symmetry with gauge parameter, Λ an arbitrary chiral field:

$$V \rightarrow V + i(\Lambda - \bar{\Lambda}) \quad (3.54)$$

B, χ, and C are gauge artifacts and can be gauged away. The symmetry is actually $U(1)_{\mathbf{C}}$ as opposed to the usual $U(1)_{\mathbf{R}}$ because although the vector field transforms with a real gauge parameter, the other fields transform with gauge parameters that depend on the imaginary part of Λ.

It is possible to construct a chiral superfield, W_α, from V as follows

$$W_\alpha = -\frac{1}{4}\bar{D}\bar{D}D_\alpha V\ , \quad \bar{D}_{\dot{\beta}}W_\alpha = 0. \quad (3.55)$$

One may choose a gauge (called the Wess Zumino gauge) in which B, C and $\chi = 0$ and then expand in terms of component fields,

$$V(y) = -\theta\sigma^\mu\bar{\theta}A_\mu + i\theta^2\bar{\theta}\bar{\lambda} - i\bar{\theta}^2\theta\lambda + \frac{1}{2}\theta^2\bar{\theta}^2 D \tag{3.56}$$

$$W_\alpha(y) = -i\lambda_\alpha + \left(\delta_\alpha^\beta D - \frac{i}{2}(\sigma^\mu\bar{\sigma}^\nu)_\alpha^\beta F_{\mu\nu}\right)\theta_\beta + (\sigma^\mu\partial_\mu\bar{\lambda})_\alpha\theta^2. \tag{3.57}$$

Where A_μ is the vector field, $F_{\mu\nu}$ its field strength, λ is the spin $\frac{1}{2}$ field and D is an auxiliary scalar field. Under the symmetry (3.54), the component fields transform under a now $U(1)_\mathbf{R}$ symmetry as:

$$A_\mu \to A_\mu - i\partial_\mu(B - B^*)\ , \quad \lambda \to \lambda\ , \quad D \to D. \tag{3.58}$$

Note, W is gauge invariant. The following supersymmetric gauge invariant Lagrangian is then constructed:

$$\mathcal{L} = \int \mathrm{d}^2\theta \left(\frac{-i\tau}{16\pi}\right) W^\alpha W_\alpha + h.c. \tag{3.59}$$

where the coupling constant τ is now complex,

$$\tau = \frac{\theta}{2\pi} + i\frac{4\pi}{g^2}\,. \tag{3.60}$$

Expanding this in component fields produces,

$$\mathcal{L} = \frac{-1}{4g^2}F_{\mu\nu}F^{\mu\nu} + \frac{1}{2g^2}D^2 - \frac{i}{g^2}\lambda\sigma D\bar{\lambda} + \frac{\theta}{32\pi^2}(*F)^{\mu\nu}F_{\mu\nu}. \tag{3.61}$$

D is clearly a non propagating field. The *theta* term couples to the instanton number density (this vanishes for abelian fields in a non-compact space). A monopole in the presence of such a coupling will get electric charge through the Witten effect. The supersymmetries acting on the component fields are, (up to total derivatives):

$$\delta_\epsilon A = \sqrt{2}\epsilon\psi \tag{3.62}$$

$$\delta_\epsilon\psi = i\sqrt{2}\sigma^\mu\bar{\epsilon}\partial_\mu A + \sqrt{2}\epsilon F \tag{3.63}$$

$$\delta_\epsilon F = i\sqrt{2}\bar{\epsilon}\bar{\sigma}^\mu\partial_\mu\psi \tag{3.64}$$

$$\delta_\epsilon F_{\mu\nu} = i(\epsilon\sigma_\mu\partial_\nu\bar{\lambda} + \bar{\epsilon}\bar{\sigma}_\mu\partial_\nu\lambda) - (\mu \leftrightarrow \nu) \tag{3.65}$$

$$\delta_\epsilon\lambda = i\epsilon D + \sigma^{\mu\nu}\epsilon F_{\mu\nu} \tag{3.66}$$

$$\delta_\epsilon D = \bar{\epsilon}\bar{\sigma}^\mu\partial_\mu\lambda - \epsilon\sigma^\mu\partial_\mu\bar{\lambda}\ . \tag{3.67}$$

One may also add to the action a term linear in the vector field V, known as a Fayet-Iliopoulos term [37]:

$$2K\int \mathrm{d}^2\theta\mathrm{d}^2\bar{\theta}V = KD = \int \mathrm{d}\theta^\alpha W_\alpha + h.c. \tag{3.68}$$

Its role will be discussed later. The $U(1)$ gauge fields couple to charged chiral matter through the following term

$$\mathcal{L} = \sum_i \int \mathrm{d}^2\theta \mathrm{d}^2\bar{\theta} \bar{\Phi}_i \exp(q_i V)\Phi_i. \tag{3.69}$$

Under the gauge transformation

$$V \to V + i(\Lambda - \bar{\Lambda}) \ , \quad \Phi_i \to \exp(-iq_i\Lambda)\Phi_i. \tag{3.70}$$

Since there are chiral Fermions there is the the possibility for chiral anomalies. In order that the theory is free from chiral anomalies one requires:

$$\sum q_i = \sum q_i^3 = 0. \tag{3.71}$$

Writing out the term (3.69) in components produces:

$$\begin{aligned}
\mathcal{L} \quad &= \quad F^*F - |\partial_\mu \phi + \frac{iq}{2} A_\mu \phi|^2 - i\bar{\phi}\bar{\sigma}\left(\partial_\mu + \frac{iq}{2} q A_\mu\right)\psi && (3.72)\\
& \quad -\frac{iq}{\sqrt{2}}\left(\phi\bar{\lambda}\bar{\psi} - \bar{\phi}\lambda\psi\right) + \frac{1}{2} q D\bar{\phi}\phi. && (3.73)
\end{aligned}$$

There are two auxiliary fields, the D and F fields.

Adding the kinetic term (3.61) for the vector field and a potential, $\tilde{W}(\Phi)$ for the matter, gives the total Lagrangian,

$$\mathcal{L} = \int \mathrm{d}^2\theta \left(W^\alpha W_\alpha + \int \mathrm{d}^2\bar{\theta}\bar{\Phi}^i \exp(q_i V)\Phi_i + \tilde{W}(\Phi)\right) \tag{3.74}$$

this produces the following potential,

$$V = \sum_i |\frac{\partial \tilde{W}}{\partial \phi^i}|^2 + \frac{q^2}{4}\left((2K + \sum |\phi_i|^2\right)^2. \tag{3.75}$$

The first term is the F-term and the second is the D-term. Both these terms need to vanish for unbroken supersymmetry.

Some remarks about this potential:

Generically the F-terms should vanish since there are n equations for n unknowns. If $\langle\phi_i\rangle = 0$, that is if the $U(1)$ is not spontaneously broken then supersymmetry is broken if and only if $K_{\mathrm{F.I.}} \neq 0$. When $K = 0$ and the F-terms have a solution then so will the D-term and there will be no supersymmetry breaking.

These ideas are demonstrated by the following example. Consider fields Φ_1, Φ_2 with opposite $U(1)$ charges and Lagrangian given by:

$$\begin{aligned}
\mathcal{L} \quad &= \quad \frac{1}{4}(W^\alpha W_\alpha + h.c.) + \bar{\Phi}_1 \exp(eV)\Phi_1 && (3.76)\\
& \quad +\bar{\Phi}_2 \exp(-eV)\Phi_2 + m\Phi_1\Phi_2 + h.c. + 2KV. && (3.77)
\end{aligned}$$

This leads to the potential:

$$V = \frac{1}{2}D^2 + F_1 F_1^* + F_2 F_2^* \tag{3.78}$$

where

$$D + K + \frac{e}{2}(A_1^* A_1 - A_2^* A_2) = 0 \tag{3.79}$$
$$F_1 + m A_2 * = 0 \tag{3.80}$$
$$F_2 + m A_1^* = 0. \tag{3.81}$$

Leading to the following expression for the potential:

$$V = \frac{1}{2}K^2 + \left(m^2 + \frac{1}{2}eK\right) A_1^* A_1 + \left(m^2 - \frac{1}{2}eK\right) A_2^* A_2 + \frac{1}{8}e^2 \left(A_1^* A_1 - A_2^* A_2\right)^2 . \tag{3.82}$$

Consider the case, $m^2 > \frac{1}{2}eK$. The scalars have mass, $\sqrt{m^2 + \frac{1}{2}eK}$ and $\sqrt{m^2 - \frac{1}{2}eK}$. The vector field has zero mass. Two Fermions have mass m and one Fermion is massless. Since the vector field remains massless then the $U(1)$ symmetry remains unbroken. For $K \neq 0$, supersymmetry is broken as the Bosons and Fermions have different masses. (For $K = 0$ though the symmetry is restored.) The massless Fermion (the Photino) is now a Goldstino. Note that a of the supersymmetry remains as $\mathrm{Tr}M_{\mathrm{B}}^2 = \mathrm{Tr}M_{\mathrm{F}}^2$ even after the breaking of supersymmetry.

Next, consider the case, $m^2 < \frac{1}{2}eK$; at the minimum, $A_1 = 0, A_2 = v$ where $v^2 \equiv 4\frac{\frac{1}{2}eK - m^2}{e^2}$. The potential expanded around this minimum becomes, with $A \equiv A_1$ and $\tilde{A} \equiv A_2 - v$:

$$V = \frac{2m^2}{e^2}(eK - m^2) + \frac{1}{2}\left(\frac{1}{2}e^2 v^2\right) A_\mu A^\mu \tag{3.83}$$
$$+2m^2 A^* A = \frac{1}{2}(\frac{1}{2}e^2 v^2)\left(\frac{1}{\sqrt{2}}\left(\tilde{A} + \tilde{A}^*\right)\right)^2 \tag{3.84}$$
$$+\sqrt{m^2 + \frac{1}{2}e^2 v^2}\left(\psi\tilde{\psi} + \bar{\psi}\bar{\tilde{\psi}}\right) + 0 \times \lambda\bar{\lambda}. \tag{3.85}$$

The first term implies that supersymmetry is broken for $m > 0$. The photon is now massive, $m_\gamma^2 = \frac{1}{2}e^2 v^2$ implying that the $U(1)$ symmetry is broken as well. The Higgs field, $\frac{1}{\sqrt{2}}(\tilde{A} + \tilde{A}^*))^2$ has the same mass as the photon. Two Fermions have non-zero mass and there is one massless Fermion, the Goldstino.

In the above example there is both supersymmetry breaking and $U(1)$ symmetry breaking except when $m = 0$ in which case the supersymmetry remains unbroken.

Next consider a more generic example where there is $U(1)$ breaking but no supersymmetry breaking, Φ is neutral under the $U(1)$ while Φ_+ has charge $+1$ and Φ_- has charge -1. The potential is given by:

$$\mathcal{L} = \frac{1}{2}m\Phi^2 + \mu\Phi_+\Phi_- + \lambda\Phi\Phi_+\Phi_- + h.c. \tag{3.86}$$

There are two branches of solutions to the vacuum equations (a denotes the vacuum expectation value of A etc.):

$$a_+a_- = 0\ ,\quad a = -\frac{\lambda}{m} \tag{3.87}$$

which leads to no $U(1)$ breaking and

$$a_+a_- = -\frac{1}{8}\left(\lambda - \frac{m\mu}{g}\right), a = -\frac{\mu}{g} \tag{3.88}$$

which breaks the $U(1)$ symmetry. Note, the presence of a flat direction:

$$a_+ \to e^{\alpha}a_+\ ,\quad a_- \to e^{-\alpha a_-} \tag{3.89}$$

leaves a_-a_+ fixed and the vacuum equations are still satisfied.

Typical generic supersymmetry breaking leads to the breaking of R-symmetry. Since there is a broken global symmetry this would lead to the presence of Goldstone Boson corresponding to the broken $U(1)_{\mathrm{R}}$. Inverting this argument leads to the conclusion that supersymmetric breaking in nature cannot be generic since we do not observe such a particle.

So far we have only dealt with $U(1)$ vector fields. One can also consider non-abelian groups. The fields are in an adjoint representation of the group, A^a_μ, λ^a, D^a, the index a is the group index, $a = 1..\mathrm{dim(group)}$ and $D^a = \sum_i \bar{\phi}_i T^a_{R(\Phi_i)}\Phi_i$. Only if there is a $U(1)$ factor will the Fayet-Iliopoulos term be non-vanishing.

We wish to examine the properties of potentials with flat directions. Take the simple example of an abelian theory with no Fayet-Iliopoulos terms and no F-terms. This gives a $U(1)$ theory with oppositely charged fields Φ_+, Φ_-. The vanishing of the D-term implies:

$$D = |\Phi_+|^2 - |\Phi_-|^2 = 0. \tag{3.90}$$

Four real fields obeying the constraint:

$$\Phi_+ = \exp(i\alpha)\Phi_- \tag{3.91}$$

implies a three dimensional configuration space. However there is still the $U(1)$ gauge symmetry that one must mod out by. This leaves a two dimensional moduli space. A convenient gauge fixing is,

$$\Phi_+ = \Phi_- . \tag{3.92}$$

There remains a Z_2 residual gauge symmetry however:

$$\Phi_+ \to -\Phi_+ . \tag{3.93}$$

After moding out by this action, the classical moduli space is given by the orbifold, $\frac{C}{Z_2}$. There is therefore a fixed point at $\Phi_+ = 0$, which will be a singularity. What is the physical interpretation of this singularity in moduli space?

Let us study the space in terms of gauge invariant variables:

$$M \equiv \Phi_+ \Phi_- . \tag{3.94}$$

Using $\Phi_- = \Phi_+$, the Kahler potential becomes:

$$K = \bar{\Phi}_+ \Phi_+ + \bar{\Phi}_- \Phi_- = 2\bar{\Phi}_+ \Phi_+ = +2\sqrt{M\bar{M}} . \tag{3.95}$$

The metric will then be:

$$\mathrm{d}s^2 = \frac{1}{2} \frac{\mathrm{d}M \mathrm{d}\bar{M}}{\sqrt{M\bar{M}}} \cdot \tag{3.96}$$

There is a singularity in moduli space when $M = 0$. By expanding the D-term, ones sees that M is a parameter that determines the mass of the matter fields. Thus the singularity in moduli space is a signature of a particle becoming massless. Note, there is no nonrenormalisation theorem for the Kahler potential.

When $K = 0$, (as is the case above) the moduli space is determined in the following way. In the absence of an F-term then there is always a solution to the D-term equations. The moduli space is the space of all fields Φ_i, such that $D^a = 0$, modulo $G_{\mathbf{R}}$ gauge transformations. This is equivalent to the space of all constant Φ fields modulo the complex $G_{\mathbf{C}}$ gauge transformations. When $K = 0$ but in the presence of an F-term, then provided there are solutions to the F-terms equations, the D-terms automatically vanish. The moduli space is then given by the space of fields that solve the F-term equations modulo complex $G_{\mathbf{C}}$ gauge transformations. Moreover the moduli space, $\frac{\Phi_i}{G_{\mathbf{C}}}$ is spanned by a basis of the independent gauge singlets (such as $m = \phi_+ \phi_-$).

4 Phases of gauge theories

First we will gain our intuition from $D = 4$ lattice gauge theories, Z_N valued gauge fields with coupling g [38–43]

The effective "temperature" of the system will be given by, $T = \frac{Ng^2}{2\pi}\cdot$

For a given theory there is a lattice of electric and magnetically charged operators where the electric charge is denoted by n and the magnetic charge by m. An operator with charges $(n,\ m)$ is it is an irrelevant operator and weakly coupled to system, so long as the free energy, $F > 0$, that is,

$$n^2 T + \frac{m^2}{T} > \frac{C}{N}\ , \tag{4.1}$$

however when the free energy is negative for the operator $(n,\ m)$, it condenses indicating the presence of a relevant operator and hence an infra-red instability, this occur when,

$$n^2 T + \frac{m^2}{T} < \frac{C}{N}\ \cdot \tag{4.2}$$

Keeping N, C fixed and vary T. How does the theory change?

There are three phases depending which operators condense. At small T, there is electric condensation which implies that there is electric charge screening, magnetic charges are confined, and the log of the Wilson loop is proportional to the length of the perimeter of the loop. (This is called the Higgs phase.)

At high T, magnetic condensation occurs, this is the dual of electric condensation. Magnetic charges are screened, electric charges are confined and the log of the Wilson loop is proportional to the area. (This is called the confinement phase.) For intermediate values of T it is possible that there is neither screening of charges nor confinement, this is the Coulomb phase.

In the presence of a *theta* parameter, an electric charge picks up a magnetic charge and becomes dyonic [44].

$$n' = n + \frac{\theta}{2\pi} m. \tag{4.3}$$

This lead to a tilted lattice of dyonic charges and one may condense dyons with charges (n_0, m_0). This leads to what is called oblique confinement with the charges commensurate with (n_0, m_0) being screened and all other charges being confined [43].

How does this relate to QCD? There are ideas that confinement in QCD occurs due the condensation of QCD monopoles [45–47]. It is difficult to

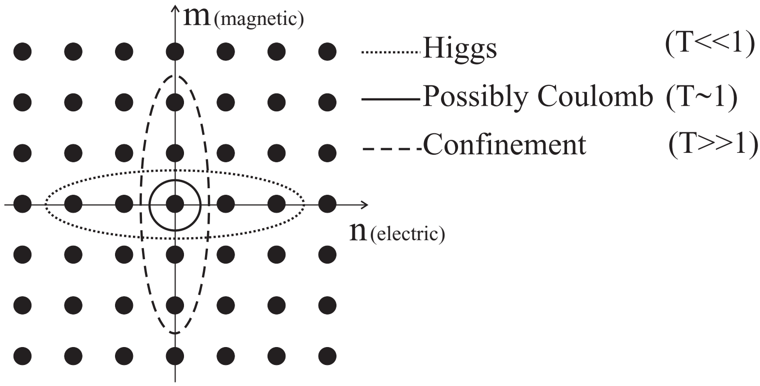


Fig. 4. The different possible phases.

study this phenomenon directly. The Dirac monopole in a $U(1)$ gauge theory is a singular object; however by embedding the monopole in a spontaneously non-abelian theory with an additional scalar field one may smooth out the core of the monopole and remove the singularity. One may proceed analogously, by enriching QCD; adding scalars and making the the theory supersymmetric one can calculate the condensation of monopoles in a four dimensional gauge theory. This has been achieved for gauge theories with $N = 1,\ 2$ supersymmetries. There are many new methods that have been utilized and the phase structures of these theories have been well investigated [15, 48, 49]. Novel properties of these theories have been discovered such as new types of conformal field theories and new sorts of infra-red duality. To this we turn next.

5 Supersymmetric gauge theories/super QCD

The goal will be to examine theories that are simple supersymmetric extensions of QCD. Consider the case of an $N = 1$ vector multiplet with gauge group $SU(N_{\mathrm{C}})$, and N_{F} chiral multiplets in the fundamental representation of $SU(N_{\mathrm{C}})$ and N_{F} chiral multiplets in the antifundamental representation. The Lagrangian is:

$$\begin{aligned}\mathcal{L} \;&=\; \int(-i\tau)\mathrm{Tr}W^{\alpha}W_{\alpha}\mathrm{d}^2\theta + h.c.\\ &\quad + Q_{\mathrm{F}}^{+}\exp(-2V)Q_{\mathrm{F}} + \tilde{Q}\exp(2V)Q_{\mathrm{F}}^{+}|_{\bar\theta\bar\theta\theta\theta} + m_{\mathrm{F}}\tilde{Q}_{\mathrm{F}}Q|_{\theta\theta} \qquad (5.1)\end{aligned}$$

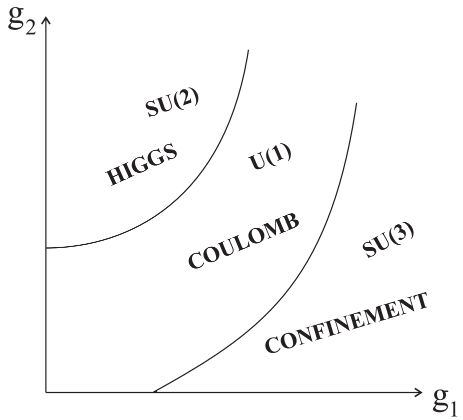


Fig. 5. Possible phases of gauge theories (g_1 and g_2 are some relevant/marginal couplings).

where the coupling is:

$$\tau = \frac{\theta}{2\pi} + i\frac{4\pi}{g^2} \cdot \tag{5.2}$$

Apart from the local $SU(N_\mathrm{C})$ gauge symmetry, the fields are charged under the following global symmetries.

$$\begin{array}{cccccccccc} & SU(N_\mathrm{F})_\mathrm{L} & \times & SU(N_\mathrm{F})_\mathrm{R} & \times & U(1)_\mathrm{V} & \times & U(1)_\mathrm{A} & \times & U(1)_{R_\mathrm{C}} \\ Q^i_a & N_\mathrm{F} & & 1 & & 1 & & 1 & & 1 \\ \tilde{Q}^a_i & 1 & & \bar{N}_\mathrm{F} & & -1 & & 1 & & 1 \\ W_\alpha & 1 & & 1 & & 0 & & 0 & & 1\,. \end{array} \tag{5.3}$$

Note, when $N_\mathrm{C} = 2$, because $2 \sim \bar{2}$, the global flavor symmetry is enhanced to $SO(2N_\mathrm{F})_\mathrm{L} \times SO(N_\mathrm{F})_\mathrm{R}$.

There is an anomaly of the $U(1)_\mathrm{A} \times U(1)_\mathrm{R}$ symmetry. A single $U(1)$ symmetry survives the anomaly. This is denoted as $U(1)_\mathrm{R}$ and is a full quantum symmetry. The adjoint Fermion contributes $2N_\mathrm{C} \times R(\lambda)$ to the anomaly. The Chiral Fermions contribute, $2N_\mathrm{F} \times R_\mathrm{F}$. $R(\lambda) = 1$, while R_F is now chosen so that the total anomaly vanishes,

$$2N_\mathrm{C} + 2R_\mathrm{F}N_\mathrm{F} = 0. \tag{5.4}$$

This leads to

$$R_\mathrm{F} = -\frac{N_\mathrm{C}}{N_\mathrm{F}} \cdot \tag{5.5}$$

The Bosons in the chiral multiplet have an R charge one greater than the Fermions in the multiplet. Thus,

$$R_{\mathrm{B}} = 1 - \frac{N_{\mathrm{C}}}{N_{\mathrm{F}}} = \frac{N_{\mathrm{F}} - N_{\mathrm{C}}}{N_{\mathrm{F}}} \,. \tag{5.6}$$

The non-anomalous R-charge, is given by:

$$R = R_{\mathrm{C}} - \frac{N_{\mathrm{C}}}{N_{\mathrm{F}}} Q_{\mathrm{A}} \,, \tag{5.7}$$

where R_{C} is the classical R-charge and A is the classical $U(1)_{\mathrm{A}}$ charge. This leads to the following non-anomalous global charges:

$$\begin{array}{ccccccccc} & SU(N_{\mathrm{F}})_{\mathrm{L}} & \times & SU(N_{\mathrm{F}})_{\mathrm{R}} & \times & U(1)_{\mathrm{V}} & \times & U(1)_{\mathrm{R}} \\ Q_a^i & N_{\mathrm{F}} & & 1 & & 1 & & \frac{N_{\mathrm{F}}-N_{\mathrm{C}}}{N_{\mathrm{F}}} \\ \tilde{Q}_i^a & 1 & & \bar{N}_{\mathrm{F}} & & -1 & & \frac{N_{\mathrm{F}}-N_{\mathrm{C}}}{N_{\mathrm{F}}} \\ W_\alpha & 1 & & 1 & & 0 & & 1\,. \end{array} \tag{5.8}$$

One is now ready to identify the classical moduli space.

5.1 The classical moduli space

The classical moduli space is given by solving the D-term and F-term equations:

$$D^a = Q_{\mathrm{F}}^{+} T^a Q_{\mathrm{F}} - \tilde{Q_{\mathrm{F}}} T^a \tilde{Q_{\mathrm{F}}}^{+} \tag{5.9}$$

$$\bar{F}_{Q_{\mathrm{F}}} = -m_{\mathrm{F}} \tilde{Q} \tag{5.10}$$

$$\bar{F}_{\tilde{Q}_{\mathrm{F}}} = -m_{\mathrm{F}} Q. \tag{5.11}$$

For $N_{\mathrm{F}} = 0$ or for $N_{\mathrm{F}} \neq 0$ and $m_{\mathrm{F}} \neq 0$, there is no moduli space. Note, the vacuum structure is an infra-red property of the system hence having $m_{\mathrm{F}} \neq 0$ is equivalent to setting $N_{\mathrm{F}} = 0$ in the infra-red.

Consider the quantum moduli space of the case where $N_{\mathrm{F}} = 0$. The Witten index, $\mathrm{Tr}(-1)^{\mathrm{F}} = N_{\mathrm{C}}$ the rank of the group $+1$. This indicates there is no supersymmetry breaking. There are $2N_{\mathrm{C}}$ Fermionic zero modes (from the vector multiplet). These Fermionic zero modes break through instanton effects the original $U(1)_{\mathrm{R}}$ down to $Z_{2N_{\mathrm{C}}}$. Further breaking occurs because the gluino two point function acquires a vacuum expectation value which breaks the symmetry down to Z_2. This leaves N_{C} vacua. The gluino condensate is:

$$\langle \lambda\lambda \rangle = \exp\left(\frac{2\pi i k}{N_{\mathrm{C}}}\right) \Lambda^3_{N_{\mathrm{C}}} \tag{5.12}$$

where Λ_{N_C} is the dynamically generated scale of the gauge theory and $k = 1,..,N_C-1$ label the vacua. Chiral symmetry breaking produces a mass gap. Note, because chiral symmetry is discrete there are no Goldstone Bosons. Further details of quantum moduli spaces will be discussed later.

Consider the case where $m_F = 0$ and $0 < N_F < N_C$. The classical moduli space is determined by the following solutions to the D-term equations:

$$Q = \tilde{Q} = \begin{pmatrix} a_1 & 0 & & & & 0..0 \\ 0 & a_2 & & & & 0..0 \\ & & \cdot & & & 0..0 \\ & & & \cdot & & 0..0 \\ & & & & a_{N_F} & 0..0 \end{pmatrix}_{N_F \times N_C.} \tag{5.13}$$

Where the row indicates the the flavor and the column indicates the colour. There are N_F diagonal non-zero real entries, a_i. (To validate this classical analysis the vacuum expectation values must be much larger than any dynamically generated scale, $a_i >> \Lambda$.) The gauge symmetry is partially broken:

$$SU(N_C) \to SU(N_C - N_F). \tag{5.14}$$

This is for generic values of a_i. By setting some subset of a_i to zero one may break to a subgroup of $SU(N_C)$ that is larger than $SU(N_C - N_F)$. Also, if $N_F = N_C - 1$ then the gauge group is complete broken. This is called the Higgs phase.

The number of massless vector Bosons becomes

$$N_C^2 - ((N_C - N_F)^2 - 1) = 2N_C N_F - N_F^2 \,, \tag{5.15}$$

the number of massless scalar fields becomes,

$$2N_C N_F - (2N_C N_F - N_F^2) = N_F^2. \tag{5.16}$$

The matrix

$$M_{\tilde{i}j} \equiv \tilde{Q}_{\tilde{i}} Q_j \tag{5.17}$$

forms a gauge invariant basis. The Kahler potential is then,

$$K = 2\mathrm{Tr}\sqrt{(M\bar{M})} \tag{5.18}$$

and as before when singularities appear $\det M = 0$ this signals the presence of enhanced symmetries.

When $N_{\rm F} \geq N_{\rm C}$, one has the following classical moduli space.

$$Q = \begin{pmatrix} a_1 & 0 & & & \\ 0 & a_2 & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & a_{N_{\rm C}} \\ 0 & 0 & \cdot\cdot & \cdot\cdot & 0 \\ \cdot & \cdot & \cdot\cdot & \cdot\cdot & 0 \\ 0 & 0 & \cdot\cdot & \cdot\cdot & 0 \end{pmatrix}_{N_{\rm F}\times N_{\rm C}} , \quad \tilde{Q} = \begin{pmatrix} \tilde{a}_1 & 0 & & & \\ 0 & \tilde{a}_2 & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \tilde{a}_{N_{\rm C}} \\ 0 & 0 & \cdot\cdot & \cdot\cdot & 0 \\ \cdot & \cdot & \cdot\cdot & \cdot\cdot & 0 \\ 0 & 0 & \cdot\cdot & \cdot\cdot & 0 \end{pmatrix}_{N_{\rm F}\times N_{\rm C}} \tag{5.19}$$

with the constraint that

$$|a_i|^2 - |\tilde{a}_i|^2 = \rho. \tag{5.20}$$

Generically the $SU(N_{\rm C})$ symmetry is completely broken. However, when $a_i = \tilde{a}_i = 0$ then a subgroup of the $SU(N_{\rm C})$ can remain.

We will now consider some special cases, first the classical moduli space for $N_{\rm F} = N_{\rm C}$. The dimension of the moduli space is given by:

$$2N_{\rm C}^2 - (N_{\rm C}^2 - 1) = N_{\rm C}^2 + 1 = N_{\rm F}^2 + 1. \tag{5.21}$$

There are $N_{\rm F}^2$ degrees of freedom from $M_{\tilde{i}j}$ and naively one would have two further degrees of freedom from:

$$B = \epsilon_{i_1 \cdots i_{N_{\rm C}}} Q^{i_1}_{j_1} \cdots Q^{i_{N_{\rm C}}}_{j_{N_{\rm F}}} \ , \ \ \tilde{B} = \epsilon_{i_1 \cdots i_{N_{\rm C}}} \tilde{Q}^{i_1}_{j_1} \cdots \tilde{Q}^{i_{N_{\rm C}}}_{j_{N_{\rm F}}}. \tag{5.22}$$

There is however a classical constraint:

$$\det M - B\tilde{B} = 0 \tag{5.23}$$

which means M, B and $\tilde{B}$ are classically dependent. This leaves only $N_{\rm F}^2+1$ independent moduli.

Generically, as well as the gauge symmetry being completely broken the global flavor symmetry is also broken. There is a singular point in the moduli space where $M = 0 = B = \tilde{B}$.

Next, consider the case, $N_{\rm f} = N_{\rm C} + 1$, again there are $N_{\rm F}^2$ moduli from $M_{i\tilde{j}}$. There are also, $2(N_{\rm C}+1)$ degrees of freedom given by:

$$B_i = \epsilon_{ii_1 \cdots i_{N_{\rm C}}} Q^{i_1}_{j_1} \cdots Q^{i_{N_{\rm C}}}_{j_{N_{\rm F}}} \ , \ \ \tilde{B}_{\tilde{i}} = \epsilon_{i_1 \cdots i_{N_{\rm C}}} \tilde{Q}^{i_1}_{j_1} \cdots \tilde{Q}^{i_{N_{\rm C}}}_{j_{N_{\rm F}}}. \tag{5.24}$$

However there are again the classical constraints:

$$\det M \quad - \quad M_{i\tilde{j}} B^i B^{\tilde{j}} = 0 \tag{5.25}$$

$$M_{\tilde{j}i} B^i \quad = \quad M_{i\tilde{j}} B^{\tilde{j}} = 0 \tag{5.26}$$

giving again an $N_{\rm F}^2+1$ dimension moduli space. (The moduli space is not smooth.) There is a generic breaking of gauge symmetry.

5.2 Quantum moduli spaces

One is required to examine on a case by case basis the role that quantum effects play in determining the exact moduli space. Quantum effects both perturbative and nonperturbative can lift moduli. In what follows we examine the quantum moduli space for the separate cases: $1 \leq N_{\rm F} \leq N_{\rm C}-1$, $N_{\rm F} = N_{\rm C}$, $N_{\rm F} = N_{\rm C}+1$, $N_{\rm C}+1 < N_{\rm F} \leq \frac{3N_{\rm C}}{2}$, $\frac{3N_{\rm C}}{2} < N_{\rm F} < 3N_{\rm C}$, $N_{\rm F} = 3N_{\rm C}$, $N_{\rm F} > 3N_{\rm C}$.

5.3 Quantum moduli space for $0 < N_F < N_C$

Classically, the dimension of the moduli space is $N_{\rm F}^2$ from $Q, \tilde{Q}$. The following table summarizes the charges under the various groups [51–53].

	$SU(N_{\rm C})$	$SU(N_{\rm F})_{\rm L}$	$SU(N_{\rm F})_{\rm R}$	$U(1)_{\rm V}$	$U(1)_{\rm A}$	$U(1)_{R_{\rm cl}}$	$U(1)_{\rm R}$
Q_a^i	$N_{\rm C}$	$N_{\rm F}$	1	1	1	1	$\frac{N_{\rm F}-N_{\rm C}}{N_{\rm F}}$
$\tilde{Q}_i^a$	$\bar{N_{\rm C}}$	1	$\bar{N}_{\rm F}$	-1	1	1	$\frac{N_{\rm F}-N_{\rm C}}{N_{\rm F}}$
$\Lambda^{3N_{\rm C}-N_{\rm F}}$	1	1	1	0	$2N_{\rm F}$	$2N_{\rm C}$	0
M	1	$N_{\rm F}$	$\bar{N}_{\rm F}$	0	2	2	$2-\frac{2N_{\rm C}}{N_{\rm F}}$
$\det M$	1	1	1	0	$2N_{\rm F}$	$2N_{\rm F}$	$2(N_{\rm F}-N_{\rm C})$

(5.27)

Λ, the dynamically generated QCD scale is assigned charges as m and g were before. The power $3N_{\rm C} - N_{\rm F}$ is the coefficient in the one loop beta function. There is no Coulomb phase so W_α does not appear.

The symmetries imply, the superpotential, W, has the following form:

$$W = (\Lambda^{3N_{\rm C}-N_{\rm F}})^a \, (\det M)^b \, c \tag{5.28}$$

a, b are to be determined. c is a numerical coefficient, If c does not vanish, the classical moduli space gets completely lifted by these nonperturbative effects.

One examines the charges of W under the various symmetries. Automatically, the charges of W for the flavor symmetries, $SU(N_{\rm F})_{\rm L} \times SU(N_{\rm F})_{\rm R}$ and the $U(1)_{\rm V}$ vanish.

If one requires the $U(1)_{\rm A}$ charge to vanish then this implies $a = -b$. Requiring the $U(1)_{\rm R}$ charge to vanish implies that $b = \frac{1}{N_{\rm F}-N_{\rm C}}$. These restrictions fix:

$$W = c \left(\frac{\Lambda^{3N_{\rm C}-N_{\rm F}}}{\det M} \right)^{\frac{1}{N_{\rm C}-N_{\rm F}}} . \tag{5.29}$$

For non vanishing c, all the moduli are now lifted and there is no ground state.

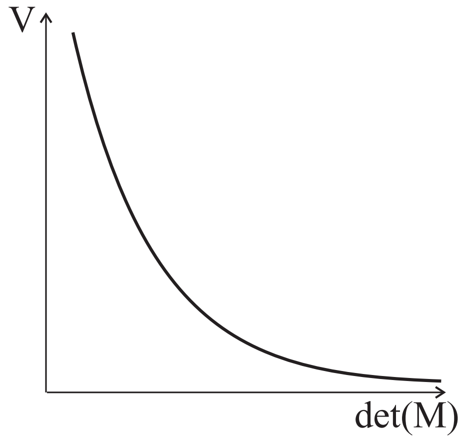


Fig. 6. The potential for $1 < N_\mathrm{F} < N_\mathrm{C}$, it has no ground state.

What is the value of c? This is a difficult to calculate directly unless there is complete Brouting. For $N_\mathrm{F} = N_\mathrm{C} - 1$ there is complete symmetry breaking and one can turn to weak coupling. From instanton calculations one calculates that $c \neq 0$ and the prepotential for the matter fields is

$$W \sim \left(\frac{\Lambda^{2N_\mathrm{C}+1}}{\det M}\right) . \tag{5.30}$$

One may now go to $N_\mathrm{F} < N_\mathrm{C} - 1$ by adding masses and integrating out the heavy degrees of freedom. This produces,

$$\langle M^i{}_j \rangle_{\min} = (m^{-1})^i{}_j (\Lambda^{3N_\mathrm{C}-N_\mathrm{F}} \det m)^{\frac{1}{N_\mathrm{C}}} . \tag{5.31}$$

5.4 Integrating in

This method involves the addition of very massive fields to known effective actions and extrapolating to the case where the additional degrees of freedom are massless, see [54–62]. It is rather surprising that anything useful can be learned by this flow in the opposite direction to the usual infra-red. We will show that under certain circumstances it is possible to derive in a rather straight forward way the potential for light fields. We thus give some of the flavour of this possibility. It gives results for the phase structure in many cases. We will also discuss when these conditions are met. We begin by reviewing the conventional method of integrating out; heavy degrees of freedom are integrated out to obtain an effective potential for the light degrees of freedom.

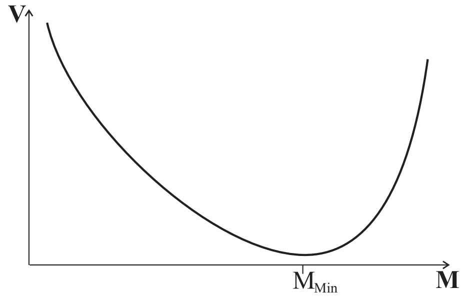


Fig. 7. The potential with finite masses has a ground state. $M_{\min} \to \infty$ as $m \to 0$.

Consider a theory containing gauge invariant macroscopic light fields of the following nature:

Fields X: built out of d_{A} degrees of freedom

Fields M: built out of u_i degrees of freedom

Fields Z: built out of both d_a and u_i.

The d_a dofs will be kept light throughout all the discussion. The u_i and with them the macroscopic fields M and Z will be considered as heavy in part of the discussion. Assume that one is given the effective potential $W_u(X, M, Z, \Lambda_u)$ describing all the light macroscopic degrees of freedom. (Λ_u is the dynamically generated scale of the theory. The effective potential is the Legendre transform

$$W_u(\Phi) = (\tilde{W}(g_i) - \sum g_i \Phi_i)_{\langle g_i \rangle} \tag{5.32}$$

where

$$\frac{\partial \tilde{W}(g_i)}{\partial g_i} = \langle \Phi_i \rangle \cdot \tag{5.33}$$

Consider next making the microscopic "up" fields massive and integrating them out retaining only the light degrees of freedom and the couplings $\tilde{m}, \lambda$ to the macroscopic degrees of freedom containing the removed fields: M and Z. One thus obtains:

$$\tilde{\mathrm{W}}_d(X, \tilde{m}, \lambda, \Lambda_u) = (W_u(X, M, Z, \Lambda_u) = \tilde{m} M + \lambda Z)_{\langle M \rangle, \langle Z \rangle}. \tag{5.34}$$

For the case $\tilde{m} \to \infty$ one may tune the scale Λ_u so as to replace an appropriate combination of $\lambda, \tilde{m}$ and Λ_u by Λ_d the scale of the theory of the remaining light degrees of freedom. One ends with $W_d(X, \Lambda_d)$. It is convenient for general $\tilde{m} \neq 0$ to write,

$$\tilde{W}_d(X, \tilde{m}, \lambda, \Lambda_u) = W_d(X, \Lambda_d) + W_I(X, \tilde{m}, \lambda, \Lambda_d) \tag{5.35}$$

where $W_d(X,\Lambda_d)$ is the exact result for infinitely heavy U_i, and to partition again,

$$W_{\rm I} = W_{{\rm tree},d} + W_\Delta \tag{5.36}$$

where

$$W_{{\rm tree},d} = (W_{\rm tree})_{\langle u_i\rangle} \tag{5.37}$$

has by definition no scale dependence. W_Δ should obeys the constraints:

$$W_\Delta \to 0\ , \qquad \text{when} \quad \Lambda_u \to 0\ , \quad or\ \tilde{m} \to \infty\ . \tag{5.38}$$

Now one reverses the direction and integrates in. Namely, given an exact form for $W_d(X,\Lambda_d)$ one can obtain:

$$W_u(X,M,Z,\Lambda_u) = (W_d(X,\Lambda_d) + W_\Delta + W_{{\rm tree},d} - W_{\rm tree})_{\langle\tilde{m}\rangle,\langle\lambda\rangle}. \tag{5.39}$$

All the essential complexities of the flow lie in the term W_Δ. Imagine however that this term would vanish. In this case the calculations become much simpler. For the cases corresponding to the colour group $SU(2)$ this indeed turns out to be the case. This can be seen in the following manner.

One starts with $SU(2)$ without adjoint fields ($N_{\rm a}=0$), and quark flavors Q_i, $i=1,..N_{\rm F}$ ($N_{\rm F}\leq 4$) for the down theory. one brings down infinitely heavy fields in the adjoint representation, that is, one resuscitates the full up theory, which contains $N_{\rm a}$ adjoint fields Φ_α with the superpotential:

$$W_{\rm tree} = mX + \tilde{m}M + \lambda Z, \tag{5.40}$$

where $M=\Phi\Phi$ and $Z=Q\Phi Q$. For convenience one writes,

$$W_\Delta(,\tilde{m},\lambda,\Lambda) = W_{{\rm tree},d} f(t) \tag{5.41}$$

where $X=WW$ and t being any possible singlet of $SU(2N_{\rm F})\times U(1)_{\rm Q}\times U(1)_\Phi\times U(1)_{\rm R}$; Φ is the adjoint field we add. The quantum numbers of all relevant fields and parameters are given by:

$$\begin{array}{cccc} & U(1)_{\rm Q} & U(1)_\Phi & U(1)_{\rm R} \\ X & 2 & 0 & 0 \\ \lambda & -2 & -1 & 2 \\ \Lambda^{b_1} & 2N_{\rm F} & 4N_{\rm a} & 4-4N_{\rm a}-2N_{\rm F} \\ \tilde{m} & 0 & -2 & 2 \\ W_\Delta & 0 & 0 & 2 \end{array} \tag{5.42}$$

where $b_1 = 6-2N_{\rm a}-N_{\rm F}$. Writing t as,

$$t \sim (\Lambda^{b_1})^a \tilde{m}^b X^c \lambda^d. \tag{5.43}$$

It is a singlet provided that

$$b = (2N_{\rm a} + 2 - N_{\rm F})a \ , \qquad c = (N_{\rm F} - 4)a \ , \qquad d = (2N_{\rm F} - 4)a \ . \tag{5.44}$$

Recall the constraints

$$W_\Delta \to 0 \quad \text{for} \quad \tilde{m} \to \infty \quad \textit{and} \quad W_\Delta \to 0 \quad \text{for} \quad \Lambda \to 0 \tag{5.45}$$

as well as the fact that in the Higgs phase one can decompose W as:

$$W(\Lambda^{b_1}) = \sum_{n=1}^{\infty} a_n (\Lambda^{b_1})^n. \tag{5.46}$$

One shows that

$$W_{\rm tree,d} \sim \frac{1}{\tilde{m}} \tag{5.47}$$

which implies that for all values of $N_{\rm f}$ the vanishing of W_Δ. We see this explicitly in the following cases.

For $N_{\rm F} = 0$ and $N_{\rm a} = 1$

$$b = 4a \tag{5.48}$$

leading to

$$W_\Delta = \sum_{a=1}^{\infty} r_{\rm a} (\Lambda^{b_1})^a \tilde{m}^{a-1}. \tag{5.49}$$

The constraints imply that $r - a$ indeed vanishes for all a's.

For $N_{\rm a} + 1, \ N_{\rm F} = 1, 2, 3, \ b_1 = 4 - N_{\rm F}$,

$$W_\Delta = \frac{r(\tilde{m}\Lambda)^{b_1}}{\tilde{m}} + ... \tag{5.50}$$

implying that $r = 0$ as well.

Starting from known results for $N_{\rm a} = N_{\rm F} = 0$ one can now obtain the effective potential for all relevant values of $N_{\rm a}$ and $N_{\rm F}$. The equations of motion of these potentials can be arranged in such a manner as to coincide with the singularity equations of the appropriate elliptic curves derived for systems with $N = 2$ supersymmetry and $SU(2)$ gauge group. (The role of these elliptic curves in $N = 2$ theories will be described later.)

5.5 Quantum moduli space for $N_F \geq N_C$

There is a surviving moduli space. In the presence of a mass matrix, m_{ij} for matter one obtains,

$$\langle M^i{}_j \rangle = (m^{-1})^i{}_j (\Lambda^{3N_\mathrm{C} - N_\mathrm{F}} \det m)^{\frac{1}{N_\mathrm{C}}}. \tag{5.51}$$

Previously, for the case of $N_\mathrm{F} < N_\mathrm{C}$, it turned out that $m \to 0$ implied $\langle M^i{}_j \rangle \to \infty$ thus explicitly lifting the classical moduli space. For $N_\mathrm{F} \geq N_\mathrm{C}$ it is possible to have $m \to 0$ while keeping $\langle M^i{}_j \rangle$ fixed.

5.6 $N_F = N_C$

Quantum effects alter the classical constraint to:

$$\det M - B\tilde{B} = \Lambda^{2N_\mathrm{C}}. \tag{5.52}$$

This has the effect of resolving the singularity in moduli space. The absence of a singularity means there will not be additional massless particles. In this case,

$$M_i{}^j = (m^{-1})_i{}^j (\det m)^{\frac{1}{N_\mathrm{C}}} \Lambda^2. \tag{5.53}$$

This implies,

$$\det M = \Lambda^{2N_\mathrm{C}} \tag{5.54}$$

since m cancels, thus this also holds in the limit $m \to 0$. On the other hand, $\langle B \rangle = \langle \tilde{B} \rangle = 0$ if $\det m \neq 0$ because all fields carrying B number are integrated out. Therefore $\det M \neq B\tilde{B}$ through quantum effects.

Note, $R(M_\mathrm{IJ}) = 0$ for $N_\mathrm{F} = N_\mathrm{C}$. Writing out an expansion that obeys the R-charge conservation:

$$\det M - B\tilde{B} = \Lambda^{2N_\mathrm{C}} \left(1 + \sum_{ij} c_{ij} \frac{(B\tilde{B})^i \Lambda^{2N_\mathrm{C} j}}{(\det M)^{i+j}} \right) \tag{5.55}$$

then by demanding that there be no singularities at small $\langle M \rangle$ or at large $\langle B \rangle$ implies that all c_{ij} must vanish and the hence $\det M - B\tilde{B} = \Lambda^{2N_\mathrm{C}}$ obeys a nonrenormalisation theorem. Note,

$$R(M) = 0 \Rightarrow W = 0\,. \tag{5.56}$$

There are allowed . . . perturbations, mass terms given by:

$$W = \mathrm{tr}(mM) + bB + \tilde{b}\tilde{B}. \tag{5.57}$$

One check is to integrate out to give the case $N_{\rm F} = N_{\rm C} - 1$ yielding,

$$W = \frac{m\Lambda^{2N_{\rm C}}}{\det M} = \frac{\tilde{\Lambda}^{2N_{\rm C}+1}}{\det M} \, . \tag{5.58}$$

What is the physics of this theory, is it in a Higgs/confinement phase? For large, $M/B/\tilde{B}$ one is sitting in the Higgs regime; however, for small $M/B/\tilde{B}$ one is in the confining regime. Note that M cannot be taken smaller than Λ.

Global symmetries need to be broken in order to satisfy the modified constraint equation.

Consider some examples: With the following expectation value,

$$\langle M^i{}_j\rangle = \delta^i{}_j\Lambda^2 \, , \qquad \langle B\tilde{B}\rangle = 0 \tag{5.59}$$

the global symmetries are broken to:

$$SU(N_{\rm F})_{\rm V} \times U_{\rm B}(1) \times U_{\rm R}(1) \tag{5.60}$$

and there is chiral symmetry breaking. When,

$$\langle M^i{}_j\rangle = 0 \, , \qquad \langle B\tilde{B}\rangle \neq 0 \tag{5.61}$$

then the group is broken to:

$$SU(N_{\rm F})_{\rm L} \times SU(N_{\rm F})_{\rm R} \times U_{\rm R}(1) \tag{5.62}$$

which has chiral symmetry and also has confinement. This is an interesting situation because there is a dogma that as soon as a system has a bound state there will be chiral symmetry breaking [65].

In both cases the 't Hooft anomaly conditions [66] are satisfied. These will be discussed later.

5.7 $N_F = N_C + 1$

The moduli space remains unchanged. The classical and quantum moduli spaces are the same and hence the singularity when $M = B = \tilde{B} = 0$ remains. This is not a theory of massless gluons but a theory of massless mesons and baryons. When, $M, B, \tilde{B} \neq 0$ then one is in a Higgs/confining phase. At the singular point when, $M = B = \tilde{B} = 0$ there is no global symmetry breaking but there is "confinement" with light baryons.

There is a suggestion that in this situation, $M, B, \tilde{B}$ become dynamically independent. The analogy is from the nonlinear sigma model, where because of strong infra-red fluctuations there are n independent fields even though there is a classical constraint. The effective potential is:

$$W_{\rm eff} = \frac{1}{\Lambda^{2N_{\rm C}-1}}(M^i{}_jB_i\tilde{B}^j - \det M) \tag{5.63}$$

the classical limit is taken by:

$$\Lambda \to 0 \tag{5.64}$$

which in turn imposes the classical constraint. Again the system obeys the 't Hooft anomaly matching conditions.

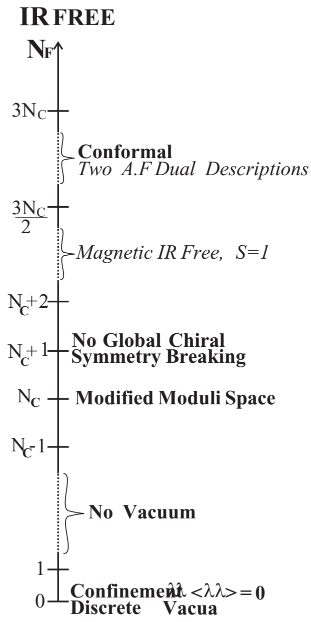


Fig. 8. The phases of super QCD.

5.8 Higgs and confinement phases

This section is discussed in [63]. While one is discussing the confinement phase in supersymmetric gauge theories one should recall that for gauge theories such as $SU(N)$ Yang Mills, with matter in a nontrivial representation of the center of the group, which is Z_N for $SU(N)$, the difference between the Higgs and confinement phases is purely quantitative. There is no phase boundary. This contrasts the situation of pure QCD or super QCD where all the particles are in the adjoint representation which is trivial under the center. In such a case there is a phase transition and there is a qualitative difference between the phases. So what about the standard model, $SU(2) \times U(1)$. Is it in a Higgs or confinement phase? Below we present the spectrum in the two pictures. In the Englert picture:

$$s = 0 \qquad I = \frac{1}{2} \qquad \phi \to \phi_{\text{real}} \tag{5.65}$$

$$s = \frac{1}{2} \qquad I = \frac{1}{2} \qquad (l)_{\text{L}}\ (q)_{\text{L}} \tag{5.66}$$

$$s = \frac{1}{2} \qquad I = 0 \qquad (l)_{\text{R}}\ (q)_{\text{R}}. \tag{5.67}$$

In the confinement picture,

$$(l)_{\text{R}}, (q)_{\text{R}} \tag{5.68}$$

are $SU(2)_{\text{L}}$ singlets. Along with,

$$s = \frac{1}{2} \qquad \phi^+\psi_i\ , \qquad \epsilon_{ij}\psi_i\phi_j \tag{5.69}$$

$$s = 0 \qquad \phi_i^+\phi_i \tag{5.70}$$

$$s = 1 \qquad \phi_i D_\mu \phi_j \epsilon_{ij}\ , \qquad \phi_i^+ D_\mu \phi_j^+ \epsilon_{ij}\ , \qquad \phi_i^+ D_\mu \phi_i. \tag{5.71}$$

One may choose a gauge,

$$\phi(x) = \Omega(x) \begin{pmatrix} \rho(x) \\ 0 \end{pmatrix} \tag{5.72}$$

then

$$B_\mu = \Omega^+ D_\mu \Omega \tag{5.73}$$

leading to the Lagrangian,

$$\mathcal{L} = \text{tr} F_{\mu\nu}(B) F^{\mu\nu}(B) + \partial_\mu \rho \partial^\mu \rho + \rho^2 (B_\mu^+ B^\mu)_{||} + V(\rho^2). \tag{5.74}$$

Unitary gauge is $\Omega = 1$. The Higgs picture also contains the operators:

$$\psi_1 = \frac{\phi_i^+\psi_i}{|\phi|}\ , \quad \psi_2 = \frac{\phi_i\psi_j\epsilon_{ij}}{|\phi|}\ , \quad \tilde{W}_\mu^+ = \frac{\phi_i^+}{|\phi|} D_\mu \frac{\phi_j^+}{|\phi|}\epsilon_{ij}\ , \quad \tilde{W}^0 = \frac{\phi_i^+}{|\phi|} D_\mu \frac{\phi_i^+}{|\phi|}\ . \tag{5.75}$$

Like confinement but with the scale determined by: $|\phi|$. At finite temperature the two phases are qualitatively indistinguishable.

Examine the charges of the fields with respect to the unbroken $U(1)$. In the Higgs picture,

$$Q(\psi) = \frac{1}{2}e \quad Q(\phi) = \frac{1}{2}e \quad Q(W^0) = 0. \tag{5.76}$$

However, the confined objects have integral charge. The appropraite conserved charge is actually:

$$Q' = Q + T_3. \tag{5.77}$$

Then,

$$Q'(e) = 1 \quad Q'(W^0) = 0 \quad Q'(v) = 0 \quad Q'(W^{\pm}) = \pm 1 \quad Q'(\rho) = 0 \tag{5.78}$$

and

$$Q'(\psi\phi) = 1 \quad Q'(\phi^+ D_\mu \phi) = 0 \quad Q'(\phi + \psi) = 0 \tag{5.79}$$

$$Q'(\phi^+ D_\mu \phi^+) = 1 \quad Q'(\phi^+ \phi) = 0 \quad Q'(\phi D_\mu \phi) = -1. \tag{5.80}$$

This matches the charges of operators in the confined picture.

Is it possible to have strong "weak" interactions at high energies? This is theoretically possible but it is not the course chosen by nature for the standard model. This is seen empirically by the absence of radial excitations of the Z particles [64].

Note, this will not work for the Georgi-Glashow model where ϕ is a triplet in $SO(3)$.

5.9 *Infra-red duality*

Two systems are called infra-red dual if, when observed at longer and longer length scales, they become more and more similar.

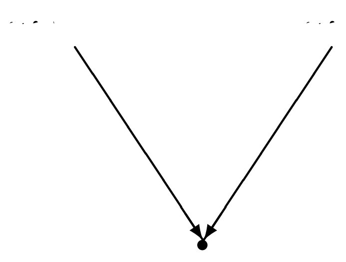


Fig. 9. Two systems with a different ultra-violet behavior flowing to the same infra-red fixed point.

Seiberg has observed and has given very strong arguments that the following set of $N = 1$ supersymmetric gauge theories are pairwise infra-red dual [88].

System		*Dual System*		
Gauge Group	#flavors	*Gauge Group*	#flavor	#singlets
$SU(N_{\mathrm{C}})$	N_{F}	$SU(N_{\mathrm{F}}-N_{\mathrm{C}})$	N_{F}	N_{F}^2
$SO(N_{\mathrm{C}})$	N_{F}	$SO(N_{\mathrm{F}}-N_{\mathrm{C}}+4)$	N_{F}	N_{F}^2
$Sp(N_{\mathrm{C}})$	$2N_{\mathrm{F}}$	$Sp(N_{\mathrm{F}}-N_{\mathrm{C}}-2)$	$2N_{\mathrm{F}}$	N_{F}^2.

For a given number of colors, N_{C}, the number of flavors, N_{F}, for which the infra-red duality holds is always large enough so as to make the entries in the table meaningful. Note that the rank of the dual pairs is usually different. Lets explain why this result is so powerful. In general, it has been known for quite a long time that two systems which differ by irrelevant operator have the same infra-red behavior. We have no indication whatsoever, that this is the case with Seiberg's duality, where groups with different number of colors are infra-red dual. Nevertheless, the common wisdom in hadronic physics has already identified very important cases of infra-red duality. For example, QCD, whose gauge group is $SU(N_{\mathrm{C}})$ and whose flavor group is $SU(N_{\mathrm{F}}) \times SU(N_{\mathrm{F}}) \times U(1)$, is expected to be infra-red dual to a theory of massless pions which are all color singlets. The pions, being the spin-0 Goldstone Bosons of the spontaneously broken chiral symmetry, are actually infra-red free in four dimensions. We have thus relearned that free spin-0 massless particles can actually be the infra-red ashes of a strongly-interacting theory, QCD whose ultraviolet behavior is described by other particles. By using supersymmetry, one can realize a situation where free massless spin-$\frac{1}{2}$ particles are also the infra-red resolution of another theory. Seiberg's duality allows for the first time to ascribe a similar role to massless infra-red free spin-1 particles. Massless spin-1 particles play a very special role in our understanding of the basic interactions. This comes about in the following way: consider, for example, the $N=1$ supersymmetric model with N_{C} colors and N_{F} flavors. It is infra-red dual to a theory with $N_{\mathrm{F}}-N_{\mathrm{C}}$ colors and N_{F} flavors and N_{F}^2 color singlets. For a given N_{C}, if the number of flavors is in the interval $N_{\mathrm{C}}+1 < N_{\mathrm{F}} < \frac{3N_{\mathrm{C}}}{2}$, the original theory is strongly coupled in the infra-red, while the dual theory has such a large number of flavors that it becomes infra-red free. Thus the infra-red behavior of the strongly-coupled system is described by infra-red free spin-1 massless fields (as well as its superpartners), that is, Seiberg's work has shown that infra-red free massless spin-1 particles(for example photons in a SUSY system) could be, under certain circumstances, just the infra-red limit of a much more complicated ultraviolet theory. Seiberg's duality has passed a large number of consistency checks under many circumstances.

The infra-red duality relates two disconnected systems. From the point of view of string theory the two systems are embedded in a larger space of models, such that a continuous trajectory relates them. We will describe the ingredients of such an embedding [68, 69] later. In order to be able to appreciate how that is derived, we will need to learn to use some tools of string theory.

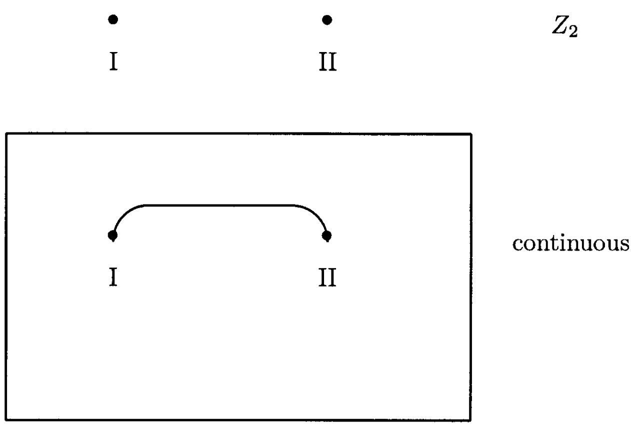


Fig. 10. In string theory, a continuous path in parameter space relates a pair of two disjoint infra-red-dual field theories.

First we will describe some more details of the Seiberg infra-red duality in field theory. Consider the example of $N = 1$ supersymmetric Yang-Mills theory with gauge group $SU(N_{\rm C})$ and $N_{\rm F}$, $\bar{N}_{\rm F}$ fundamental, anti-fundamental matter. The charges of the matter fields are given by the table below:

$$\begin{array}{ccccc} & SU_{\rm L}(N_{\rm F}) & SU_{\rm R}(N_{\rm F}) & U_{\rm B}(1) & U_{\rm R}(1) \\ Q & N_{\rm F} & 1 & 1 & 1-\frac{N_{\rm C}}{N_{\rm F}} \\ \tilde{Q} & 1 & \bar{N}_{\rm F} & -1 & 1-\frac{N_{\rm C}}{N_{\rm F}}. \end{array} \tag{5.81}$$

The infra-red dual is $N = 1$ supersymmetric Yang-Mills with gauge group $S(N_{\rm F} - N_{\rm C})$ and $N_{\rm F}$, $\bar{N}_{\rm F}$ fundamental and antifundamental matter and $N_{\rm F}^2$

gauge singlets. The charges are given by:

$$\begin{array}{ccccc} & SU_{\rm L}(N_{\rm F}) & SU_{\rm R}(N_{\rm F}) & U_{\rm B}(1) & U_{\rm R}(1) \\ q & \bar{N}_{\rm F} & 1 & \frac{N_{\rm C}}{N_{\rm F}-N_{\rm C}} & 1-\frac{N_{\rm C}}{N_{\rm F}} \\ \tilde{q} & 1 & N_{\rm F} & -\frac{N_{\rm C}}{N_{\rm F}-N_{\rm C}} & 1-\frac{N_{\rm C}}{N_{\rm F}} \\ M & N_{\rm F} & \bar{N}_{\rm F} & 0 & 2\frac{N_{\rm F}-N_{\rm C}}{N_{\rm F}}. \end{array} \tag{5.82}$$

One must also add an interaction term in the dual theory described by:

$$W = \frac{1}{\mu} M^i{}_{\tilde{j}} q_i q^{\tilde{j}}. \tag{5.83}$$

- The dual theories have different gauge groups. If one regards a gauge symmetry as a redundancy in the description of the theory then this is not important. What does matter is that the two dual theories share the same global symmetries;

- Note, it is not possible to build a meson out of $q, \tilde{q}$ that has the same R-charge as a meson built from $Q, \tilde{Q}$. The M field in the dual theory does have the same charges as a meson built from $Q, \tilde{Q}$;

- The Baryons built from $Q, \tilde{Q}$ have the same charges as those built from $q, \tilde{q}$. For the case $N_{\rm F} = N_{\rm C} + 1$ then the Baryon of the $SU(N_{\rm C})$ theory becomes the q in the dual theory (which is a singlet in this case). This looks like there is a solitonic dual for the quarks in this case;

- If M is fundamental there should be an associated $U_{\rm M}(1)$ charge which does not appear in the original $SU(N_{\rm C})$ theory;

- Where are the $q, \tilde{q}$ mesons in the original $SU(N_{\rm C})$ theory?

- the resolution to the previous two points is provided by the interaction term (5.83). This term breaks the $U_{\rm M}(1)$ symmetry and provides a mass to the $q, \tilde{q}$ mesons, which implies one may ignore them in the infra-red.

For the case, $\frac{3N_{\rm C}}{2} < N_{\rm F} < 3N_{\rm C}$, and for $\frac{3}{2}(N_{\rm F} - N_{\rm C}) < \tilde{N}_{\rm F} < 3(N_{\rm F} - N_{\rm C})$. The operator, $Mq\tilde{q}$ has dimension:

$$D(Mq\tilde{q}) = 1 + \frac{3N_{\rm C}}{N_{\rm F}} < 3 \tag{5.84}$$

and so it is a relevant operator. In both dual pictures there is an Infra red fixed point; both are asymptotically free and in the center of moduli space the theories will be a conformal.

The checks of the duality are as follows:

- They have the same global symmetries;
- They obey the 't Hooft anomaly matching conditions;
- It is a Z_2 operation;
- There are the same number of flat directions;
- There is the same reaction to a mass deformation. Adding a mass in one theory is like an Englerting in the other;
- There is a construction of the duality by embedding the field theory in string theory. This will be the subject of the next section.

The 't Hooft anomaly matching conditions are determined as follows. One takes the global symmetries in the theory and then make them local symmetries. One then calculates their anomalies. Both dual theories must share the same anomalies. In the above example there are anomalies for:

$$SU(N_{\mathrm{F}})^3\ ,\ \ SU(N_{\mathrm{F}})^2U_{\mathrm{R}}(1)\ ,\ \ SU(N_{\mathrm{F}})^2U_{\mathrm{B}}(1)\ ,\ \ U_{\mathrm{R}}(1)\ , \tag{5.85}$$

$$U_{\mathrm{R}}(1)^3\ ,\ \ U_{\mathrm{B}}(1)^2U_{\mathrm{R}}(1)\ ,\ \ U_{\mathrm{B}}(1)^3\ ,\ \ U_{\mathrm{B}}(1)^2U_{\mathrm{R}}(1). \tag{5.86}$$

All these anomalies match between the dual theories. There was no reason for them to do so.

Let us examine some of the consequences of this duality. For the case $\frac{3N_{\mathrm{C}}}{2} < N_{\mathrm{F}} < 3N_{\mathrm{C}}$ the two dual theories are both asymptotically free. It is symmetric around $N_{\mathrm{F}} = 2N_{\mathrm{C}}$. Perhaps one can more learn about this system since it is a fixed point under duality. At the origin of moduli space one may have obtained a new conformal theory- this will be discussed later. For $N_{\mathrm{C}} + 2 \leq N_{\mathrm{F}} \leq 3\frac{N_{\mathrm{C}}}{2}$, the theory is an infra-red free gauge theory plus free singlets. This is the first example of a weakly interacting theory with spin one particles that in the infra-red one may view as bound states of the dual theory. The panorama of these structures is given in Figure 8.

Let us now enrich the structure of the theory by adding N_{a} particles in the adjoint representation. At first we will have no matter in the fundamental representation and scalar multiplets which are adjoint valued. The potential for the scalars, ϕ_i is given by:

$$V = ([\phi, \phi])^2\ . \tag{5.87}$$

This potential obviously has a flat direction for diagonal ϕ. The gauge invariant macroscopic moduli would be $\mathrm{Tr}\phi^k$. Consider the non generic example of $N_{\mathrm{C}} = 2$ and $N_{\mathrm{a}} = 1$, the supersymmetry is now increased to

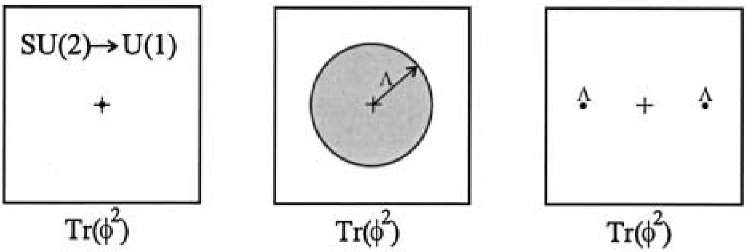


Fig. 11. The classical, naive quantum and exact quantum moduli spaces.

$N = 2$. There is a single complex modulus, $\mathrm{Tr}\phi^2$. Classically, $SU(2)$ is broken to $U(1)$ for $\mathrm{Tr}\phi^2 \neq 0$. One would expect a singularity at $\mathrm{Tr}\phi^2 = 0$. The exact quantum potential vanishes in this case [48, 49].

Naively, one would expect the following. When $\mathrm{Tr}\phi^2$ is of order Λ or smaller, one would expect that the strong infra-red fluctuations would wash away the expectation value for $\mathrm{Tr}\phi^2$ and the theory would be confining. The surprising thing is that when $SU(2)$ breaks down to $U(1)$, because of the very strong constraints that supersymmetry imposes on the system, there are only two special points in moduli space and even there the theory is only on the verge of confinement. Everywhere else the theory is in the Coulomb phase. At the special points in the moduli space, new particles will become massless. This is illustrated in the Figure 11.

We will now examine the effective theory at a generic point in moduli space where the theory is broken down to $U(1)$. The Lagrangian is given by,

$$\mathcal{L} = \int \mathrm{d}^2\theta Im(\tau_{\mathrm{eff}}(\mathrm{tr}\phi^2, g, \Lambda) W_\alpha W^\alpha). \tag{5.88}$$

The τ_{eff} is the effective complex coupling which is a function of the modulus, $\mathrm{tr}\phi^2$, the original couplings and the scale, Λ. This theory has an $SL(2, Z)$ duality symmetry. The generators of the $SL(2, Z)$ act on τ, defined by 5.2, as follows:

$$\tau \to -\frac{1}{\tau} , \quad \tau \to \tau + 1. \tag{5.89}$$

This is a generalization of the usual $U(1)$ duality that occurs with electromagnetism to the case of a complex coupling. Recall the usual electromagnetic duality for Maxwell theory in the presence of charged matter is:

$$E \to B , \;\; B \to -E , \;\; e \to m , \;\; m \to -e. \tag{5.90}$$

This generalizes to a $U(1)$ symmetry by defining:

$$E + iB , \quad e + im. \tag{5.91}$$

The duality symmetry now acts by:

$$E + iB \to \exp(i\alpha)(E + iB)\ ,\quad e + im \to \exp(i\alpha)(e + im). \tag{5.92}$$

Previously for the $SU(2)$ case the moduli were given by $u = \mathrm{Tr}\phi^2$ for $SU(N_C)$ the moduli are given by $u_k = \mathrm{Tr}\phi^k$, $k = 2, .., N_C$. Again the classical moduli space is singular at times, there is no perturbative or non-perturbative corrections.

How does one find τ as a function of the u? There is a great deal of literature on the subject here we will just sketch the ideas [48–50].

The following complex equation,

$$y^2 = ax^3 + bx^2 + cx + d \tag{5.93}$$

determines a torus. The complex structure of the torus, τ_{torus} will be identified with the complex coupling τ_{eff}. a, b, c, d are holomorphic functions of the moduli, couplings and scale and so will implicitly determine τ_{torus}.

When $y(x)$ and $y'(x)$ vanish for the same value of x then τ is singular. Therefore,

$$\tau_{\text{eff}} = i\infty\ ,\quad g^2_{\text{eff}} = 0 \tag{5.94}$$

and the effective coupling vanishes. This reflects the presence of massless charged objects. This occurs for definite values of u in the moduli space. These new massless particles are monopoles or dyons. The theory is on the verge of confinement. For $N = 2$ supersymmetry that is the best one can do. The monopoles are massless but they have not condensed [54–57, 59, 60, 62]. For condensation to occur the monopoles should become tachyonic indicating an instability that produces a condensation. One can push this to confinement by adding a mass term: $\tilde{m}\mathrm{Tr}\phi^2$, or generally for $SU(N_C)$ the term:

$$\delta W = g_k u_k\ . \tag{5.95}$$

The effective prepotential is now:

$$W = M(u_k)q\tilde{q} + g_k u_k \tag{5.96}$$

then

$$\frac{\partial W}{\partial u_k} = 0\ ,\quad \frac{\partial W}{\partial(q\tilde{q})} = 0 \Rightarrow M(\langle u_k\rangle) = 0\ ,\quad \partial_{u_k} M(\langle u_k\rangle)\langle q\tilde{q}\rangle\ = -g_k. \tag{5.97}$$

Since generically,

$$\partial_{u_k} M(\langle u_k \rangle) \neq 0 \tag{5.98}$$

then there will be condensation.

We now describe how the complex elliptic curve arises using more physical terms. This is achieved here by using the integrating in method discussed earlier. Consider the case, $N_{\rm C} = 2$ with arbitrary $N_{\rm F}$ and $N_{\rm a}$. The fields that are the moduli in the system are:

$$\begin{aligned} X_{\rm IJ} &= \epsilon_{ab} Q_i^a Q_j^b && (5.99)\\ M_{\alpha\beta} &= \epsilon_{aa'}\epsilon_{bb'}\phi_\alpha^{ab}\phi_\beta^{a'b'} && (5.100)\\ Z_{ij} &= \epsilon_{aa'}\epsilon_{bb'} Q_i^a \phi_\alpha^{a'b'} Q_j^b && (5.101) \end{aligned}$$

where Q are fundamental and ϕ are adjoint fields. $\alpha, \beta = 1.., N_{\rm a}, i, j = 1.., 2N_{\rm F}, a, b = 1, 2$. We define the quantity,

$$\Gamma_{\alpha\beta}(M, X, Z) = M_{\alpha\beta} + {\rm Tr}_{2N_{\rm F}}(Z_\alpha X^{-1} Z_\beta X^{-1}) \tag{5.102}$$

which we will use to write the prepotential as follows,

$$\begin{aligned} W_{N_{\rm F},N_{\rm a}}(M, X, Z) &= (b_1 - 4)\left(\Lambda^{-b_1} PfX(\det(\Gamma_{\alpha\beta}))^2\right)^{\frac{1}{4-b_1}} && (5.103)\\ &\quad + {\rm Tr}_{N_{\rm a}}\tilde{m}M\frac{1}{2}{\rm Tr}_{2N_{\rm F}}mX && (5.104)\\ &\quad + \frac{1}{\sqrt{2}}{\rm Tr}_{2N_{\rm F}}\lambda^\alpha Z_\alpha. && (5.105) \end{aligned}$$

This respects the necessary symmetries and can be checked semiclassically. Take the case $N_{\rm a} = 1$, $N_{\rm F} = 2$. The equations of motion from minimizing the superpotential are:

$$\frac{\partial W_{2,1}}{\partial M} = \frac{\partial W_{2,1}}{\partial X} = \frac{\partial W_{2,1}}{\partial Z} = 0 \tag{5.106}$$

which imply:

$$\begin{aligned} \tilde{m} &= 2\Lambda^{-1}(PfX)^{\frac{1}{2}} && (5.107)\\ m &= R^{-1}(X^{-1} - 8\Gamma^{-1}X^{-1}(ZX^{-1})^2) && (5.108)\\ \frac{1}{\sqrt{2}}\lambda &= 4R^{-1}\Gamma^{-1}X^{-1}ZX^{-1} && (5.109) \end{aligned}$$

where

$$R^{-1} \equiv \Lambda^{-}(PfX)^{\frac{1}{2}}\Gamma\,, \quad X \equiv \frac{1}{2}\Gamma. \tag{5.110}$$

The following equations are then obeyed:

$$X^3 - MX^2 + bX - \frac{1}{128}(c - 8M)\tilde{c} = 0 \tag{5.111}$$

$$X^2 - 2MX + b = 0. \tag{5.112}$$

Taking y and y' to vanish we can compare with the elliptic cure,

$$y^2 = x^3 + ax^2 + bx + c. \tag{5.113}$$

One can therefore identify the parameters as:

$$a = -M \qquad b = -\frac{\alpha}{4} + \frac{\Lambda^2}{4} Pfm \tag{5.114}$$

$$c = \frac{\alpha}{8}(2M + \mathrm{Tr}(\mu^2)) \qquad \alpha \equiv \frac{\Lambda^4}{16}\,, \quad \mu \equiv \lambda^{-1} m. \tag{5.115}$$

Identifying the modular parameter of the torus from the elliptic equation involves standard techniques in algebraic geometry. This modular parameter will then be the effective coupling of the theory.

Some comments:

- Some points in moduli space when $2 + N_{\mathrm{F}} = 4$, are degenerate vacua which are possibly non-local with respect to each other. These are Argyres Douglas points [71];
- As you move in moduli space monopoles turn smoothly into dyons and electric charge. This is an indication of the Higgs/confinement complementarity;
- These techniques may be extended to obtain curves for other more complicated groups.

We now rexamine some special properties of the region $\frac{3}{2}N_{\mathrm{C}} < N_{\mathrm{F}} < 3N_{\mathrm{C}}$.

5.10 Superconformal invariance in $d = 4$

For the case of $\frac{3}{2}N_{\mathrm{C}} < N_{\mathrm{F}} < 3N_{\mathrm{C}}$, at the center of moduli space when all expectation values vanish, it is claimed that the theory is described by a non-trivial conformal field theory [73]. There are several motivations for reaching this conclusion. Examine for example, the exact β function:

$$\beta(g) = -\frac{g^3}{16\pi^2} \frac{3N_{\mathrm{C}} - N_{\mathrm{F}} + N_{\mathrm{F}}\gamma(g^2)}{1 - \frac{g^2}{8\pi^2} N_{\mathrm{C}}} \tag{5.116}$$

where $\gamma(g^2)$ is the anomalous dimension given perturbatively by:

$$\gamma(g^2) = -\frac{g^2}{8\pi^2}\frac{N_{\rm C}^2 - 1}{N_{\rm C}} + O(g^4)\ . \tag{5.117}$$

If one now considers a limit where $N_{\rm F}, N_{\rm C}$ are both taken to be large but their ratio is kept fixed then the fixed point of the β function, $\beta(g^*) = 0$, is at

$$g^{*2} = \frac{1}{N_{\rm C}}\frac{8\pi^2}{3}\epsilon\ , \tag{5.118}$$

where $\epsilon = 3 - \frac{N_{\rm F}}{N_{\rm C}} << 1$. Since the coupling at the fixed point is proportional to $\epsilon << 1$, both the existence of a fixed point and the perturbative evaluation for γ is justified. The anomalous dimension at the fixed point is:

$$\gamma(g^*) = 1 - \frac{3N_{\rm C}}{N_{\rm F}}\ \cdot \tag{5.119}$$

The dimension of the $Q\tilde{Q}$ is

$$D(Q\tilde{Q}) = 2 + \gamma = 3\frac{N_{\rm F} - N_{\rm C}}{N_{\rm F}}\ \cdot \tag{5.120}$$

This observatoin will lead to interesting conclusions.

In $d = 2$ the conformal group is infinite dimensional and so provides very powerful constraints on the theory. In $d = 4$ the conformal group is finite dimensional and so the conformal symmetry does not constrain the theory in the same way [74, 75]. Nevertheless there are still interesting properties of $d = 4$ conformal theories arising from the conformal invariance. One can prove that the scaling dimension of a field is bounded by its R-charge as follows:

$$D \geq \frac{3}{2}|R| \tag{5.121}$$

where D is the scaling dimension and R is the R-charge. The bound is saturated for (anti-)chiral fields. Consider the operator product:

$$O_1(x^1)O_2(x^2) = \sum_i O_{12}^i f^i(x^1 - x^2). \tag{5.122}$$

Generally, the dimension of the product of the operator appearing on the righthand side of equation (5.122) is not a sum of the dimensions of the two operators appearing on the left hand side of equation (5.122). The product contains a superposition of operators with different dimensions. For chiral fields the situation is simpler. Since the R-charge is additive and chiral fields saturate the bound (5.121)

$$R(O^i_{12}) = R(O_1(x^1)) + R(O_2(x^2)) \Rightarrow D(O_1O_2) = D(O_1) + D(O_2) \tag{5.123}$$

and $f^i(x^1 - x^2)$ are thus all constants. By (5.123) O^i_{12} is also a chiral operator. The closure property of the chiral fields under the operator product expansion leads to the name "chiral ring".

There are reasons to expect that at the fixed point, the infra-red non-anomalous R-charge equals the non-anomalous R-charge of the ultra violet. At the fixed point of the dimension of $Q\tilde{Q}$ is given by, see equation (5.120):

$$D(Q\tilde{Q}) = \frac{3}{2}R(Q\tilde{Q}) = 3\frac{N_\mathrm{F} - N_\mathrm{C}}{N_\mathrm{F}}\,. \tag{5.124}$$

The dimension of the Baryon and anti-Baryon are:

$$D(B) = D(\tilde{B}) = \frac{3N_\mathrm{C}(N_\mathrm{F} - N_\mathrm{C})}{2N_\mathrm{F}}\,. \tag{5.125}$$

For unitary representations (of spin$\,=0$) fields, where I is the identity operator and O is an operator $\neq I$,

$$D(I) = 0\,, \quad D(O) \geq 1. \tag{5.126}$$

For a free field the bound is saturated,

$$D(O) = 1. \tag{5.127}$$

When $N_\mathrm{F} = \frac{3}{2}N_\mathrm{C}$, the operator $Q\tilde{Q}$ becomes free, since $D(Q\tilde{Q}) = 1$. For $N_\mathrm{C} + 1 < N_\mathrm{F} < \frac{3}{2}N_\mathrm{C}$ it appears that $D(Q\tilde{Q}) < 1$ which is forbidden. This is an indication that one is using the wrong degrees of freedom and a dual description is required. This will be elaborated later.

Non-trivial superconformal $N = 2$ theories in $d = 4$ occur for $N = 2$, $SU(3)$ theories without matter or $SU(2)$ theories with matter [76]. The key point is that a non-trivial conformal theory with vector fields contains both massless electric and magnetic excitations (these are mutually non-local).

The definition of a primary state is that it is annihilated by the generator of special conformal transformations, K. The descendents are obtained by acting on primary states with momentum operators. The Lorentz group decomposes into $SU(2)_\mathrm{L} \oplus SU(2)_\mathrm{R}$ with charges $j, \tilde{j}$ respectively. An operator with non-zero spin will be carry $j, \tilde{j}$ charge as well as D the dimension.

Unitary chiral primaries obey: $j\tilde{j} = 0$. The dimension of a chiral filed is:

$$D(O) \geq j + \tilde{j} + 1 \tag{5.128}$$

as compared with a non-chiral operator:

$$D(O) \geq j + \tilde{j} + 2. \tag{5.129}$$

For a free field:

$$D(O) = j + \tilde{j} + 1. \tag{5.130}$$

This generalizes the previous results for scalar chiral fields. $F_{\mu\nu}$ decomposes into self-dual and anti self-dual parts that form irreducible representations of $SU(2)_{\mathrm{L}} \oplus SU(2)_{\mathrm{R}}$.

$$F = F^+ + F^-, F^\pm = F \pm *F. \tag{5.131}$$

One can show that there are states associated with the conserved currents:

$$J^\pm_\mu = \partial^\nu F^\pm_{\mu\nu}. \tag{5.132}$$

These states are the decendents of $F^\pm$, obtained by applying the momentum operator P on the chiral field, $F^\pm$. The norm is then calculated using the following: the commutator of the $d = 4$ conformal algebra,

$$[P^{\alpha\dot{\alpha}}, K^{\beta\dot{\beta}}] = \frac{i}{2} M^{\alpha\beta}\epsilon^{\dot{\alpha}\dot{\beta}} + \bar{M}^{\dot{\alpha}\dot{\beta}}\epsilon^{\alpha\beta} + D\epsilon^{\alpha\beta}\epsilon^{\dot{\alpha}\dot{\beta}} , \tag{5.133}$$

where the undotted indices are $SU(2)_{\mathrm{L}}$ indices, dotted indices are $SU(2)_{\mathrm{R}}$ indices and $\frac{i}{2}M^{12} = J^3$; the Hermitian conjugate relation,

$$(P^{\alpha\dot{\alpha}})^+ = -\epsilon^{\alpha\beta}\epsilon^{\dot{\alpha}\dot{\beta}} K^{\beta\dot{\beta}}; \tag{5.134}$$

and that K annihilates primary states. After some algebra this produces,

$$||J^\pm\rangle| = 2(D-2). \tag{5.135}$$

For $D = 2$, these are null states and $F^\pm$ is free. If F is not free and $D > 2$, then

$$||J^\pm\rangle| > 0 \tag{5.136}$$

and therefore both,

$$J^{\mathrm{e}} \equiv J^+ + J^- \ , \quad J^m \equiv J^+ - J^- \tag{5.137}$$

are non-vanishing [76].

At the point in moduli space (discussed previously) where there are both electric and magnetic charges, called the Argyres Douglas point [71], this condition is met and it is possible to have a non-trivial conformal field theory with spin one particles.

Other conformal theories occur for $N = 1$ supersymmetry when $N_{\rm a} = 3$ and the couplings are appropriately tuned; this is actually a deformed $N = 4$ theory. For $N = 2$ supersymmetry, the theory is conformal when $N_{\rm F} = 2N_{\rm C}$ and the one loop β function vanishes; this is known to be an exact result [77].

One can deform $N = 4$ theories with marginal operators such that the global symmetries are broken but the theory remains conformal. Hence, consider,

$$\mathcal{L} = \mathcal{L}_0 + \sum_i g_i O^i \tag{5.138}$$

where O^i are the set of operators with dimension 4.

Consider $N = 1$ with $N_{\rm a} = 3$ and an interaction for the adjoint fields X_1, X_2, X_3:

$$L_{\rm int} = hX_1X_2X_3. \tag{5.139}$$

When $h = g$, g being the gauge coupling, the theory has full $N = 4$ supersymmetry.

For a general $h\phi_1..\phi_n$ perturbation one has

$$\beta_h = h(\mu)\left(-d_w + \sum_{k=1}^{n}\left(d\,(\phi_k) + \frac{1}{2}\gamma(\phi_k)\right)\right) \tag{5.140}$$

where d_w is the engineering dimension of the composite perturbing term, $d(\phi_k)$ is the engineering dimension of the field, ϕ_k and $\gamma(\phi_k)$ is the associated anomalous dimension of the field ϕ_k. By symmetry all fields have the same anomalous dimension and so $\gamma(\phi_k)$ is independent of k and so is denoted as simply $\gamma(\phi)$. In the case at hand

$$d_w = 3\ , \qquad \sum_{k=1}^{3} d(\phi_k) = 3\ , \qquad \sum_k \frac{1}{2}\gamma(\phi_k) = \frac{3}{2}\gamma \tag{5.141}$$

and therefore

$$\beta_h = h(\mu)\frac{3}{2}\gamma(\phi) \tag{5.142}$$

and

$$\beta_g = -f(g(\mu))((3C_2(G) - \sum_k T(R_k)) + \sum_k T(R_k)\gamma(\phi_k)) \tag{5.143}$$

$$= -f(g(\mu))3T(R)\gamma(\phi) \tag{5.144}$$

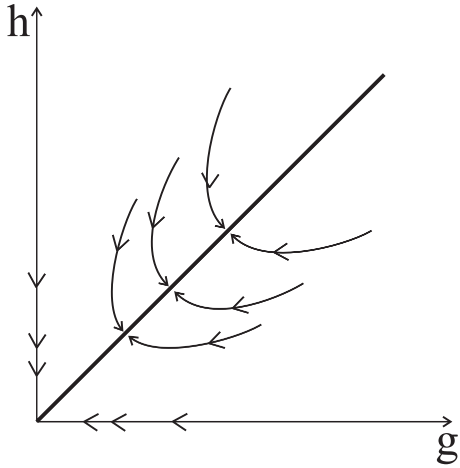


Fig. 12. RG flow to the fixed line.

which is non vanishing. $C_2(G)$ is the second Casimir of the group G and $T(R_k)$ is associated with the representation of ϕ.

Therefore, both β_h and β_g vanish if γ vanishes. This means there is a fixed line, rather than a fixed point, where $g = h$ and the supersymmetry is enhanced to $N = 4$. (This is different from the more generic situation where the β functions are not related and there are isolated fixed points.) This fixed line is infra-red stable. Also the relations between relevant deformations follow a similar pattern [78].

At the fixed line, one can ask if the $N = 4$ theory has any discrete symmetries that relate theories with different moduli. The answer is yes [79]. There is a great deal of evidence that $N = 4$ theories possess an $SL(2,\ Z)$ duality [80] that identifies theories with coupling τ given by equation (5.2):

$$\tau \to \frac{a\tau + b}{c\tau + d}\,, \qquad ad - bc = 1\ ,\quad a, b, c, d \in Z. \tag{5.145}$$

To conclude, supersymmetric gauge theories have a very rich phase structure and many outstanding dynamical issues can be discussed reliably in the supersymmetric arena that are hard to address elsewhere.

6 Comments on vacuum energies in scale invariant theories

The puzzle to be addressed in this section is that of the cosmological constant problem [81, 82]. The problem was originally stated as follows. Given that there is an effective scale below which the physics is known then (such as QED or the standard model) integrating out the more energetic degrees of freedom above this scale leads to a vacuum energy that is proportional to the cut off. For either choice of the above physics, the vacuum energy calculated in such a manner however would give a cosmological constant many orders of magnitude above the observed value. One should observe some important caveats in this argument [83,84] and also [78,85]. When one writes down a low energy effective action for a spontaneously broken theory one should respect all the symmetries that appear in the original action. If the original theory is scale invariant then the effective action of the spontaneously broken theory should reflect this symmetry. The consequence of scale invariance is a zero vacuum energy whether or not scale invariance is spontaneously broken.

As an example consdier $N = 4$ super Yang-Mills in four dimensions. The potential is given by:

$$V = [\phi, \phi]^2 \tag{6.1}$$

thus the theory has flat directions. This result is exact to all orders in perturbation theory and non-perturbatively. Giving an expectation value to a field will break the scale invariance spontaneously as well as generically breaking the gauge symmetry down to $U(1)^r$ where r is the rank of the unbroken gauge group. An analysis of the spectrum shows that a gauge singlet particle emerges, which is the dilaton (the Goldstone Boson associated with spontaneously broken scale invariance). The vacuum energy remains zero.

Thus provided there is a translationally invariant ground state, global supersymmetry can't be spontaneously broken whether or not scale invariance is spontaneoulsy broken. The presence of a ground state is crucial; recall the example discussed in quantum mechanics where the action scale invariance and supersymmetry were spontaneously broken but the system had no ground state.

The next example is the $O(N)$ model in three dimensions, described by the Lagrangian,

$$\mathcal{L} = \partial_\mu \vec{\phi} \cdot \partial^\mu \vec{\phi} - g_6 \left(|\vec{\phi}|^2 \right)^3 \tag{6.2}$$

where the fields $\vec{\phi}$ are in the vector representation of $O(N)$. In the limit, $N \to \infty$,

$$\beta_{g_6} = 0 \tag{6.3}$$

($\frac{1}{N}$ corrections break the conformality). g_6 is a modulus. An effective potential can be written down for the $O(N)$ invariant field,

$$\sigma = |\vec{\phi}|^2 \tag{6.4}$$

the effective potential is:

$$V(\sigma) = f(g)|\sigma|^3. \tag{6.5}$$

One has the following possibilities, summarized in the table below:

$$\begin{array}{cccccc} & \langle\sigma\rangle & \text{S.B.} & \text{masses} & V \\ f(g) > 0 & 0 & \text{No} & 0 & 0 \\ f(g) = 0 & \langle\sigma\rangle = 0 & \text{No} & 0 & 0 \\ f(g) = 0 & \langle\sigma\rangle \neq 0 & \text{Yes} & \langle\sigma\rangle,\, 0 & 0 \end{array} \tag{6.6}$$

where S.B. indicates spontaneous symmetry breaking of scale invariance and V is the vacuum energy. (For $f(g) < 0$ the theory is unstable). Note, the vacuum energy always vanishes. To summarize, in the situation where $\langle\sigma\rangle \neq 0$ and the scale invariance is spontaneously broken, one could write down the effective theory for energy scales below $\langle\sigma\rangle$ and integrate out the physics above that scale. The vacuum energy remains zero however and not proportional to $\langle\sigma\rangle^3$ as is the naive expectation.

The final example will be the $O(N) \times O(N)$ model with two fields in the vector representation of $O(N)$, with Lagrangian:

$$\begin{aligned} \mathcal{L} &= \partial_\mu\vec{\phi_1}\cdot\partial^\mu\vec{\phi_1} + \partial_\mu\vec{\phi_2}\cdot\partial^\mu\vec{\phi_2} - \lambda_{6,0}\left(|\vec{\phi_1}|^2\right)^3 & (6.7)\\ &\quad -\lambda_{4,2}\left(|\vec{\phi_1}|^2\right)^2\left(|\vec{\phi_2}|^2\right) - \lambda_{2,4}\left(|\vec{\phi_1}|^2\right)\left(|\vec{\phi_2}|^2\right)^2 & (6.8)\\ &\quad -\lambda_{0,6}\left(|\vec{\phi_2}|^2\right)^3 . \end{aligned}$$

Again the β functions vanish in the strict $N \to \infty$ limit. There are now two possible scales, one associated to the break down of a global symmetry and another with the break down of scale invariance. The possibilities are summarized by the table below:

$$\begin{array}{cccccc} O(N) & O(N) & \text{scale} & \text{massless} & \text{massive} & V \\ + & + & + & \text{all} & \text{none} & 0 \\ - & + & - & (N-1)\pi' s, D & N, \sigma & 0 \\ + & - & - & (N-1)\pi' s, D & N, \sigma & 0 \\ - & - & - & 2(N-1)\pi' s, D & \sigma & 0. \end{array} \tag{6.9}$$

Again, in all cases the vacuum energy vanishes. Assume a hierarchy of scales where the scale invariance is broken at a scale much above the scale at which the $O(N)$ symmetries are broken. One would have argued that one would have had a low energy effective Lagrangian for the pions and dilaton with a vacuum energy given by the scale at which the global symmetry is broken. This is not true, the vacuum energy remains zero.

It was proposed that the underlying physics of nature is scale invariant and the scale invariance is only removed by spontaneous symmetry breaking [83, 84] and also [78, 85]. This would explain the vanishing of the cosmological constant. (The data in 2001 seems to indicate the presence of a small cosmological constant.) A key problem with this scenario however is that we do not observe a dilaton in nature that would be expected if scale invariance is spontaneously broken. The hope is that this dilaton may somehow be given a mass with a further independent (Higgs like?) mechanism.

The total energy of the universe can be augmented even in a scale invariant theory. Such theories may have many superselection sectors (such as magnetic monopoles in $N = 4$ super Yang-Mills). In each sector the total energy will differ from zero by the energy of these particles. If there are many of them they may mimic a very small cosmological constant [87].

7 Supersymmetric gauge theories and string theory

We now view the supersymmetric gauge theories from a different point of view. We obtain them as the low energy limit of various string backgrounds. Many properties of gauge theories are obtained in this fashion [93, 94]. In the following discussion we will have as our goal to:

- Construct:
  - a $D = 4$ dimensional effective theory;
  - with $N = 1$ supersymmetry (SUSY);
  - with $U(N_c)$ gauge symmetry;
  - with $SU(N_F) \times SU(N_F) \times U(1)$ global symmetry.
- Identify dualities in a pedestrian way.

Our tools for this project will be... comic strips [95]. The unabridged original novel, from which this "comics illustrated" is derived is yet to be written. The scaffolding for this construction will be extended objects called branes.

7.1 Branes in string theory

Branes are extended object solutions which emerge non perturbatively in string theory in a very similar way that solitons emerge in field theory. Magnetic monopoles and vortices are examples of solitonic configurations in gauge theories. What are they good for? First, there existence answers a yearning to search for more than meets the eye (or the equations). A yearning, which seems to be engraved in at least part of our community. It, of course, also answers positively the important question of the existence of a magnetic monopole, as well as that of other interesting topological excitations. These excitations are often very heavy and have little direct impact on the low-energy dynamics. However, there are circumstances, in which they can condense and thus take over the control and drive the infra-red dynamics. In the context we will discuss here, the branes will basically serve as regulators for some field theories, playing the same role as string sizes and lattice cut-offs. In lattice gauge theories, such regulators have granted a very rapid access to identifying non-perturbative features, such as confinement, in strong-coupling approximations. Similar consequences will occur here. In string theory, at this stage, the regulator will also learn how to behave from the very theory it regulates.

7.2 Branes in IIA and IIB string theories

For some years it has been known that solitons exist in the low-energy effective theory of superstring theory/supergravity [96]. To appreciate that, we first review the spectrum of massless particles in various string theories [97]. In the closed bosonic string theory, the spectrum consists of a graviton, an anti-symmetric tensor, and a dilaton, denoted by $G_{\mu\nu}$, $B_{\mu\nu}$, Φ, $\mu = 1, 2, \ldots, 24$, respectively. The massless spectrum of the open bosonic string consists of photons A_μ. The bosonic sector of type-IIA theory consists of two sectors. The first is called NS-NS sector and consists of the same spectrum as that of the bosonic closed string theory, namely $G_{\mu\nu}$, $B_{\mu\nu}$, Φ. The other sector is called the RR sector, consisting of one-form and three-form vector potentials, denoted by A_μ and $A_{\mu\nu}$. To each of them is associated an "electric field", carrying two and four indices, respectively. In type-IIB the NS-NS sector also consists of $G_{\mu\nu}$, $B_{\mu\nu}$, Φ. The RR sector consists of vector potentials which are 0-forms, 2-forms, and 4-forms. In addition, duality relations in ten dimensions between the electric field $E_{\mu_1,\ldots,\mu_{p+1}}$, derived from the p-form A, and the electric field $E_{\mu_1,\ldots,\mu_{8-p-1}}$, derived from the (6-$p$)-form A lead to further vector potentials, a 5-form in type-IIA and 4-forms and 6-forms in RR sector of type-IIB. Gravitons and the other particles in the NS-NS sector have a perturbative string realization. What about the objects in the RR sector? First we note, that in the NS-NS sector, there exist solitons, discovered in supergravity, whose tension T is: $T_{\mathrm{NS-NS}} = \frac{1}{g_{\mathrm{s}}^2 l_{\mathrm{s}}^6}$, where g_{s} is the string coupling and l_{s} is the string scale. These solitons spread over five space and one time dimensions. The exact background corresponding to this configuration is not yet known as the dilaton background seem to contain singularities. In the RR sector, there are solitons as well [92]. The $p+1$ dimensional solitons have a tension which is: $T_{\rm RR} = \frac{1}{g_{\rm s} l_{\rm s}^{p+1}}$. These are called D-branes. The D stands for Dirichlet. It turns out that open strings, obeying appropriate Dirichlet boundary conditions, may end on these branes. The solitonic sector of closed string theory contains configurations of open strings. One may inquire as to the effective theory on the six-dimensional world volume of the NS-NS soliton, the NS5 brane. One may also inquire what are the supersymmetry properties of that theory. It is easier to answer the second question. The supersymmetry algebra:

$$\begin{aligned}\{Q^i_\alpha, \bar{Q}_{\beta\, j}\} &= 2\sigma^\mu_{\alpha\beta} P_\mu \delta^i{}_j \\ \{Q^i_\alpha, Q^j_\beta\} &= \{\bar{Q}_{\alpha\, i}, \bar{Q}_{\beta\, j}\} = 0\end{aligned} \tag{7.1}$$

contains on its right-hand side the generators of translation. A soliton, existing in ten dimensions, whose extension is in $p+1$ space-time dimensions breaks translational invariance in 9-p directions. Thus it is not clear if any of the supersymmetric generators, residing in the left-hand side of the algebra, will survive intact. It turns out that, in flat space-time, the system has 32 SUSY charges. In the presence of the NS5 configuration sixteen (half) of the SUSY charges survive. They are of the form:

$$\epsilon_{\rm L} Q_{\rm L} + \epsilon_{\rm R} Q_{\rm R} \tag{7.2}$$

where $\epsilon_{\rm L}$ and $\epsilon_{\rm R}$ obey the constraints:

$$\begin{aligned}\epsilon_{\rm L} &= \Gamma^0 \ldots \Gamma^5 \epsilon_{\rm L} \\ \epsilon_{\rm R} &= \pm\Gamma^0 \ldots \Gamma^5 \epsilon_{\rm R},\end{aligned} \tag{7.3}$$

where the sign $\pm$ corresponds to type-IIA and type-IIB, respectively, and the Γ-matrices are respective ten-dimensional Dirac Γ-matrices.

The answer to the first question is more complicated. The massless sector on the NS5 in the type-IIA theory is a chiral system consisting of a self-dual anti-symmetric vector potential, five scalars, and their supersymmetric partners. The massless sector on the NS5 in the type-IIB theory is a non-chiral system consisting of spin-1 particles, spin-0 particles, and their supersymmetric partners.

Let us return now to the RR sector. The D-brane solitons in this sector are denoted Dp, where p denotes the number of dimensions of the branes world volume (the spatial volume in which the brane extends). This is our first comic strip (Fig. 13).

The coordinates $x_0, x_1, \ldots, x_p$ are unconstrained and span the brane's space-time $p+1$ dimensional world volume. The other coordinates $x_{p+1}, .., x_9$ are fixed. The translational non-invariance of the Dp-brane

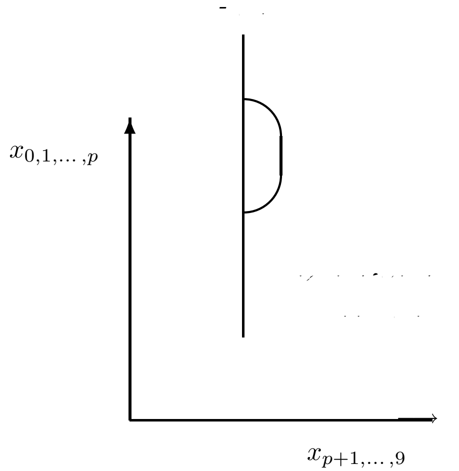


Fig. 13. A Dp brane with an open string ending on it.

reduces the number of SUSY generators by half for any value of p. For Dp-branes ϵ_{R} and ϵ_{L} in the surviving supersymmetric charge (Eq. (7.2)) obey the constraint

$$\epsilon_{\mathrm{L}} = \Gamma^0 \dots \Gamma^p \epsilon_{\mathrm{R}}. \tag{7.4}$$

Thus, for any p, there are sixteen surviving supercharges. Note that, applying naively Newton's law, one can estimate the effective gravitational coupling G_{N} "M". G_{N}, Newton's constant, is proportional to g_{s}^2. "M" is proportional to $\frac{1}{g_{\mathrm{s}}}$ for Dp branes and to $\frac{1}{g_{\mathrm{s}}^2}$ for the NS5 brane. Thus G_{N} "M" vanishes at weak coupling for Dp branes, unaltering the large distance geometry.

7.3 The effective field theory on branes

The effective theory on the Dp-brane can be identified in more detail in this case. It can be shown to contain, in addition to the sixteen conserved supercharges, 9-p massless scalars (corresponding to the Goldstone Bosons resulting from the spontaneous breaking of some of the translational invariances), spin-1 massless particles, and their spin-$\frac{1}{2}$ superpartners. The effective theory is invariant under local $U(1)$ gauge transformations (Fig. 14).

Similarly, one can construct a configuration containing N_{C} parallel Dp-branes. This theory still contains sixteen conserved supercharges, and the gauge symmetry has increased to $U(1)^{N_{\mathrm{C}}}$. The next comic strip describes this configuration for $N_{\mathrm{C}} = 4$ (Fig. 15).

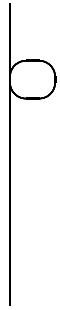


Fig. 14. A massless state propagating on a Dp brane.

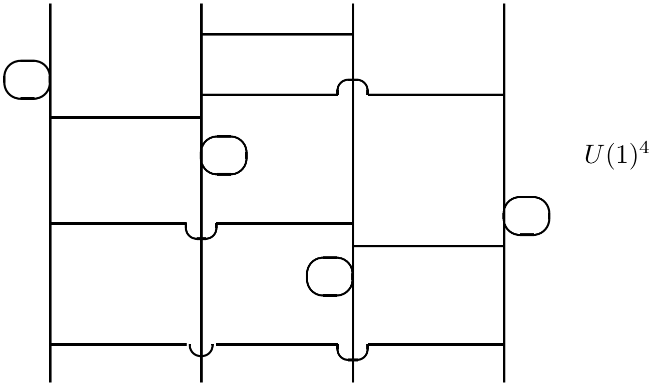


Fig. 15. The effective field theory describing this D4 brane configuration is a $U(4)$ gauge theory, spontaneously broken down to $U(1)^4$. Four massless states (corresponding to open strings ending on the same brane) and six massive states (corresponding to open strings ending on different branes) are shown.

It turns out that the masses of the particles can be directly associated with the minimal distance between the end points of the strings stretched between different branes. For example, a string stretched between the i-th and j-th brane represents a particle which has a mass $m_{ij} = \frac{1}{l_s^2}|\vec{x}_i - \vec{x}_j|$, where $\vec{x}_i$ represents the value of the coordinates at which the brane is set. A parameter in field theory, the mass of a particle has a very simple geometrical meaning. Imagine now bringing the parallel branes together (Fig. 16). This, according to the above relation, will lead to the emergence of massless particles. In fact, one can show that these massless particles can enhance the gauge symmetry all the way from $U(1)^4$ to $U(4)$.

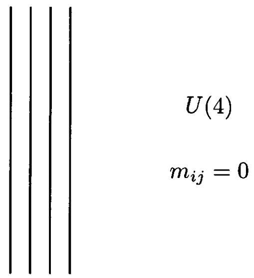


Fig. 16. Four parallel D-branes, piled on top of each other. The sixteen massless states formed are not shown in the figure.

The minimal gauge symmetry for N_{C} Dp-branes is $U(1)^{N_{\mathrm{C}}}$. The maximal gauge symmetry is $U(N_{\mathrm{C}})$. This has a correspondence in SUSY field theory. The Lagrangian describing the bosonic sector of a supersymmetric gauge theory with sixteen supercharges is of the form:

$$\mathcal{L} = \frac{1}{g^2_{\mathrm{YM},p+1}}\left(\mathrm{Tr}F^2_{\mu\nu} + \frac{1}{l_{\mathrm{s}}^4}D_\mu X^{\mathrm{I}} D^\mu X_{\mathrm{I}}\right) + \frac{1}{l_{\mathrm{s}}^8 g^2_{\mathrm{YM},p+1}}\mathrm{Tr}[X^I, X^J]^2, \quad (7.5)$$

where

$$D_\mu X^I \quad = \quad \partial_\mu X^I - i[A_\mu, X^I] \quad (7.6)$$

X^{I} is a scalar field in the adjoint representation of the gauge group and $g^2_{\mathrm{YM},p+1}$ is the effective Yang-Mills coupling in $p+1$ dimensions, it is proportional to g_{s}. The system has indeed the same particle content and gauge symmetries as can be inferred from Figure 15. Moreover, the Higgs mechanism in the presence of scalars in the adjoint representation is known to conserve the rank of the group. Indeed, only such residual gauge symmetry groups that preserve the rank are allowed according to the comic strip. The form of the potential term appearing in equation (7.5) is generic for supersymmetric theories. The classical potential is flat and allows for an infinite set of vacua, parameterized by those expectation values of the scalar fields X^{I} for which the potential term vanishes. For example, in a $N = 1$ supersymmetric gauge theory, containing two multiplets with opposite electric charges, the potential is given by:

$$V = (q^\dagger q - \tilde{q}^\dagger \tilde{q})^2. \quad (7.7)$$

It is a property of supersymmetric theories that the flat potential is retained to all orders in perturbation theory. For some theories with a small number of SUSY charges, this result is modified non-perturbatively [15, 52]. For theories with sixteen SUSY charges, the potential remains flat also non-perturbatively. The vacua of the system consist of those expectation values for which $\langle q \rangle = \langle \tilde{q} \rangle \neq 0$. In each of these vacua, the system will have massless scalar excitations. They are denoted moduli and their number is called the dimension of moduli space.

Returning to the configuration of N_{C} Dp-branes, we have seen that $U(1)$ symmetries remain unbroken and therefore the theory is said to be in the Coulomb phase. The Coulomb phase is that phase of gauge theory for which the force between both electric and magnetic charges is a Coulomb force. The number of massless particles is at most N_{C} (in complex notation). This is again apparent from the geometry of the open strings ending on the brane configuration. The expectation values of the Higgs fields in the adjoint representation can be shown to have themselves a very transparent geometrical meaning:

$$\vec{x}_i = \langle \vec{X}_{ii} \rangle \quad (i = 1, \ldots, N_{\mathrm{C}}). \tag{7.8}$$

$\vec{x}_i$ on the left-hand side of the equation denotes the location of the i-th brane, $\vec{X}_{ii}$ on the right-hand side of the equation denotes the component of the Higgs field in the i-th element of the Cartan subalgebra of $U(N_{\mathrm{C}})$. In this context, the mass formula

$$m_{ij} = \frac{1}{l_{\mathrm{s}}^2} |\vec{x}_i - \vec{x}_j| \tag{7.9}$$

is just the usual mass obtained by the Higgs mechanism. To keep fixed the mass of these "W" particles in the limit of the decoupling of the string states ($l_{\mathrm{s}} \to 0$), the separations between the branes should vanish themselves in that limit, that is, these separation should be sub-stringy. In this limit, one cannot resolve the N_{C} different world volumes, so the theory is perceived as a $U(1)^{N_{\mathrm{C}}}$ gauge theory on a single $p+1$ dimensional world volume. We are by now well on our way to obtain that brane configurations will help accomplish our goals, namely the brane configuration leading to an effective $D = 4$, $N = 1$ supersymmetrical $U(N_{\mathrm{C}})$ gauge theory containing in addition matter fields.

7.4 Effective $D = 4$ dimensional systems with $N = 2$ supersymmetry

To obtain an effective $D = 4$ description one can either set up a single D3-brane configuration in type IIB string theory or build a more complicated configuration in type IIA string theory. It turns out that for our goals the latter is more useful. One constructs a brane configuration which has a world volume of the form $M^{3,1} \times I_{[\Delta x_6]}$ where $M^{3,1}$ is the four-dimensional Minkowski space-time and $I_{[\Delta x_6]}$ is an interval of length Δx_6. This is realized by the following configuration:

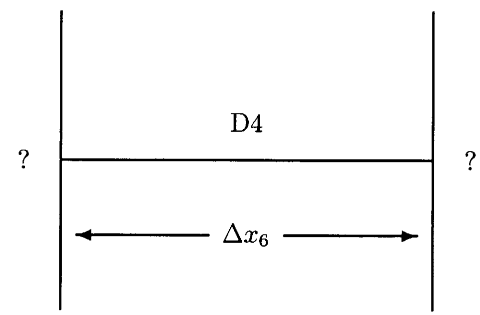


Fig. 17. The effective field theory on the D4 brane will be four-dimensional for energies much smaller than $\frac{1}{\Delta x_6}$. The content of the field theory will depend on the branes "?" on which the D4 brane ends.

In this configuration a D4-brane (whose world volume is 5 dimensional and of the type $M^{3,1} \times I_{[\Delta x_6]}$) the two branes between which the D4-brane is suspended would be chosen such that the effective field theory has $U(N_C)$ local gauge symmetry and $N = 1$ supersymmetry. The candidates for "heavier" such branes would be either NS5-branes or D6-branes. For either choice, the effective field theory, that is the field theory at energy scales much smaller than $1/\Delta x_6$, is effectively four dimensional. Before analyzing the various effective theories resulting from the different choices of the branes on which the D4-brane ends we discuss which are the allowed "vertices", that is on which branes are the D4-branes actually allowed to end. In the first example we will show that a D4 is allowed to end on a D6-brane, that is the vertex appearing in Figure 18 is allowed.

A fundamental string (F1) can by definition end on any Dp-brane, in particular in type IIB string theory it can end on a D3-brane. Performing what is called an S-duality transformation validates that also a D1-brane may end on a D3-brane in type IIB theory. The world-volume of the D3-brane is chosen to extend in the x_0, x_7, x_8, x_9 directions and that of the D1-brane extends in x_0, x_6. Establishing that the D1 may end on a D3-brane, we pause now to briefly discuss several types of useful discrete symmetries in string theory, called S- and T-dualities.

S-duality is a symmetry which is familiar already in some field theories. For example, in an $N = 4$ supersymmetric gauge theory in $D = 4$ dimensions the gauge coupling constant g is a real parameter. The field

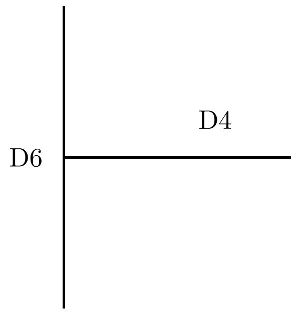


Fig. 18. A "vertex" in which a D4 brane ends on a D6 brane.

theory is finite, the gauge coupling does not run under the renormalization group and thus different values of g correspond to different theories. There is evidence that the theory with coupling g is isomorphic to the theory with coupling $1/g$. Type IIB string theory has similar properties with the string coupling playing the role of the gauge coupling. This non-perturbative symmetry, called S-duality, has a generalization involving also the value of the θ parameter in field theory and an additional corresponding field in string theory. In its implementation in field theory electric and magnetic excitations were interchanged, similarly in string theory different types of branes are interchanged under S-duality. An F1 is interchanged with a D1, a D3 is left invariant and a D5 is interchanged with an NS5. We have used some of these properties in the derivation above.

T-duality is a symmetry which has aspects peculiar to string theory [98]. In particular, a closed bosonic string theory with one compact dimension whose radius in string units is R, is identical to another bosonic string theory whose compact dimension in string units is of radius $1/R$. It is the extended nature of the string which leads to this result. The mass M of the particles depends on the compactification radius through the formula:

$$M^2 = \frac{n^2}{R^2} + m^2 R^2. \tag{7.10}$$

n/R denotes the quantized momentum of the center of mass of the string. The term $\frac{n^2}{R^2}$ is not particular to string theory, it describes also a point particle in a Kaluza-Klein compactification. The second term m^2R^2 reflects the extended nature of the string. It describes those excitations in which the closed bosonic string extends and winds around the compact dimensions m times. For a small radius R these are very low energy excitations. All in all, an interchange of n and m simultaneously with an interchange of R and $1/R$ in equation (7.10) gives an indication of how T-duality works. T-duality can be generalized to an infinite discrete symmetry and can be shown to actually be a gauge symmetry in the bosonic case. This indicates that it persists non-perturbatively. For supersymmetric string theories T-duality has some different manifestations. In particular the transformation $R \to 1/R$ maps a type IIA string theory background with radius R to a type IIB background with radius $1/R$ and [illegible]. In the presence of D-branes one naturally distinguishes between two types of compact dimensions: "longitudinal" dimensions, which are part of the world-volume of the brane, and "transverse" dimensions, those dimensions which are not part of the world-volume of the brane. A T-duality involving a longitudinal dimension will transform a Dp-brane into a D$(p-1)$-brane and will leave an NS5-brane intact. T-duality involving a transverse direction transforms a Dp-brane into a D$(p{+}1)$-brane. Its effect on a NS5-brane is more complicated and we will not need it in this lecture (Figs. 19, 20).

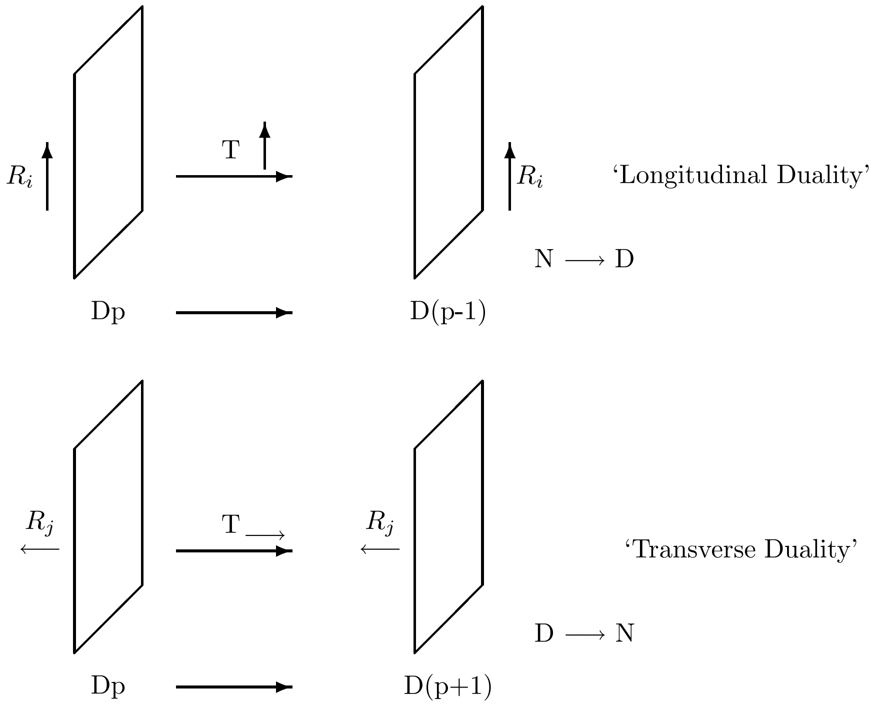


Fig. 19. T-duality acting on Dp branes. D and N denote Dirichlet and Neumann boundary conditions in the compact directions, respectively.

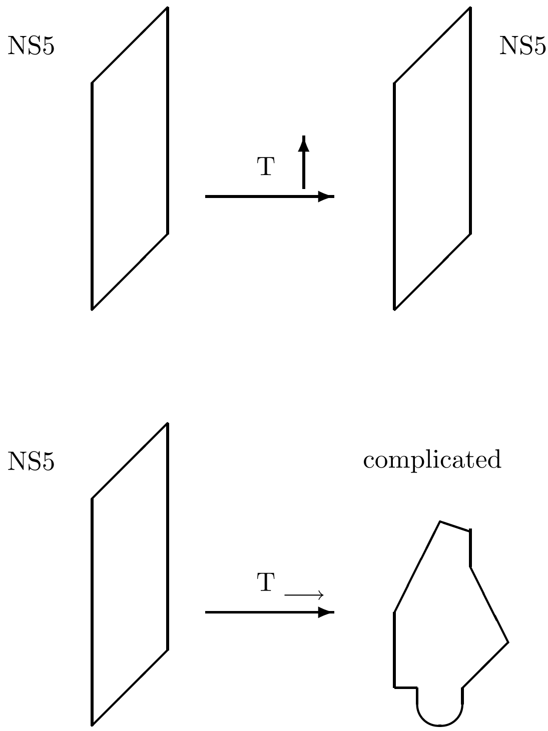


Fig. 20. T-duality acting on a NS5 brane. The detailed action of the transverse duality is not indicated.

Equipped with this information we can continue the proof of the existence of a D4 configuration ending on a D6-brane. By performing a T-duality along three directions transverse to both the D3 and the D1-branes, for example the directions x_1, x_2 and x_3, we obtain a D4-brane ending on a D6-brane. Due to the odd number of T-duality transformations, one passes from a IIB background to a type IIA background. The proof thus rests on the validity of both S- and T-duality. There are many indications that the former is correct, and there is firmer evidence of the validity of T-duality. The construction sketched in this proof shows that any Dp-brane can end on any D$(p+2)$-brane. The steps used in the proof are summarized in Figure 21.

In a somewhat similar manner one can show that a D4-brane can end on a NS5-brane (Fig. 22).

Starting from the by now established configuration of a D1-brane ending on a D3-brane in type IIB string theory, one performs T-duality along two transverse direction, x_1 and x_2 to obtain a D3-brane ending on a D5-brane.

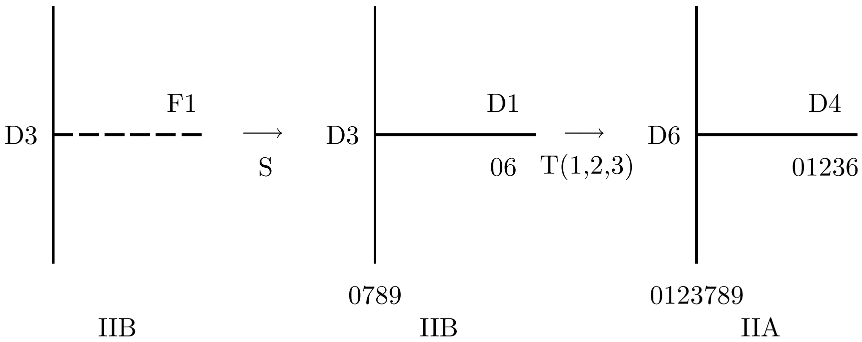


Fig. 21. A combination of S and T duality transformations, establishing that a D4 brane can indeed end on a D6 brane.

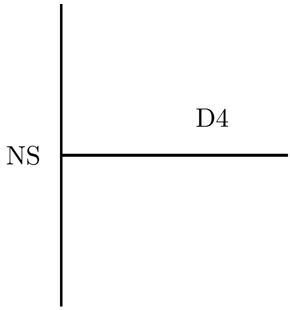


Fig. 22. A D4 brane ends on a NS5 brane.

An S-duality transformation brings us to a D3-brane ending on a NS5-brane. A T-duality along x_3, which is transverse to the newly formed D3-brane but longitudinal to the NS5-brane, leads to the desired configuration in which a D4-brane ends on a NS5-brane. This series of S- and T-duality transformations (called U-duality) is shown in Figure 23.

Let us consider first the effective low energy field theory related to the configuration in which the D4-brane is suspended between a NS5-brane and a D6-brane (Fig. 24).

The spatial extension of the various branes is summarized in the table below, where + and − denote respectively directions longitudinal and

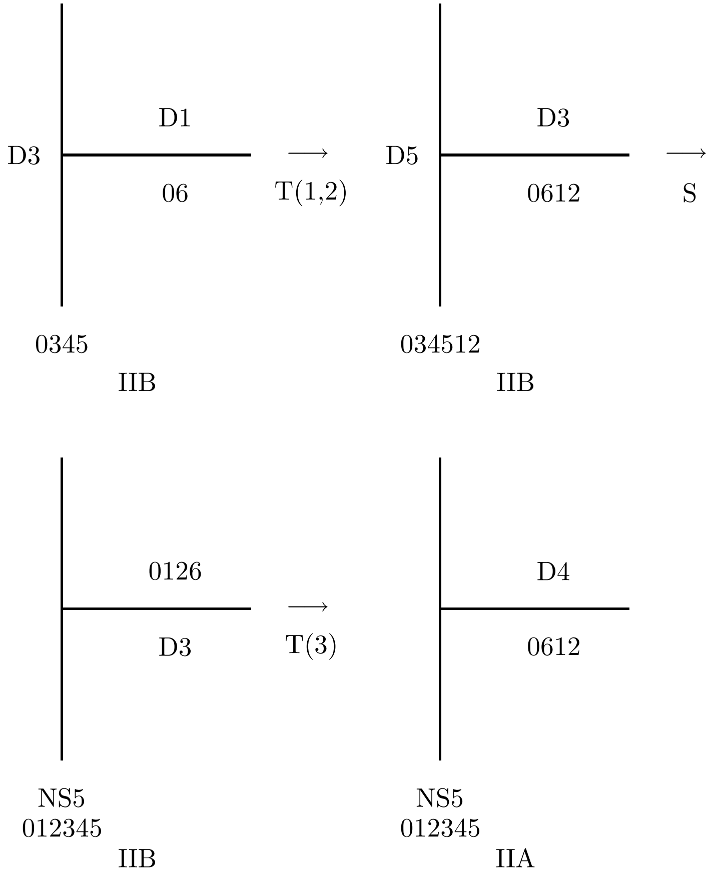


Fig. 23. S- and T-duality transformations establish that a D4 brane may end on a NS5 brane.

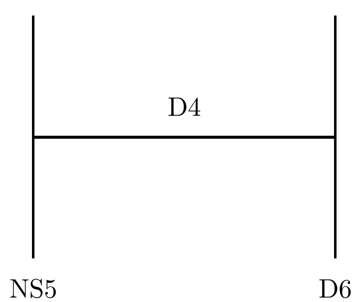


Fig. 24. A D4 brane suspended between and NS5 and D6 branes. The effective field theory contains no massless particles and is thus, at best, some topological theory.

transverse to the relevant brane, and = indicates that the brane is of finite extent in that direction.

$$\begin{array}{cccccccc} & x_{0123} & x_4 & x_5 & x_6 & x_7 & x_8 & x_9 \\ D4 & + & - & - & = & - & - & - \\ D6 & + & - & - & - & + & + & + \\ NS5 & + & + & + & - & - & - & -. \end{array} \tag{7.11}$$

The effective field theory should contain the massless degrees of freedom of the system. Massless degrees of freedom can also be identified in a geometrical manner. Each possibility to displace the D4-brane along the D6 and the NS5 branes maintaining the shape of the configuration corresponds to a massless particle. The D4-brane left on its own could be displaced along the directions x_4, x_5, x_6, x_7, x_8 and x_9. On the D6 side, the D4-brane is locked in the x_4, x_5 and x_6 directions. On the NS5 side the D4 is locked along x_6, x_7, x_8 and x_9. All in all, the D4-brane is frozen. It cannot be displaced in a parallel fashion and therefore the effective theory contains no massless particles whatsoever. Although the D4-brane would have allowed the propagation of 5-dimensional photons had it been left on its own, suspended between the D6 and the NS5 it allows no massless degrees of freedom to propagate on it. It is at best a topological field theory.

We thus turn to another attempt to build an effective four dimensional field theory along the suspended D4-brane. We now suspend it between two NS5-branes which are extended in the same directions as the NS5 of the previous example [95] (Fig. 25).

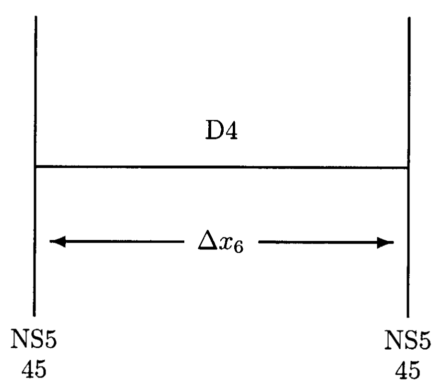


Fig. 25. A D4 brane suspended between two NS5 branes. The effective field theory is $N = 2$ SUSY $U(1)$ gauge theory.

At both ends, the D4-branes is locked in the x_6, x_7, x_8 and x_9 directions. Therefore now it can be displaced in the x_4 and x_5 directions, as shown in Figure 26.

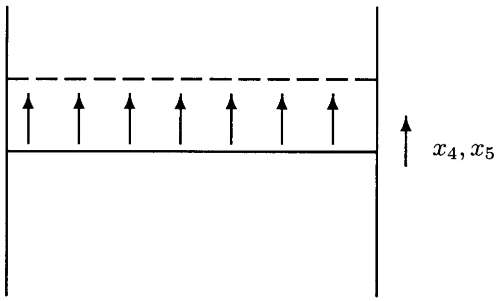


Fig. 26. The D4 brane can be parallely displaced along the x_4 and x_5 directions. This corresponds to two massless spin-0 particles appearing in the effective field theory. $N = 2$ supersymmetry implies the existence of massless spin-$\frac{1}{2}$ and spin-1 particles as well.

Thus the effective field theory contains at least two massless spin 0 particles. An analysis using (Eqs. (7.2), (7.3)) and (Eq. (7.4)) shows that in this configuration one half of the supersymmetry generators of the single brane configuration survive. That is, 8 supercharges survive as symmetries, implementing an $N = 2$ supersymmetry in the effective four dimensional theory. The two scalar particle identified geometrically form part of the $N = 2$ vector multiplet. That multiplet consists of spin 1, spin 1/2 and spin 0 particles. Thus the configuration above describes an effective $D = 4$, $N = 2$, $U(1)$ supersymmetric gauge theory. The effective four dimensional gauge coupling constant is related to the effective five dimensional gauge coupling constant in the usual Kaluza-Klein manner:

$$g^2_{\mathrm{YM},4} = \frac{g^2_{\mathrm{YM},5}}{\Delta x_6} \,. \tag{7.12}$$

Changing the value of the separation Δx_6 amounts to rescaling $g^2_{\mathrm{YM},4}$. As in the single brane case, the gauge symmetry can be enhanced to $U(N_\mathrm{C})$ by suspending N_C D4-branes between the NS5-branes[1] (Fig. 27).

Rotating one of the NS5-branes from the x_4, x_5 directions to the x_8, x_9 directions will lead to the desired $D = 4$, $N = 1$, $U(N_\mathrm{C})$ effective gauge theory. Before pursuing this, it will be useful to study the effective field theory on a D4-brane suspended between two D6-branes (Fig. 28).

[1]The actual symmetry turns out to be $SU(N_\mathrm{C})$. This is discussed in [93, 99].

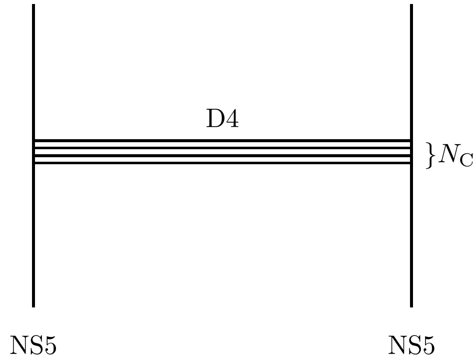


Fig. 27. The effective field theory is a $D = 4$, $N = 2$ $U(N_{\mathrm{C}})$ SUSY gauge theory.

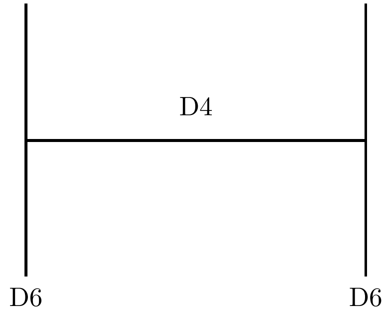


Fig. 28. The effective field theory on the D4 brane is a $D = 4$, $N = 2$ SUSY field theory, containg an $N = 2$ matter hypermultiplet. It contains no massless gauge particles.

The spatial extension of the branes is:

$$\begin{array}{cccccccc} & x_{0123} & x_4 & x_5 & x_6 & x_7 & x_8 & x_9 \\ D4 & + & - & - & = & - & - & - \\ D6 & + & - & - & - & + & + & +\,. \end{array} \tag{7.13}$$

This configuration also has 8 surviving supercharges and thus the effective field theory has $N = 2$ supersymmetry in $D = 4$. Similar considerations to the ones used above show that the effective theory contains at least three massless spin 0 particles. These correspond to the allowed translations in the directions x_7, x_8 and x_9. An $N = 2$ supersymmetric multiplet would require either two or four spin 0 particles. At the case at hand, the multiplet is actually the $N = 2$ hypermultiplet which contains four spin 0 and four spin 1/2 degrees of freedom. It is important to note that the system contains no massless spin 1 degrees of freedom, that is the low energy effective theory is not an unbroken gauge theory. Actually, later we will show that this configuration will be part of the description of the Higgs phase of the supersymmetric gauge theory. The fact that the geometrical considerations were not sufficient to identify all four spin 0 particles shows us one of the limitations of the simple geometrical analysis. The fourth spin 0 particle can be identified in this case with the component of a compactified gauge field, namely A_6. At this stage of understanding of the gauge theory–brane correspondence, one finds parameters in the brane picture with no clearly known field theoretical interpretation and .

Returning to the $N = 2$ $U(N_{\mathrm{C}})$ gauge configuration, one notes that a separation of the D4-branes along the directions x_4, x_5 leads to all rank preserving possible breakings of the gauge symmetry. This is similar to what was described before in the case of the separations of N_{C} infinitely extended parallel Dp-branes. The mass of the W particles is again proportional to $\Delta\mathrm{D4}(x_4, x_5)$, which denotes the separation of two D4-branes in the x_4, x_5 directions. The complex number of moduli is immediately read out of the geometrical picture to be N_{C}. This coincides with algebraic analysis determining the complex number of massless spin 0 particles surviving the breaking of the gauge group.

In Figure 29 we summarize the complex number of massless spin 0 particles appearing in any of the four configurations discussed until now.

7.5 An effective $D = 4$, $N = 1$, $U(N_{\mathrm{C}})$ gauge theory with matter

We construct now the configuration leading to $N = 1$ supersymmetry. The rotation of one of the NS5 from x_4, x_5 to the x_8, x_9 directions corresponds to adding an infinite mass term to the scalar fields in the adjoint representation, a rotation by different angles would have given rise in field theoretical language to a finite mass term. The rotation leads to a configuration with 4 surviving supercharges (Fig. 30). The effective four dimensional theory is a $U(1)$ gauge symmetry and has no moduli, as can be seen from the by now familiar geometrical considerations. This agrees with a description by an effective $D = 4$, $N = 1$ supersymmetric $U(1)$ gauge theory.

The gauge symmetry can be enhanced to $U(N_{\mathrm{C}})$ by suspending N_{C} D4-branes between the two different NS5-branes (Fig. 31).

The final ingredient needed is to add some flavor to the effective field theory. This is done by distributing N_{F} D6-branes along the x_6 direction (Fig. 32).

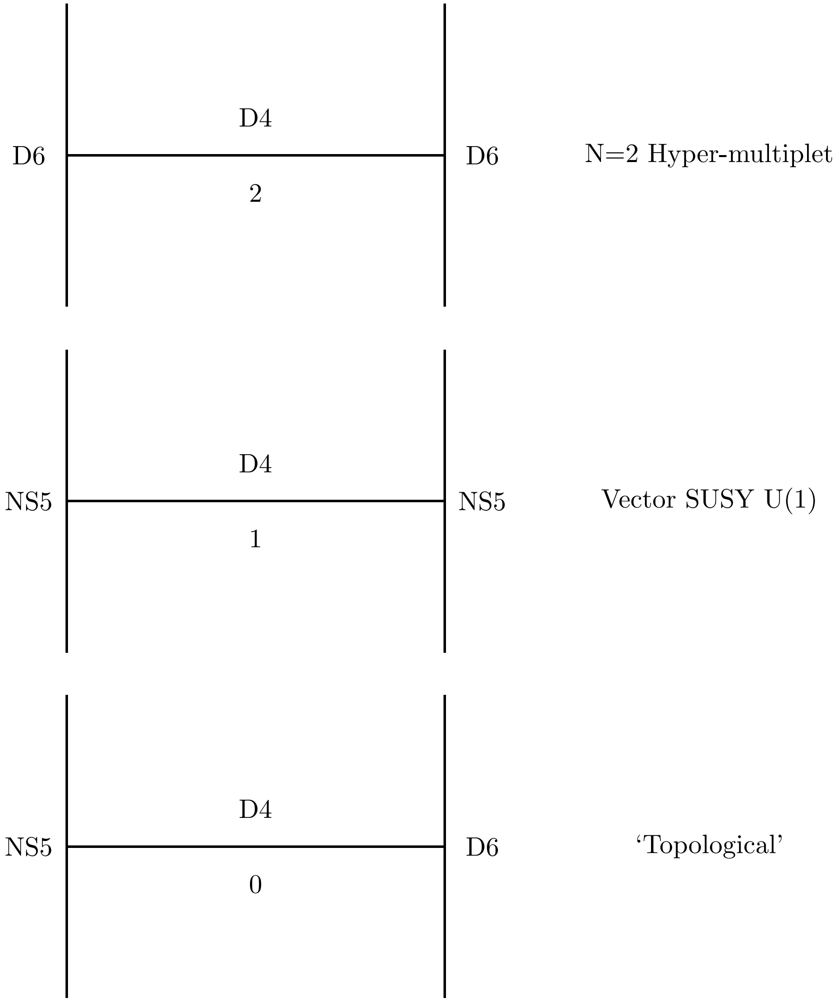


Fig. 29. A summary of the particle content of the low-energy effective theories described in Section 7.4.

The spatial extension of the various branes is:

$$\begin{array}{lccccccc} & x_{0123} & x_4 & x_5 & x_6 & x_7 & x_8 & x_9 \\ D4 & + & - & - & = & - & - & - \\ D6 & + & - & - & - & + & + & + \\ NS5 & + & + & + & - & - & - & - \\ NS5' & + & - & - & - & - & + & +\,. \end{array} \tag{7.14}$$

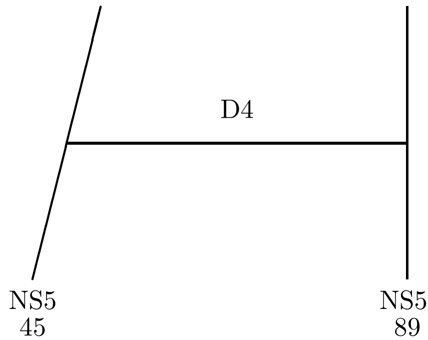


Fig. 30. The configuration leading to a $D = 4$, $N = 1$ $U(1)$ SUSY gauge theory. The NS5 extending also in the directions x_8, x_9 is labeled NS5'.

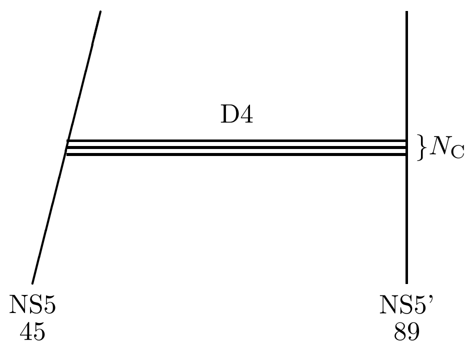


Fig. 31. The configuration leading to a $D = 4$, $N = 1$ $U(N_C)$ SUSY gauge theory.

The matter fields appear as representations of diagonal vectorial subgroup of the flavor group $SU(N_F) \times SU(N_F) \times U(1)$. Geometrically they are associated with strings connecting the D6-branes with the D4-branes. Touching the D6-branes endows the open strings with flavor, and touching the D4-branes endows them with color. Their directionality is responsible for the appearance of both the N_F and $\bar{N}_F$. In [93] the reader will find references for attempts to construct chiral configurations and to identify the full flavor group. The masses of the squarks have a geometrical interpretation: they are proportional to the distance between the D4 and D6-branes along the x_4, x_5 directions. The effective field theory is thus a four dimensional $N = 1$ $U(N_C)$ supersymmetric gauge theory which has matter in the fundamental representation of the vector subgroup of the flavor symmetry. This is as close to our goal as we will reach in this lecture.

The way to obtain the Seiberg dual configuration is essentially to move the position of the NS5-brane residing on the left hand part of the

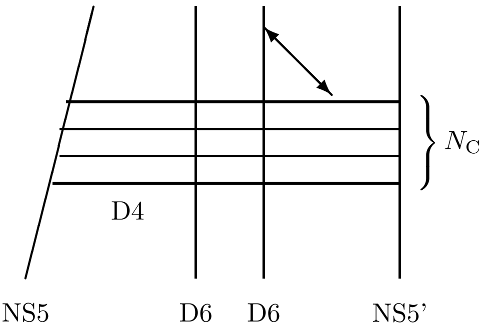


Fig. 32. Adding D6 branes allows the existence of matter in the effective field theory. Open strings ending on D4 and D6 branes carry both color and flavor.

configuration all the way to the right hand part. The resulting new configuration will have $SU(N_\mathrm{F} - N_\mathrm{C})$ gauge symmetry and N_F and $\bar{N}_\mathrm{F}$ colored matter as well as N_F^2, flavor adjoint color singlets, which is exactly the result of Seiberg (Fig. 33). The N_F^2 fields are essentially the color-singlet particles appearing in the effective field theory corresponding to the brane configuration (Fig. 28).

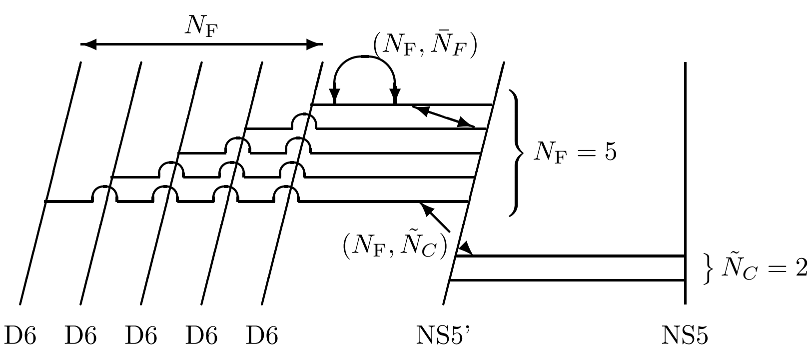


Fig. 33. This is the dual configuration to that of a $N = 1$ SUSY gauge theory with $N_\mathrm{C} = 3$ and $N_\mathrm{F} = 5$ (Fig. 39). The configuration shown has $N_\mathrm{C} = 5 - 3 = 2$, $N_\mathrm{F} = 5$ and, in addition, N_F^2 color singlet massless particles.

7.6 More pieces of information

In order to be able to perform this displacement of the NS5-brane, one needs three more pieces of "information" as to the behavior of branes.

1)

Changes of the positions of branes correspond in some cases to smooth changes in field theoretical parameters. For example, a change in the relative displacement in the x_4, x_5 directions of the D4-branes in the $N = 4$ and $N = 2$ supersymmetric configurations led to a smooth change of the mass of the W's (or equivalently to smooth changes in the Higgs expectation values). A change in the relative position in the D4 and D6-branes in the x_4, x_5 directions led to a change in the masses of the squarks (it is actually equivalent to a smooth change in the expectation value of the scalar coupled to fermions by a Yukawa coupling) (Fig. 34).

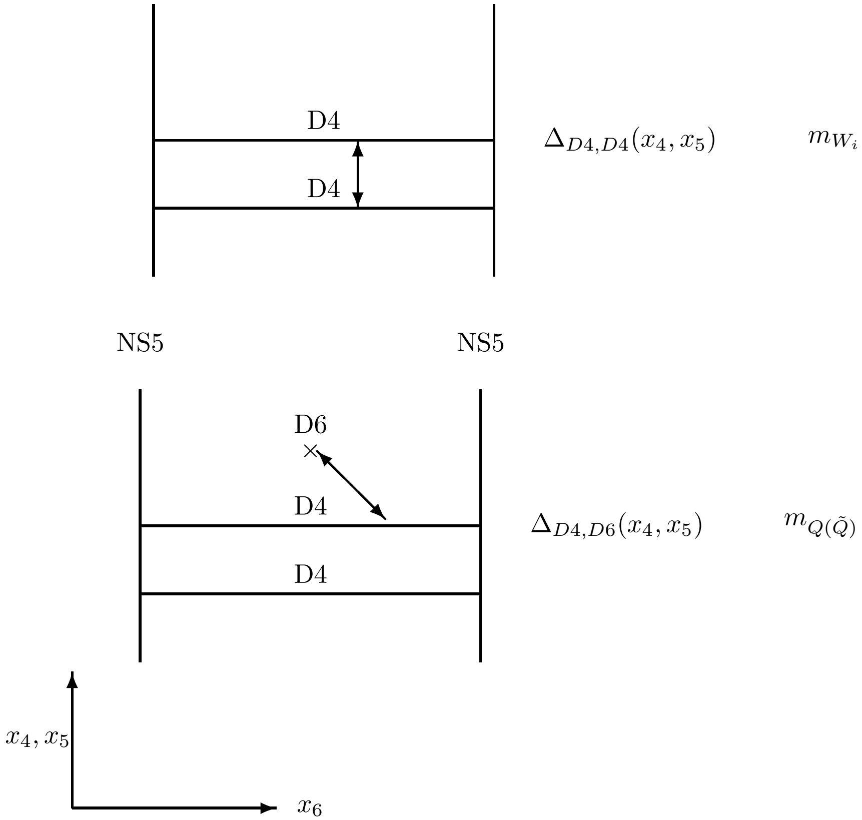


Fig. 34. The values of the masses of the W and squark particles is encoded in this figure. They are proportional to the distances $\Delta_{\mathrm{D4,D4}}(x_4, x_5)$ and $\Delta_{\mathrm{D4,D6}}(x_4, x_5)$, respectively. Both distances are indicated in the figure. Varying these distances smoothly changes the parameters of the field theory.

A change in the relative distance between the NS5-branes corresponded to a change in the coupling of the gauge theory on the D4-branes suspended between them. A change in the relative position along the x_7, x_8 and x_9 directions of the positions of two NS5-branes corresponds to adding a Fayet-Iliopoulos term in an $N = 2$ field theoretical interpretation (Fig. 35). A similar interpretation can be given for the separation along the x_7 direction between the NS5 and NS5′ in the $N = 1$ case supersymmetric case.

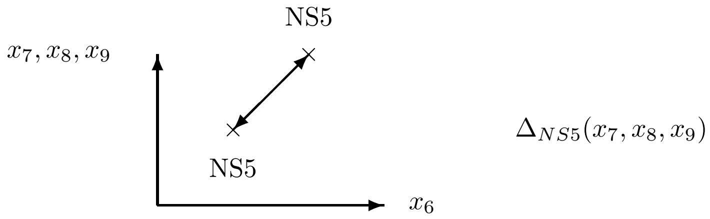


Fig. 35. The NS5 branes can be separated in the x_7, x_8, x_9 directions. Each such separation corresponds to appropriate components of the Fayet-Iliopoulos term in $N = 2$ SUSY $U(N_{\mathrm{C}})$ gauge theory.

In other cases changes in the position of branes could have more abrupt consequences. Consider a configuration consisting only of a NS5-brane and a D6-brane separated in the x_6 direction (Fig. 36). Displacing the NS5-brane for example in the x_6 direction is a smooth motion as long as the two branes do not intersect (this is an example where a brane parameter, the relative separation in the x_6 direction, has no clear impact on the field theory description). Eventually, the two branes have to intersect. However once they do, it is clear that something more singular may result. What actually can be shown to occur [93, 95] is that as the NS5-brane "crosses" the D6-brane a D4-brane is created and suspended between the NS5 and D6-brane. Note that there is no particle content to the effective theory on that D4-brane and thus no new degrees of freedom are created as the D4-brane is formed.

2)

Supersymmetry allows that only a single D4-brane can be suspended between a NS5-brane and a D6-brane [93, 95] (Fig. 37).

3)

We have seen that the Coulomb phase of $N = 4$ gauge theories is spanned by the separation among the D4-brane. The same is true the $N = 2$ case. The $N = 1$ case has no Coulomb phase. Also the Higgs phase has a simple geometrical realization in terms of branes. Consider an $N = 2$

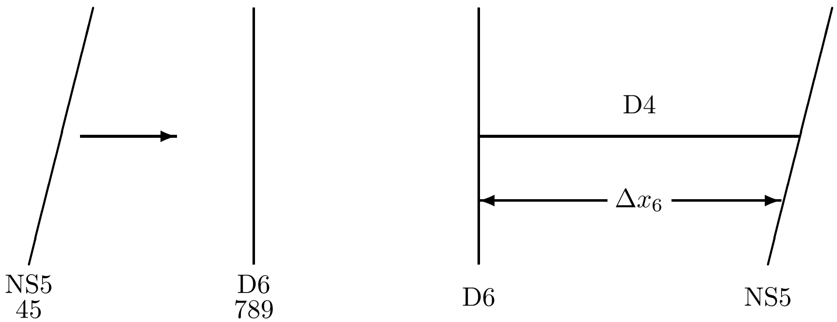


Fig. 36. A D4 brane is created when the NS5 brane crosses the D6 brane, from left to right in this figure. The D4 brane carries no massless degrees of freedom.

configuration in which in addition of having a D4-brane suspended between two NS5-branes a D6-brane is inserted between the NS5-branes (Fig. 38). The separation of the D4-branes on the D6-brane along the x_4, x_5 directions maintains the supersymmetry. However the effective field theory on each of the D4-branes contains no spin 1 massless particles. One can show that this breaking indeed represents the Higgs phenomenon, the separation among the two D4-branes along the x_4, x_5 directions being proportional to the expectation value of the scalar field responsible for the Higgsing. This result is true also for the $N = 1$ configurations. For the $N = 1$ case one needs also to realize that the effective theory on a D4-brane suspended between a D6-brane and a NS5$'$-brane contains one massless complex spin 0 field, one massless spin 1/2 field and no massless spin 1 fields.

The brane geometrical picture has a classical weakly coupled taste. In order to be able to trust the pictures as one moves the branes around it is advisable to be in a weak coupling situation as long as possible. To enforce that, one starts the journey by setting the $D = 4$ $N = 1$ $U(N_\mathrm{C})$ supersymmetric gauge theory with N_F flavors in the Higgs phase. Retaining the Higgs phase, one is able not only to reproduce the particle content required by Seiberg's duality, but also his analysis of comparing the dimensions of the moduli space at both ends of the duality transformation, as well as comparing the numbers of the relevant operators. One is also able to show in the brane picture as Seiberg has done in field theory that masses as one end of the duality pair correspond to expectation values on the other side.

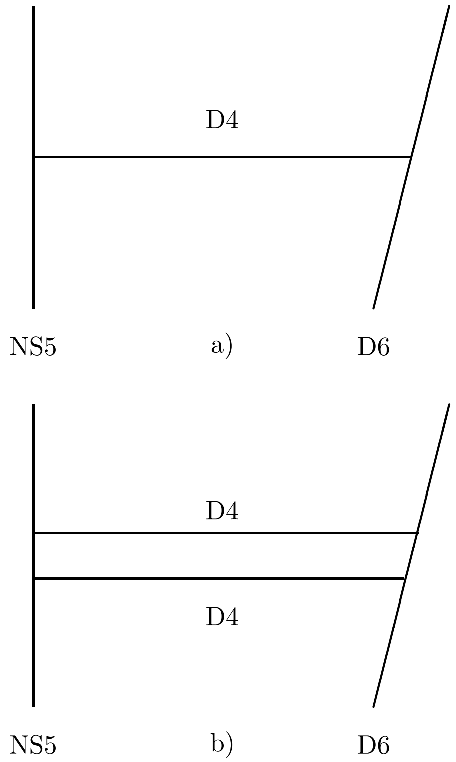


Fig. 37. SUSY allows only the existence of the configuration **a)**. Configuration **b)** is disallowed. However, if the NS5 is exchanged with an NS5', configuration **b)** is allowed.

7.7 Obtaining the dual field theory

At this stage in the lecture a movie[2] composed out of the comic strips was shown. This enabled an easy visualization of the continuous features of the duality transformation. The reader is actually equipped by now to embark on this journey on her/his own (or by reading [68, 69, 93]). As a navigational aid, I will briefly mention the different stages of the journey. The starting point is a configuration depicted in Figure 39, which has $N_{\mathrm{F}} = 5$ and $N_{\mathrm{C}} = 3$.

the NS5-brane is displaced across the D6-branes respecting the rule mentioned in 1) above concerning D4-brane creation.

[2]An attempt is made to digitalize the movie.

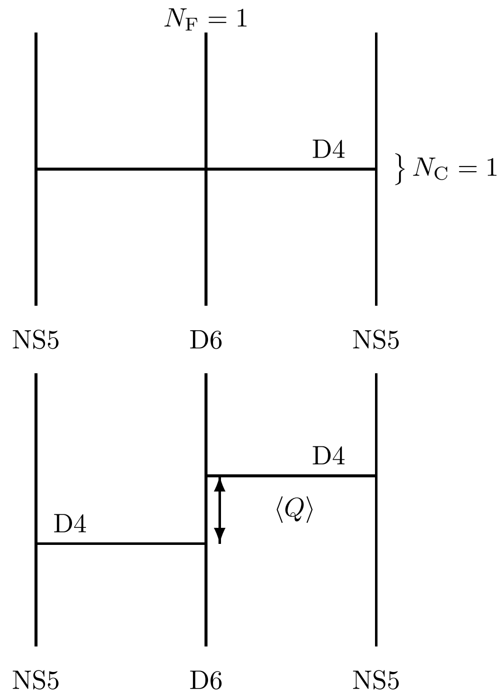


Fig. 38. The Higgs phase of field theory is realized by the D4 branes connecting the two NS5 branes, breaking and separating on the D6 brane. This indicated distance of separation corresponds to the expectation value of the Higgs matter field. The effective field theory in this example carries no massless degrees of freedom at all. In the presence of additional D6 branes, the effective theory would contain massless spin-0 and spin-$\frac{1}{2}$ degrees of freedom, but no massless spin-1 degrees of freedom, as appropriate for a Higgs phase (Fig. 28).

the system is driven into the Higgs phase according to rule 3) above.

it is realized that if the NS5-brane will directly collide with the NS5′ brane the gauge coupling will diverge, leaving the protected weak coupling regime. In order to avoid that, the branes are separated along the x_7 direction before they are made to coincide in the x_6 direction. This separation, as stated in 1) above, corresponds in some sense [68, 69, 93] to the turning on of a Fayet-Iliopoulos coupling in field theory. At this stage, the appropriate reconnections of the branes shows that the configuration represents the Higgs phase of a $U(N_{\mathrm{F}} - N_{\mathrm{C}})$ gauge theory.

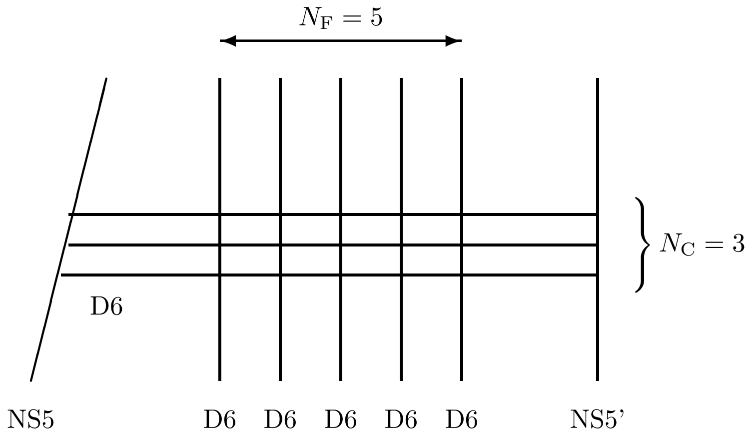


Fig. 39. A $D = 4$, $N = 1$ SUSY, $N_C = 3$, $N_F = 5$ configuration.

rejoining the branes in the x_7 direction requires some quantum adjustments which we will indicate below. The final outcome is shown in Figure 33, the gauge theory has $N_F = 5$ flavors and $N_C = 5 - 3 = 2$ colors. In addition, the theory has 5^2 singlet complex spin-0 and spin-$\frac{1}{2}$ massless particles. This is the realization of the dual theory obtained by Seiberg. The result here is simply the outcome of the NS5 moving according to the brane rules from the left of the NS5' to its right. The simple procedure allows one to obtain the infra-red dual of many other brane configurations.

7.8 Concluding remarks

What has actually been shown is that in a continuous motion starting from one $N = 1$ supersymmetric gauge theory one reaches another. That does not really demonstrate that the two are identical in the infra-red or at all. Taking a bus ride from Chamonix to Les Houches does not constitute a proof that the two cities are the same. What has emerged naturally is one gauge group out of another as well as the presence of N_F^2 color singlets. A more detailed analysis of the resulting massless particle spectrum (chiral ring) and its properties is needed to conclude infra-red duality [68, 69, 93].

It can be shown that $Sp(N)$, $O(N)$ and product gauge groups can be constructed as well [68, 69, 93].

By allowing configurations with k NS5-branes and k' NS5$'$-branes one can test new and old generalizations of Seiberg's duality in the presence of a richer matter content and various Landau-Ginzburg-like interactions [68, 69, 93].

The brane–field theory correspondence obtains even more geometrical features once embedded in M-theory [93, 99, 100].

The classical brane picture needs to be amended by quantum considerations [95, 99]. For example, recall that two D4-branes were not allowed to be suspended between an NS5 and a D6-brane. One could ascribe this to a quantum repulsive force between two D4-branes on the same side of a NS5-brane in a non-BPS configuration. The agreement between the dimensions of the moduli spaces of the two dual models was obtained in the presence of a quantum attractive interaction between two D4-branes on the opposite sides of a NS5-brane. At this stage, these are additional postulates they have however immediate consequences in allowing a unified description of $D = 4$ and $D = 3$ gauge theories with 4 supercharges. For example, consider the $D = 4$ $N = 1$ $U(N_{\mathrm{C}})$ gauge theory with no flavors. Such a system has a ground state [15, 52]. However the same system in $D = 3$ has no ground state [101]. This is indicated in the brane picture in the following manner: compactify one dimension, say x_3, of the world-volume of the D4-brane, $0 \leq x_3 \leq 2\pi R$. It is more convenient to perform a longitudinal T-duality along the x_3 direction. The resulting system is a type IIB string theory whose third direction has radius $1/R$. Thus a four dimensional IIA system corresponds to a type IIB system with vanishing radius. For any finite radius, the type IIB effective theory corresponds to a field theory with more than three dimensions in type IIA. Only an infinite type IIB radius will correspond to an effective three dimensional gauge theory. Due to the repulsive force between the D4-branes, and due to the compact nature of the x_3 coordinate, a stable configuration will form for any finite value of x_3, that is, such a vacuum state exists for any effective supersymmetric gauge theory with no flavors in more than three dimensions. Once the IIB radius is infinite, the equilibrium state exists no more, the three dimensional gauge theory has no vacuum.

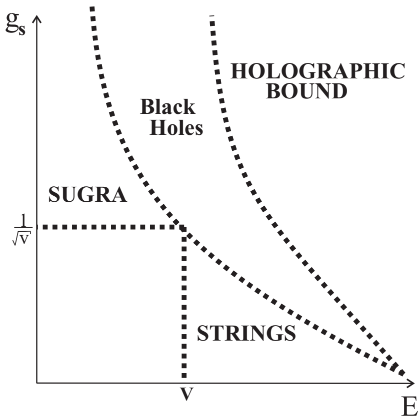


Fig. 40. A prototype of a phase diagram for a gravitational theory.

8 Final remarks

In this set of lectures we have had a panoramic vista of the rich structure of supersymmetric gauge theories. The classification of possible phases of gauge theories were understood well before concepts like confinement could have been analysed analytically in four dimensional continuum theories. Perhaps a similar course can be followed for theories whose symmetry includes general coordinate invariance in analogy to Figures 4, 5 and 8. The qualitative understanding of the phase structure of such systems will proceed their complete quantitative analysis. As a prototype for a phase diagram of gravity we propose the following Figure 40, [102].

E.R. acknowledges S. Elitzur, A. Forge, A. Giveon, D. Kutasov and A. Schwimmer for collobartions and many very valuable discussions. D.B. acknowledges the inspirational atmosphere created by the groups of the Hebrew University and the Weizmann institute.

The references given in this article though detailed are not complete.

References

[1] Y.A. Golfand and E.P. Likhtman, *JETP Lett.* **13** (1971) 323 [Pisma Zh. *Eksp. Teor. Fiz.* **13** (1971) 452].

[2] J. Wess and B. Zumino, *Phys. Lett. B* **49** (1974) 52.

[3] J. Bagger and J. Wess, *Supersymmetry and Supergravity*, 2nd edition (Princeton University Press, 1992).

[4] P.C. West, *Introduction To Supersymmetry And Supergravity*, Singapore (Singapore: World Scientific, 1990), 425.

[5] S. Weinberg, *Phys. Rev. D* **13** (1976) 974.
[6] S. Weinberg, *Phys. Rev. D* **19** (1979) 1277.
[7] L. Susskind, *Phys. Rev. D* **20** (1979) 2619.
[8] J.R. Ellis, S. Kelley and D.V. Nanopoulos, *Phys. Lett. B* **260** (1991) 131.
[9] U. Amaldi, W. de Boer and H. Furstenau, *Phys. Lett. B* **260** (1991) 447.
[10] P. Langacker and M.x. Luo, *Phys. Rev. D* **44** (1991) 817.
[11] D. Kutasov and N. Seiberg, *Nucl. Phys. B* **358** (1991) 600.
[12] Introduction to Supersymmetry `http://www.lns.cornell.edu/ argyres/phys661/`
[13] P. Fayet and S. Ferrara, *Phys. Rept.* **32** (1977) 249.
[14] H.P. Nilles, *Phys. Rept.* **110** (1984) 1.
[15] K. Intriligator and N. Seiberg, *Nucl. Phys. Proc. Suppl. BC* **45** (1996) 1 [`hep-th/9509066`].
[16] S.J. Gates, M.T. Grisaru, M. Rocek and W. Siegel, *Front. Phys.* **58** (1983) 1 [`arXiv:hep-th/0108200`].
[17] S. Weinberg, *The Quantum Theory Of Fields*, Vol. 3: Supersymmetry (Cambridge, UK: Univ. Pr., 2000), 419.
[18] J.M. Figueroa-O'Farrill, *BUSSTEPP lectures on supersymmetry* [`arXiv:hep-th/0109172`].
[19] For a review see, O. Aharony, S.S. Gubser, J.M. Maldacena, H. Ooguri and Y. Oz, *Phys. Rept.* **323** (2000) 183 [`arXiv:hep-th/9905111`].
[20] See for example, O. Aharony, J. Sonnenschein, M.E. Peskin and S. Yankielowicz, *Phys. Rev. D* **52** (1995) 6157 [`arXiv:hep-th/9507013`].
[21] For a review see, T. Banks, *Nucl. Phys. Proc. Suppl.* **62** (1998) 341; *Nucl. Phys. Proc. Suppl.* **68** (1998) 261 [`arXiv:hep-th/9706168`].
[22] S. Elitzur, E. Rabinovici and A. Schwimmer, *Phys. Lett. B* **119** (1982) 165.
[23] N. Parga, E. Rabinovici and A. Schwimmer, *Nucl. Phys. B* **255** (1985) 383.
[24] E. Witten, *Nucl. Phys. B* **188** (1981) 513.
[25] E. Witten, *Nucl. Phys. B* **202** (1982) 253.
[26] P.C. West, *Nucl. Phys. B* **106** (1976) 219.
[27] A. Forge and E. Rabinovici, *Phys. Rev. D* **32** (1985) 927.
[28] V. de Alfaro, S. Fubini and G. Furlan, *Nuovo Cim. A* **34** (1976) 569.
[29] V.P. Akulov and A.I. Pashnev, *Theor. Math. Phys.* **56** (1983) 862; *Teor. Mat. Fiz.* **56** (1983) 344.
[30] S. Fubini and E. Rabinovici, *Nucl. Phys. B* **245** (1984) 17.
[31] P. Claus, M. Derix, R. Kallosh, J. Kumar, P.K. Townsend and A. Van Proeyen, *Phys. Rev. Lett.* **81** (1998) 4553.
[32] R. Kallosh, *Black holes and quantum mechanics* [`arXiv:hep-th/9902007`].
[33] N. Seiberg, *Phys. Lett. B* **318** (1993) 469.
[34] Private communication, P. Fayet and J. Illiopoulos.
[35] P. Fayet, *Phys. Lett. B* **58** (1975) 67.
[36] L. O'Raifeartaigh, *Nucl. Phys. B* **96** (1975) 331.
[37] P. Fayet and J. Iliopoulos, *Phys. Lett. B* **51** (1974) 461.
[38] T. Banks, R. Myerson and J.B. Kogut, *Nucl. Phys. B* **129** (1977) 493.
[39] D. Horn, M. Weinstein and S. Yankielowicz, *Phys. Rev. D* **19** (1979) 3715.
[40] S. Elitzur, R.B. Pearson and J. Shigemitsu, *Phys. Rev. D* **19** (1979) 3698.
[41] A. Ukawa, P. Windey and A.H. Guth, *Phys. Rev. D* **21** (1980) 1013.

[42] For similar structures in a string theory context see, S. Elitzur, B. Pioline and E. Rabinovici, *JHEP* **0010** (2000) 011 [`arXiv:hep-th/0009009`].

[43] J.L. Cardy and E. Rabinovici, *Nucl. Phys. B* **205** (1982) 1.

[44] E. Witten, *Phys. Lett. B* **86** (1979) 283.

[45] H.B. Nielsen and P. Olesen, *Nucl. Phys. B* **61** (1973) 45.

[46] S. Mandelstam, *Phys. Rept.* **23** (1976) 245.

[47] G. 't Hooft, *Nucl. Phys. B* **138** (1978) 1.

[48] N. Seiberg and E. Witten, *Nucl. Phys. B* **431** (1994) 484 [`arXiv:hep-th/9408099`].

[49] N. Seiberg and E. Witten, *Nucl. Phys. B* **426** (1994) 19; *Erratum-ibid.* **430** (1994) 485 [`arXiv:hep-th/9407087`].

[50] A. Bilal, *Duality in $N = 2$ SUSY $SU(2)$ Yang-Mills Theory: A pedagogical introduction to the work of Seiberg and Witten* [`arXiv:hep-th/9601007`].

[51] T.R. Taylor, G. Veneziano and S. Yankielowicz, *Nucl. Phys. B* **218** (1983) 493.

[52] D. Amati, K. Konishi, Y. Meurice, G.C. Rossi and G. Veneziano, *Phys. Rept.* **162** (1988) 169.

[53] I. Affleck, M. Dine and N. Seiberg, *Nucl. Phys. B* **241** (1984) 493; *Nucl. Phys. B* **256** (1985) 557.

[54] K. Intriligator, *Phys. Lett. B* **336** (1994) 409 [`hep-th/9407106`].

[55] S. Elitzur, A. Forge, A. Giveon and E. Rabinovici, *Phys. Lett. B* **353** (1995) 79 [`hep-th/9504080`].

[56] S. Elitzur, A. Forge, A. Giveon and E. Rabinovici, *Nucl. Phys. B* **459** (1996) 160 [`hep-th/9509130`].

[57] S. Elitzur, A. Forge, A. Giveon and E. Rabinovici, *Ahrenshoop Symp.* (1995) 174 [`hep-th/9512140`].

[58] K. Intriligator, R.G. Leigh and N. Seiberg, *Phys. Rev. D* **50** (1994) 1092 [`hep-th/9403198`].

[59] K. Intriligator and N. Seiberg, *Nucl. Phys. B* **431** (1994) 551 [`hep-th/9408155`].

[60] K. Intriligator and N. Seiberg, *Nucl. Phys. B* **444** (1995) 125 [`hep-th/9503179`].

[61] K. Intriligator, N. Seiberg and S.H. Shenker, *Phys. Lett. B* **342** (1995) 152 [`hep-ph/9410203`].

[62] S. Elitzur, A. Forge, A. Giveon, K. Intriligator and E. Rabinovici, *Phys. Lett. B* **379** (1996) 121 [`hep-th/9603051`].

[63] T. Banks and E. Rabinovici, *Nucl. Phys. B* **160** (1979) 349; E. Fradkin and S. Shenker, *Phys. Rev. D* **19** (1979) 3682.

[64] L.F. Abbott and E. Farhi, *Nucl. Phys. B* **189** (1981) 547.

[65] A. Casher, *Phys. Lett. B* **83** (1979) 395.

[66] G. 't Hooft, in *C79-08-26.4* PRINT-80-0083 (UTRECHT) *Lecture given at Cargese Summer Inst.* (Cargese, France, Aug. 26 – Sep. 8, 1979).

[67] N. Seiberg, *Nucl. Phys. B* **435** (1995) 129 [`hep-th/9411149`].

[68] S. Elitzur, A. Giveon and D. Kutasov, *Phys. Lett. B* **400** (1997) 269 [`hep-th/9702014`].

[69] S. Elitzur, A. Giveon, D. Kutasov, E. Rabinovici and A. Schwimmer, *Nucl. Phys. B* **505** (1997) 202 [`hep-th/9704104`].

[70] N. Dorey, T.J. Hollowood, V.V. Khoze and M.P. Mattis, *The calculus of many instantons* [`arXiv:hep-th/0206063`].

[71] P. Argyres and M. Douglas, *Nucl. Phys. B* **448** (1995) 93 [`hep-th/9505062`].

[72] A.C. Davis, M. Dine and N. Seiberg, *Phys. Lett. B* **125** (1983) 487.

[73] T. Banks and A. Zaks, *Nucl. Phys. B* **196** (1982) 189.

[74] G. Mack, *Commun. Math. Phys.* **53** (1977) 155.

[75] V.K. Dobrev, G. Mack, V.B. Petkova, S.G. Petrova and I.T. Todorov, *Harmonic Analysis On The N-Dimensional Lorentz Group And Its Application To Conformal Quantum Field Theory* (Berlin, 1977), 280.

[76] P.C. Argyres, M. Ronen Plesser, N. Seiberg and E. Witten, *Nucl. Phys. B* **461** (1996) 71.

[77] R.G. Leigh and M.J. Strassler, *Nucl. Phys. B* **447** (1995) 95 [arXiv:hep-th/9503121].

[78] M.B. Einhorn, G. Goldberg and E. Rabinovici, *Nucl. Phys. B* **256** (1985) 499.

[79] For some recent string theory results see, O. Aharony, B. Kol and S. Yankielowicz, *JHEP* **0206** (2002) 039 [arXiv:hep-th/0205090].

[80] For a review, see: J. Harvey [hep-th/9603082], and references therein.

[81] Ya.B. Zeldovich, *Sov. Phys. Uspekhi* **11** (1968) 381.

[82] S. Weinberg, *Rev. Mod. Phys.* **61** (1989).

[83] D.J. Amit and E. Rabinovici, *Nucl. Phys. B* **257** (1985) 371.

[84] E. Rabinovici, B. Saering and W.A. Bardeen, *Phys. Rev. D* **36** (1987) 562.

[85] W.M. Alberico and S. Sciuto, *Symmetry And Simplicity in Physics*, in Proceedings, Symposium on The Occasion of Sergio Fubini's 65th Birthday, Turin, Italy, February 24-26, 1994 (Singapore, Singapore: World Scientific, 1994), 220.

[86] W.A. Bardeen, M. Moshe and M. Bander, *Phys. Rev. Lett.* **52** (1984) 1188.

[87] E. Rabinovici, unpublished.

[88] N. Seiberg, *Nucl. Phys. B* **435** (1995) 129 [hep-th/9411149].

[89] P. Argyres, M.R. Plesser, N. Seiberg and E. Witten, *Nucl. Phys. B* **461** (1996) 71 [hep-th/9511154].

[90] N. Seiberg, *Phys. Lett. B* **318** (1993) 469 [hep-th/9309335].

[91] For example, see: A. Giveon and M. Roček, *Phys. Lett. B* **363** (1995) 173 [hep-th/9508043], and references therein.

[92] J. Polchinski, *TASI Lectures on D-Branes* [hep-th/9611050].

[93] A. Giveon and D. Kutasov, *Brane Dynamics and Gauge Theory* [hep-th/9802067].

[94] E. Rabinovici, *Non-perturbative gauge dynamics and strings*, Published in "Erice 1999, Basics and highlights in fundamental physics", 284.

[95] A. Hanany and E. Witten, *Nucl. Phys. B* **492** (1997) 152 [hep-th/9611230].

[96] M. Duff, R.R. Khuri and J.X. Lu, *Phys. Rept.* **259** (1995) 213 [hep-th/9412184].

[97] M.B. Green, J.H. Schwarz and E. Witten, *Superstring Theory* (Cambridge University Press, 1987).

[98] A. Giveon, M. Porrati and E. Rabinovici, *Phys. Rept.* **244** (1994) 77 [hep-th/9401139].

[99] E. Witten, *Nucl. Phys. B* **500** (1997) 3 [hep-th/9703166].

[100] P.S. Howe, N.D. Lambert and P.C. West, *Phys. Lett. B* **418** (1998) 85 [arXiv:hep-th/9710034]; N.D. Lambert and P.C. West, *Nucl. Phys. B* **524** (1998) 141 [arXiv:hep-th/9712040].

[101] O. Aharony, A. Hanany, K. Intriligator, N. Seiberg and M.J. Strassler, *Nucl. Phys. B* **499** (1997) 67 [hep-th/9703110].

[102] J.L. Barbon, I.I. Kogan and E. Rabinovici, *Nucl. Phys. B* **544** (1999) 104 [arXiv:hep-th/9809033]; S.A. Abel, J.L.F. Barbon, I.I. Kogan and E. Rabinovici, *JHEP* **9904** (1999) 015 [arXiv:hep-th/9902058].

LECTURE 3

AN INTRODUCTION TO DUALITY SYMMETRIES IN STRING THEORY

A. SEN

Harish-Chandra Research Institute,
Chhatnag Road, Jhusi, Allahabad
211019, India

Contents

AN INTRODUCTION TO DUALITY SYMMETRIES IN STRING THEORY

A. Sen

Abstract

In this review I discuss some basic aspects of non-perturbative string theory. The topics include test of duality symmetries based on the analysis of the low energy effective action and the spectrum of BPS states, relationship between different duality symmetries, and an introduction to M-theory.

1 Introduction

During the last few years, our understanding of string theory has undergone a dramatic change. The key to this development is the discovery of duality symmetries, which relate the strong and weak coupling limits of apparently different string theories. These symmetries not only relate apparently different string theories, but give us a way to compute certain strong coupling results in one string theory by mapping it to a weak coupling result in a dual string theory. In this review I shall try to give an introduction to this exciting subject. However, instead of surveying all the important developments in this subject I shall try to explain the basic ideas with the help of a few simple examples. I apologise for the inherent bias in the choice of examples and the topics, this is solely due to the varied degree of familiarity that I have with this vast subject. I have also not attempted to give a complete list of references. Instead I have only included those references whose results have been directly used or mentioned in this article. A complete list of references may be obtained by looking at the citations to some of the original papers in spires. There are also many other reviews in this subject where more references can be found [1–17]. I hope that this review will serve the limited purpose of initiating a person with a knowledge of perturbative string theory into this area. (For an introduction to perturbative string theory, see [18].)

The review will be divided into seven main sections as described below.

1. A brief review of perturbative string theory: in this section I shall very briefly recollect some of the results of perturbative string theory which will be useful to us in the rest of this article. This will in no way constitute an introduction to this subject; at best it will serve as a reminder to a reader who is already familiar with this subject;

2. Notion of duality symmetry: in this section I shall describe the notion of duality symmetry in string theory, a few examples of duality conjectures in string theory, and the general procedure for testing these duality conjectures;

3. Analysis of the low energy effective action: in this section I shall describe how one arrives at various duality conjectures by analyzing the low energy effective action of string theory;

4. Precision test of duality based on the spectrum of BPS states: in this section I shall discuss how one can device precision tests of various duality conjectures based on the analysis of the spectrum of a certain class of supersymmetric states in string theory;

5. Interrelation between various dualities: in this section I shall try to relate the various duality conjectures introduced in the Sections 3–5 by "deriving" them from a basic set of duality conjectures. I shall also discuss what we mean by relating different dualities and try to formulate the rules that must be followed during such a derivation.

6. Duality in theories with <16 supersymmetries: the discussion in Sections 4–6 is focussed on string theories with at least 16 supersymmetry generators. In this section I consider theories with less number of supersymmetries. Specifically we shall focus our attention on theories with eight supercharges, which correspond to $N = 2$ supersymmetry in four dimensions.

7. M-theory: in this section I discuss the emergence of a new theory in eleven dimensions – now known as M-theory – from the strong coupling limit of type IIA string theory. I also discuss how compactification of M-theory gives rise to new theories that cannot be regarded as perturbative compactification of a string theory.

Throughout this article I shall work in units where $\hbar = 1$ and $c = 1$.

2 A brief review of perturbative string theory

String theory is based on the simple idea that elementary particles, which appear as point-like objects to the present day experimentalists, are actually different vibrational modes of strings. The energy per unit length of the string, known as string tension, is parametrized as $(2\pi\alpha')^{-1}$, where α' has the dimension of $(\text{length})^2$. As we shall describe later, this theory automatically contains gravitational interaction between elmentary particles, but in order to correctly reproduce the strength of this interaction, we need to choose $\sqrt{\alpha'}$ to be of the order of 10^{-33} cm. Since $\sqrt{\alpha'}$ is the only length parameter in the theory, the typical size of a string is of the order of $\sqrt{\alpha'} \sim 10^{-33}$ cm – a distance that cannot be resolved by present day experiments. Thus there is no direct way of testing string theory, and its appeal lies in its theoretical consistency.

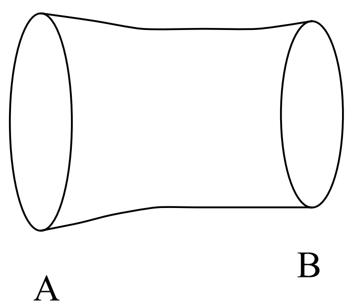


Fig. 1. Propagation of a closed string.

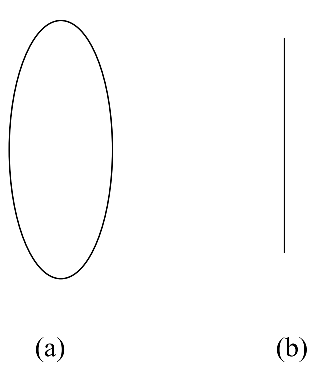


Fig. 2. a) A closed string, and **b)** an open string.

The basic principle behind constructing a quantum theory of relativistic string is quite simple. Consider propagation of a string from a space-time configuration A to a space-time configuration B. During this motion the string sweeps out a two dimensional surface in space-time, known as the string world-sheet (see Fig. 1). The amplitude for the propagation of the string from the space-time position A to space-time position B is given by the weighted sum over all world-sheet bounded by the initial and the final locations of the string. The weight factor is given by e^{-S} where S is the product of the string tension and the area of the world-sheet. It turns out that this procedure by itself does not give rise to a fully consistent string theory. In order to get a fully consistent string theory we need to add some internal fermionic degrees of freedom to the string and generalize the notion of area by adding new terms involving these fermionic degrees of freedom. The leads to five (apparently) different consistent string theories in $(9+1)$ dimensional space-time, as we shall describe.

In the first quantized formalism, the dynamics of a point particle is described by quantum mechanics. Generalizing this we see that the first quantized description of a string will involve a $(1+1)$ dimensional quantum field theory. However unlike a conventional quantum field theory where the spatial directions have infinite extent, here the spatial direction, which labels the coordinate on the string, has finite extent. It represents a compact circle if the string is closed (Fig. 2a) and a finite line interval if the string is open (Fig. 2b). This $(1 + 1)$ dimensional field theory is known as the world-sheet theory. The fields in this $(1 + 1)$ dimensional quantum field theory and the boundary conditions on these fields vary in different string theories. Since the spatial direction of the world-sheet theory has finite extent, each world-sheet field can be regarded as a collection of infinite number of harmonic oscillators labelled by the quantized momentum along this spatial direction. Different states of the string are obtained by acting on the Fock vacuum by these oscillators. This gives an infinite tower of states. Typically each string theory contains a set of massless states and an infinite tower of massive states. The massive string states typically have mass of the order of $(10^{-33}\,\mathrm{cm})^{-1} \sim 10^{19}$ GeV and are far beyond the reach of the present day accelerators. Thus the interesting part of the theory is the one involving the massless states. We shall now briefly describe the spectrum and interaction in various string theories and their compactifications.

2.1 The spectrum

There are five known fully consistent string theories in ten dimensions. They are known as type IIA, type IIB, type I, $E_8 \times E_8$ heterotic and $SO(32)$ heterotic string theories respectively. Here we give a brief description of the degrees of freedom and the spectrum of massless states in each of these theories. We shall give the description in the so called light-cone gauge which has the advantage that all states in the spectrum are physical states.

1. Type II string theories: in this case the world-sheet theory is a free field theory containing eight scalar fields and eight Majorana fermions. These eight scalar fields are in fact common to all five string theories, and represent the eight transverse coordinates of a string moving in a nine dimensional space. It is useful to regard the eight Majorana fermions as sixteen Majorana-Weyl fermions, eight of them having left-handed chirality and the other eight having right-handed chirality. We shall refer to these as left- and right-moving fermions respectively. Both the type II string theories contain only closed strings; hence the spatial component of the world-sheet is a circle. The eight scalar fields satisfy periodic boundary condition as we go around the circle. The fermions have a choice of having periodic or anti-periodic boundary conditions. It is customary to refer to periodic boundary condition as Ramond (R) boundary condition [123] and anti-periodic boundary condition as Neveu-Schwarz (NS) boundary condition [124]. It turns out that in order to get a consistent string theory we need to include in our theory different classes of string states, some of which have periodic and some of which have anti-periodic boundary condition on the fermions. In all there are four classes of states which need to be included in the spectrum:

    - NS-NS where we put anti-periodic boundary conditions on both the left- and the right-moving fermions;
    - NS-R where we put anti-periodic boundary condition on the left-moving fermions and periodic boundary condition on the right-moving fermions;
    - R-NS where we put periodic boundary condition on the left-moving fermions and anti-periodic boundary condition on the right-moving fermions;
    - R-R where we put anti-periodic boundary conditions on both the left- and the right-moving fermions.

    Finally, we keep only about (1/4)th of the states in each sector by keeping only those states in the spectrum which have in them only even number of left-moving fermions and even number of right-moving fermions. This is known as the GSO projection [125]. The procedure has some ambiguity since in each of the four sectors we have the choice of assigning to the ground state either even or odd fermion number. Consistency of string theory rules out most of these possibilities, but at the end two possibilities remain. These differ from each other in the following way. In one possibility, the assignment of the left- and the right-moving fermion number to the left- and the right-moving Ramond ground states are carried out in an identical manner. This gives type IIB string theory. In the second possibility the GSO projections in the left- and the right-moving sector differ from each other. This theory is known as type IIA string theory.

Typically states from the Ramond sector are in the spinor representation of the $SO(9,1)$ Lorentz algebra, whereas those from the NS sector are in the tensor representation. Since the product of two spinor representation gives us back a tensor representation, the states from the NS-NS and the RR sectors are bosonic, and those from the NS-R and R-NS sectors are fermionic. It will be useful to list the massless bosonic states in these two string theories. Since the two theories differ only in their R-sector, the NS sector bosonic states are the same in the two theories. They constitute a symmetric rank two tensor field, an anti-symmetric rank two tensor field, and a scalar field known as the dilaton[1]. The RR sector massless states of type IIA string theory consist of a vector, and a rank three anti-symmetric tensor. On the other hand, the massless states from the RR sector of type IIB string theory consist of a scalar, a rank two anti-symmetric tensor field, and a rank four anti-symmetric tensor gauge field satisfying the constraint that its field strength is self-dual.

The spectrum of both these theories are invariant under space-time supersymmetry transformations which transform fermionic states to bosonic states and [illegible]. The supersymmetry algebra for type IIB theory is known as the chiral $N = 2$ superalgebra and that of type IIA theory is known as the non-chiral $N = 2$ superalgebra. Both superalgebras consist of 32 supersymmetry generators.

Often it is convenient to organise the infinite tower of states in string theory by their oscillator level defined as follows. As has already been pointed out before, the world-sheet degrees of freedom of the string can be regarded as a collection of infinite number of harmonic oscillators. For the creation operator associated with each oscillator we define the level as the absolute value of the number of units of world-sheet momentum that it creates while acting on the vacuum. The total oscillator level of a state is then the sum of the levels of all the oscillators that act on the Fock vacuum to create this state. (The Fock vacuum, in turn, is characterized by several quantum numbers,

[1]Although from string theory we get the spectrum of states, it is useful to organise the spectrum in terms of fields. In other words the spectrum of massless fields in string theory is identical to that of a free field theory with these fields.

which are the momenta conjugate to the zero modes of various fields − modes carrying zero world-sheet momentum.) We can also separately define left- (right-) moving oscillator level as the contribution to the oscillator level from the left- (right-) moving bosonic and fermionic fields. Finally, if E and P denote respectively the world-sheet energy and momentum[2] then we define $L_0 = (E+P)/2$ and $\bar{L}_0 = (E-P)/2$. L_0 and $\bar{L}_0$ include contribution from the oscillators as well as from the Fock vacuum. Thus for example the total contribution to L_0 will be given by the sum of the right-moving oscillator level and the contribution to L_0 from the Fock vacuum.

2. Heterotic string theories: the world-sheet theory of the heterotic string theories consists of eight scalar fields, eight right-moving Majorana-Weyl fermions and thirty two left-moving Majorana-Weyl fermions. We have as before NS and R boundary conditions as well as GSO projection involving the right-moving fermions. Also as in the case of type II string theories, the NS sector states transform in the tensor representation and the R sector states transform in the spinor representation of the $SO(9,1)$ Lorentz algebra. However, unlike in the case of type II string theories, in this case the boundary condition on the left-moving fermions do not affect the Lorentz transformation properties of the state. Thus bosonic states come from states with NS boundary condition on the right-moving fermions and fermionic states come from states with R boundary condition on the right-moving fermions.

   There are two possible boundary conditions on the left-moving fermions which give rise to fully consistent string theories. They are:

   - $SO(32)$ heterotic string theory: in this case we have two possible boundary conditions on the left-moving fermions: either all of them have periodic boundary condition, or all of them have anti-periodic boundary condition. In each sector we also have a GSO projection that keeps only those states in the spectrum which contain even number of left-moving fermions. The massless bosonic states in this theory consist of a symmetric rank two field, an anti-symmetric rank two field, a scalar field known as the dilaton and a set of 496 gauge fields filling up the adjoint representation of the gauge group $SO(32)$;
   - $E_8 \times E_8$ heterotic string theory: in this case we divide the thirty two left-moving fermions into two groups of sixteen each and

[2]We should distinguish between world-sheet momentum, and the momenta of the $(9+1)$ dimensional theory. The latter are the the momenta conjugate to the zero modes of various bosonic fields in the world-sheet theory.

use four possible boundary conditions, 1) all the left-moving fermions have periodic boundary condition 2) all the left-moving fermions have anti-periodic boundary condition, 3) all the left-moving fermions in group 1 have periodic boundary conditions and all the left-moving fermions in group 2 have anti-periodic boundary conditions, 4) all the left-moving fermions in group 1 have anti-periodic boundary conditions and all the left-moving fermions from group 2 have periodic boundary conditions. In each sector we also have a GSO projection that keeps only those states in the spectrum which contain even number of left-moving fermions from the first group, and also even number of left-moving fermions from the second group. The massless bosonic states in this theory consist of a symmetric rank two field, an anti-symmetric rank two field, a scalar field known as the dilaton and a set of 496 gauge fields filling up the adjoint representation of the gauge group $E_8 \times E_8$.

The spectrum of states in both the heterotic string theories are invariant under a set of space-time supersymmetry transformations. The relevant superalgebra is known as the chiral $N = 1$ supersymmetry algebra, and has sixteen real generators.

Using the bose-fermi equivalence in $(1 + 1)$ dimensions, we can reformulate both the heterotic string theories by replacing the thirty two left-moving fermions by sixteen left-moving bosons. In order to get a consistent string theory the momenta conjugate to these bosons must take discrete values. It turns out that there are only two consistent ways of quantizing the momenta, giving us back the two heterotic string theories.

3. Type I string theory: the world-sheet theory of type I theory is identical to that of type IIB string theory, with the following two crucial difference.

   - Type IIB string theory has a symmetry that exchanges the left- and the right-moving sectors in the world-sheet theory. This transformation is known as the world-sheet parity transformation. (This symmetry is not present in type IIA theory since the GSO projection in the two sectors are different). In constructing type I string theory we keep only those states in the spectrum which are invariant under this world-sheet parity transformation;
   - In type I string theory we also include open string states in the spectrum. The world-sheet degrees of freedom are identical to those in the closed string sector. Specifying the theory requires

us to specify the boundary conditions on the various fields. We put Neumann boundary condition on the eight scalars, and appropriate boundary conditions on the fermions.

The spectrum of massless bosonic states in this theory consists of a symmetric rank two tensor and a scalar dilaton from the closed string NS sector, an anti-symmetric rank two tensor from the closed string RR sector, and 496 gauge fields in the adjoint representation of $SO(32)$ from the open string sector. This spectrum is also invariant under the chiral $N=1$ supersymmetry algebra with sixteen real supersymmetry generators.

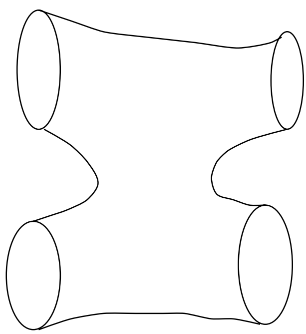


Fig. 3. A string world-sheet bounded by four external strings.

2.2 *Interactions*

So far we have discussed the spectrum of string theory, but in order to fully describe the theory we must also describe the interaction between various particles in the spectrum. In particular, we would like to know how to compute a scattering amplitude involving various string states. It turns out that there is a unique way of introducing interaction in string theory. Consider for example a scattering involving four external strings, situated along some specific curves in space-time. The prescription for computing the scattering amplitude is to compute the weighted sum over all possible string world-sheet bounded by the four strings with weight factor e^{-S}, S being the string tension multiplied by the generalized area of this surface (taking into account the fermionic degrees of freedom of the world-sheet). One such surface is shown in Figure 3. If we imagine the time axis running from left to right, then this diagram represents two strings joining into one string and then splitting into two strings, – the analog of a tree diagram in field theory. A more complicated surface is shown in Figure 4. This represents two strings joining into one string, which then splits into two and joins again, and finally splits into two strings. This is the analog of a one loop diagram in field theory. The relative normalization between the contributions from these two diagrams is not determined by any consistency requirement. This introduces an arbitrary parameter in string theory, known as the string coupling constant. However, once the relative normalization between these two diagrams is fixed, the relative normalization between all other diagrams is fixed due to various consistency requirement. Thus besides the dimensionful parameter α', string theory has a single dimensionless coupling constant. As we shall see later, both these parameters can be absorbed into definitions of various fields in the theory.

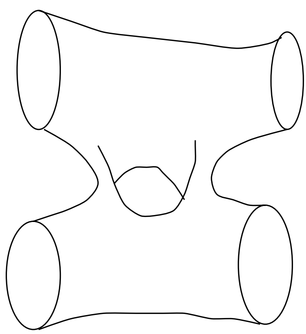


Fig. 4. A more complicated string world-sheet.

What we have described so far is the computation of the scattering amplitude with fixed locations of the external strings in space-time. The more relevant quantity is the scattering amplitude where the external strings are in the eigenstates of the energy and momenta operators conjugate to the coordinates of the $(9+1)$ dimensional space-time. This is done by simply taking the convolution of the above scattering amplitude with the wavefunctions of the strings corresponding to the external states. In practice there is an extremely efficient method of doing this computation using the so called vertex operators. It turns out that unlike in quantum field theory, all of these scattering amplitudes in string theory are ultraviolet finite. This is one of the major achievements of string theory.

Out main interest will be in the scattering involving the external massless states. The most convenient way to summarize the result of this computation in any string theory is to specify the effective action. By definition this effective action is such that if we compute the [illegible] scattering amplitude using this action, we should reproduce the S-matrix elements involving the massless states of string theory. In general such an action will have to contain infinite number of terms, but we can organise these terms by examining the number of space-time derivatives that appear in a given term in the action. Terms with the lowest number of derivatives constitute the [illegible] *ff* [illegible], – so called because this gives the dominant contribution if we want to evaluate the scattering amplitude when all the external particles have small energy and momenta.

The low energy effective action for all five string theories have been found. The actions for the type IIA and type IIB string theories correspond to those of two well known supergravity theories in ten space-time dimensions, called type IIA and type IIB supergravity theories respectively. On the other hand the actions for the three heterotic string theories correspond to another set of well-known supersymmetric theories in ten dimensions, – $N = 1$ supergravity coupled to $N = 1$ super Yang-Mills theory. For type I and the $SO(32)$ heterotic string theories the Yang-Mills gauge group is $SO(32)$ whereas for the $E_8 \times E_8$ heterotic string theory the gauge group is $E_8 \times E_8$. The emergence of gravity in all the five string theories is the most striking result in string theory. Its origin can be traced to the existence of the symmetric rank two tensor state (the graviton) in all these theories. This, combined with the result on finiteness of scattering amplitudes, shows that string theory gives us a finite quantum theory of gravity. We shall explicitly write down the low energy effective action of some of the string theories in Section 4.

The effective action of all five string theories are invariant under the transformation

$$\Phi \to \Phi - 2C, \qquad g_S \to \mathrm{e}^C g_S, \tag{2.1}$$

together with possible rescaling of other fields. Here Φ denotes the dilaton field, g_S denotes the string coupling, and C is an arbitrary constant. Using this scaling property, g_S can be absorbed in Φ. Put another way, the dimensionless coupling constant in string theory is related to the vacuum expectation value $\langle\Phi\rangle$ of Φ. The perturbative effective action does not have any potential for Φ, and hence $\langle\Phi\rangle$ can take arbitrary value. One expects that in a realistic string theory where supersymmetry is spontaneouly broken, there will be a potential for Φ, and hence $\langle\Phi\rangle$ will be determined uniquely.

In a similar vain one can argue that in string theory even the string tension, or equivalently the parameter α', has no physical significance. Since α' has the dimension of (length)^2 and is the only dimensionful parameter in the theory, the effective action will have an invariance under the simultaneous rescaling of α' and the metric $g_{\mu\nu}$:

$$\alpha' \to \lambda\alpha', \qquad g_{\mu\nu} \to \lambda g_{\mu\nu}\,, \tag{2.2}$$

together with possible rescaling of other fields. Using this scaling symmetry α' can be absorbed into the definition of $g_{\mu\nu}$. We shall discuss these two rescalings in detail in Section 4.1.

2.3 Compactification

So far we have described five different string theories, but they all live in ten space-time dimensions. Since our world is $(3+1)$ dimensional, these are not realistic string theories. However one can construct string theories in lower dimensions using the idea of compactification. The idea is to take the $(9+1)$ dimensional space-time as the product of a $(9-d)$ dimensional compact manifold $\mathcal{M}$ with euclidean signature and a $(d+1)$ dimensional Minkowski space $R^{d,1}$. Then, in the limit when the size of the compact manifold is sufficiently small so that the present day experiments cannot resolve this distance, the world will effectively appear to be $(d+1)$ dimensional. Choosing $d=3$ will give us a $(3+1)$ dimensional theory. Of course we cannot choose any arbitrary manifold $\mathcal{M}$ for this purpose; it must satisfy the equations of motion of the effective field theory that comes out of string theory. One also normally considers only those manifolds which preserve part of the space-time supersymmetry of the original ten dimensional theory, since this guarantees vanishing of the cosmological constant, and hence consistency of the corresponding string theory order by order in perturbation theory. There are many known examples of manifolds satisfying these restrictions [illegible] tori of different dimensions, K3, Calabi-Yau manifolds etc. Instead of going [illegible] the effective action, one can also directly describe these compactified theories as string theories. For this one needs to modify the string world-sheet action in such a way that it describes string propagation in the new manifold $\mathcal{M} \times R^{d,1}$, instead of in flat ten dimensional space-time. This modifies the world-sheet theory to an interacting non-linear σ-model instead of a free field theory. Consistency of string theory puts restriction on the kind of manifold on which the string can propagate. At the end both approaches yield identical results.

The simplest class of compact manifolds, on which we shall focus much of our attention in the rest of this article, are tori – product of circles. The effect of this compactification is to periodically identify some of the bosonic fields in the string world-sheet field theory – the fields which represent coordinates tangential to the compact circles. One effect of this is that the momentum carried by any string state along any of these circles is quantized in units of $1/R$ where R is the radius of the circle. But that is another novel effect: we now have new states that correspond to strings wrapped around a compact circle. For such a states, as we go once around the string, we also go once around the compact circle. These states are known as winding states and play a crucial role in the analysis of duality symmetries.

3 Notion of duality symmetries in string theory

In this section I shall elaborate the notion of duality symmetries, the difficulties in testing them, and the way of avoiding these difficulties. We begin by introducing the notion of duality in string theory.

3.1 Duality symmetries: Definition and examples

As was described in the last section, there are five consistent string theories in ten space-time dimensions. We also saw that we can get many different string theories in lower dimensions by compactifying these five theories on appropriate manifold $\mathcal{M}$. Each of these theories is parametrized by a set of parameters known as moduli[3] [illegible]

- String coupling constant (related to the vacuum expectation value of the dilaton field);
- Shape and size of $\mathcal{M}$ (information contained in the metric);
- various other background fields.

Inside the moduli space of the theory there is a certain region where the string coupling is weak and perturbation theory is valid. Elsewhere the theory is strongly coupled. This situation has been illustrated in Figure 5.

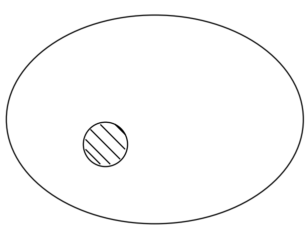


Fig. 5. A schematic representation of the moduli space of a string theory. The shaded region denotes the weak coupling region, whereas the white region denotes the strong coupling region.

String duality provides us with an equivalence map between two different string theories. In general this equivalence relation maps the weak coupling region of one theory to the strong coupling region of the second theory and [illegible]. This situation is illustrated in Figure 6.

[3]In string theory these moduli are related to vacuum expectation values of various dynamical fields and are expected to take definite values when supersymmetry is broken.

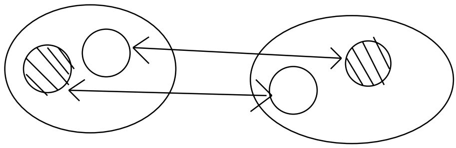


Fig. 6. A schematic representation of the duality map between the moduli spaces of two different string theories, A on K and B on K', where A and B are two of the five string theories in ten dimensions, and K, K' are two compact manifolds. Under this duality the weak coupling region of the first theory (denoted by the shaded region) gets mapped to the strong coupling region of the second theory and *vice versa.*

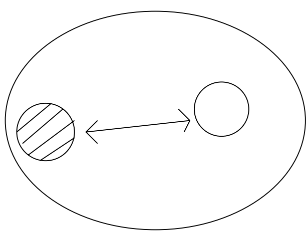


Fig. 7. Schematic representation of the moduli space of a self-dual theory. Duality relates weak and strong coupling regions of the same theory.

Before we proceed, let us give a few examples of dual pairs:

- Type I and $SO(32)$ heterotic string theories in $D = 10$ are conjectured to be dual to each other [19–22];
- Type IIA string theory compactified on K3 and heterotic string theory compactified on a four dimensional torus T^4 are conjectured to be dual to each other [19, 23–26][4].

Under duality, typically perturbation expansions get mixed up. Thus for example, tree level results in one theory might include perturbative and non-perturbative corrections in the dual theory. Also under duality, many of the elementary string states in one theory get mapped to solitons and their bound states in the dual theory.

[4]Throughout this article a string theory on $\mathcal{M}$ will mean string theory in the background $\mathcal{M} \times R^{9-n,1}$ where n is the real dimension of $\mathcal{M}$, and $R^{9-n,1}$ denotes $(10-n)$ dimensional Minkowski space.

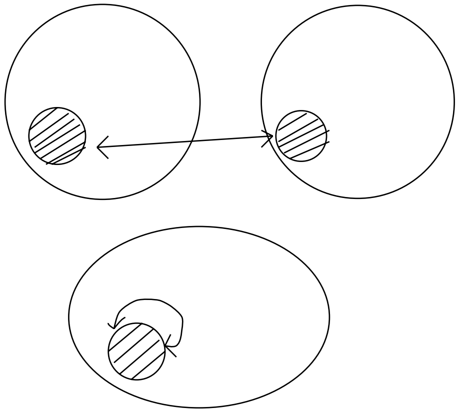


Fig. 8. Examples of T-duality relating a weakly coupled theory to a different or the same weakly coupled theory.

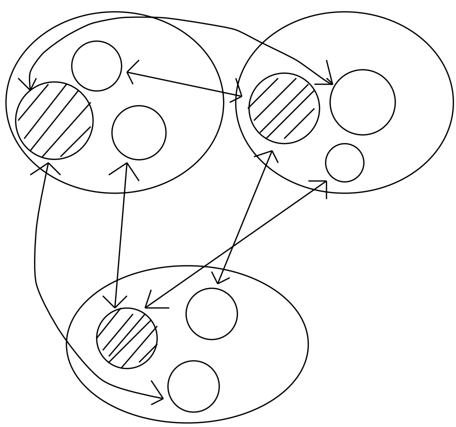


Fig. 9. A schematic representation of the moduli spaces of a chain of theories related by duality. In each case the shaded region denotes weak coupling region as usual.

Although duality in general relates the weak coupling limit of one theory to the strong coupling limit of another theory, there are special cases where the situation is a bit different. For example, we can have:

- Self-duality: here duality gives an equivalence relation between different regions of the moduli space of the same theory, as illustrated in Figure 7. In this case, duality transformations form a symmetry group that acts on the moduli space of the theory. For example, type IIB string theory in $D = 10$ is conjectured to have an $SL(2, Z)$ self-duality group [23].

- T-duality: in this case duality transformation maps the weak coupling region of one theory to the weak coupling region of another theory or the same theory as illustrated in Figure 8. For example, type IIA string theory compactified on a circle of radius R is dual to IIB string theory compactified on a circle of radius R^{-1} at the same value of the string coupling. Also, either of the two heterotic string theories compactified on a circle of radius R is dual to the same theory compactified on a circle of radius R^{-1} at the same value of the coupling constant. As a result the duality map does not mix up the perturbation expansions in the two theories. (For a review of this subject, see [27].)

In a generic situation duality can relate not just two theories, but a whole chain of theories, as illustrated in Figure 9. Thus for example, type IIA string theory compactified on K3 is related to heterotic string theory compactified on T^4. On the other hand, due to the equivalence of the $SO(32)$ heterotic and type I string theory in ten dimensions, $SO(32)$ heterotic string theory compactified on T^4 is related to type I string theory compactified on T^4. Thus these three theories are related by a chain of duality transformations.

From this discussion we see that the presence of duality in string theory has two important consequences. First of all, it reduces the degree of non-uniqueness of string theory, by relating various apparently unrelated (compactified) string theories. Furthermore, it allows us to study a strongly coupled string theory by mapping it to a weakly coupled dual theory whenever such a dual theory exists.

3.2 *Testing duality conjectures*

Let us now turn to the question of testing duality. As we have already emphasized, duality typically relates a weakly coupled string theory to a strongly coupled string theory. Thus in order to prove/test duality we must be able to analyze at least one of the theories at strong coupling. But in string theory we only know how to define the theory perturbatively at weak coupling. Thus it would seem impossible to prove or test any duality conjecture in string theory[5]. This is where supersymmetry comes to our

[5]Note that this problem is absent for T-duality transformations which relates two weakly coupled string theories, and hence can be tested using string perturbation theory. All T-duality symmetries in string theory can be “proved” this way, at least to all orders in perturbation theory.

rescue. Supersymmetry gives rise to certain non-renormalization theorems in string theory, due to which some of the weak coupling calculations can be trusted even at strong coupling. Thus we can focus our attention on such "non-renormalized" quantities and ask if they are invariant under the proposed duality transformations. Testing duality invariance of these quantities provides us with various tests of various duality conjectures, and is in fact the basis of all duality conjectures.

The precise content of these non-renormalization theorems depends on the number of supersymmetries present in the theory. The maximum number of supersymmetry generators that can be present in a string theory is 32. This gives $N = 2$ supersymmetry in ten dimensions, and $N = 8$ supersymmetry in four dimensions. Examples of such theories are types IIA or type IIB string theories compactified on n dimensional tori T^n. The next interesting class of theories are those with 16 supersymmetry generators. This corresponds to $N = 1$ supersymmetry in ten dimensions and $N = 4$ supersymmetry in four dimensions. Examples of such theories are types IIA or type IIB string theories compactified on $K3 \times T^n$, heterotic string theory compactified on T^n, etc. Another class of theories that we shall discuss are those with eight supersymmetry generators, heterotic string theory on $K3 \times T^n$, type IIA or IIB string theory on six dimensional Calabi-Yau manifolds, etc. For theories with 16 or more SUSY generators the non-renormalization theorems are particularly powerful. In particular,

- Form of the low energy effective action involving the massless states of the theory is completely fixed by the requirement of supersymmetry (and the spectrum) [28]. Thus this effective action cannot get renormalized by string loop corrections. As a result, any valid symmetry of the theory must be a symmetry of this effective field theory;

- These theories contain special class of states which are invariant under part of the supersymmetry transformations. They are known as BPS states, named after Bogomol'nyi, Prasad and Sommerfeld. The mass of a BPS state is completely determined in terms of its charge as a consequence of the supersymmetry algebra. Since this relation is derived purely from an analysis of the supersymmetry algebra, it is not modified by quantum corrections. Furthermore it can be argued that the degeneracy of BPS states of a given charge does not change as we move in the moduli space even from weak to strong coupling region [29]. Thus the spectrum of BPS states can be calculated from weak coupling analysis and the result can be continued to the strong coupling region. Since any valid symmetry of the theory must be a symmetry of the spectrum of BPS states, we can use this to design non-trivial tests of duality [1].

For theories with eight supersymmetries the non-renormalization theorems are less powerful. However, even in this case one can design non-trivial tests of various duality conjectures. We shall discuss these in Section 7.

4 Analysis of low energy effective field theory

In this section I shall discuss tests of various dualities in string theories with $\geq$16 supersymmetries based on the analysis of their low energy effective action. As has been emphasized in the previous section, the form of this low energy effective action is determined completely by the requirement of supersymmetry and the spectrum of massless states in the theory. Thus it does not receive any quantum corrections, and if a given duality transformation is to be a symmetry of a string theory, it must be a symmetry of the corresponding low energy effective action. Actually, since the low energy effective action is to be used only for deriving the equations of motion from this action, and/or computing the tree level S-matrix elements using this action, but not to perform a full-fledged path integral, it is enough that only the equations of motion derived from this action are invariant under duality transformations. (This also guarantees that the tree level S-matrix elements computed from this effective action are invariant under the duality transformations.) It is not necessary for the action itself to be invariant.

Throughout this article we shall denote by $G_{\mu\nu}$ the string metric $-$ the metric that is used in computing the area of the string world-sheet embedded in space time for calculating string scattering amplitudes. For a string theory compactified on a $(9-d)$ dimensional manifold $\mathcal{M}$, we shall denote by Φ the shifted dilaton, related to the dilaton $\Phi^{(10)}$ of the ten dimensional string theory as

$$\Phi = \Phi^{(10)} - \ln V \, , \tag{4.1}$$

where $(2\pi)^{9-d}\, V$ is the volume of $\mathcal{M}$ measured in the ten dimensional string metric. The dilaton is normalized in such a way that $\mathrm{e}^{\langle \Phi^{(10)} \rangle}$ corresponds to the square of the closed string coupling constant in ten dimensions[6]. $g_{\mu\nu}$ will denote the canonical Einstein metric which is related to the string metric by an appropriate conformal rescaling involving the dilaton field,

$$g_{\mu\nu} = \mathrm{e}^{-\frac{2}{d-1}\Phi} G_{\mu\nu} \, . \tag{4.2}$$

We shall always use this metric to raise and lower indices. The signature of space-time will be taken as $(-, +, \cdots +)$. Finally, all fields will be made dimensionless by absorbing appropriate powers of α' in them.

[6] Φ is related to the more commonly normalized dilaton ϕ by a factor of two: $\Phi = 2\phi$.

We shall now consider several examples. The discussion will closely follow references [1,19,23]. For a detailed review of the material covered in this section, see reference [15].

4.1 Type I–$SO(32)$ heterotic duality in $D = 10$

In $SO(32)$ heterotic string theory, the massless bosonic states come from the NS sector of the closed heterotic string, and contains the metric $g^{(H)}_{\mu\nu}$, the dilaton $\Phi^{(H)}$, the rank two anti-symmetric tensor field $B^{(H)}_{\mu\nu}$, and gauge fields $A^{(H)a}_{\mu}$ ($1 \leq a \leq 496$) in the adjoint representation of $SO(32)$. The low energy dynamics involving these massless bosonic fields is described by the $N = 1$ supergravity coupled to $SO(32)$ super Yang-Mills theory in ten dimensions [114]. The action is given by [102]:

$$\begin{aligned} S^{(\mathrm{H})} \quad = \quad & \frac{1}{(2\pi)^7(\alpha'_{\mathrm{H}})^4 g_{\mathrm{H}}^2} \int \mathrm{d}^{10}x\sqrt{-g^{(\mathrm{H})}}\Big[R^{(\mathrm{H})} - \frac{1}{8}g^{(\mathrm{H})\mu\nu}\partial_\mu\Phi^{(\mathrm{H})}\partial_\nu\Phi^{(\mathrm{H})} \\ & -\frac{1}{4}g^{(\mathrm{H})\mu\mu'}g^{(\mathrm{H})\nu\nu'}\mathrm{e}^{-\Phi^{(\mathrm{H})}/4}\mathrm{Tr}(F^{(\mathrm{H})}_{\mu\nu}F^{(\mathrm{H})}_{\mu'\nu'}) \\ & -\frac{1}{12}g^{(\mathrm{H})\mu\mu'}g^{(\mathrm{H})\nu\nu'}g^{(\mathrm{H})\rho\rho'}\mathrm{e}^{-\Phi^{(\mathrm{H})}/2}H^{(\mathrm{H})}_{\mu\nu\rho}H^{(\mathrm{H})}_{\mu'\nu'\rho'}\Big]\,, \end{aligned} \tag{4.3}$$

where $R^{(\mathrm{H})}$ is the Ricci scalar, $F^{(\mathrm{H})}_{\mu\nu}$ denotes the non-abelian gauge field strength,

$$F^{(\mathrm{H})}_{\mu\nu} = \partial_\mu A^{(\mathrm{H})}_\nu - \partial_\nu A^{(\mathrm{H})}_\mu + \sqrt{\frac{2}{\alpha'_{\mathrm{H}}}}\,[A^{(\mathrm{H})}_\mu, A^{(\mathrm{H})}_\nu]\,, \tag{4.4}$$

Tr denotes trace in the vector representation of $SO(32)$, and $H^{(\mathrm{H})}_{\mu\nu\rho}$ is the field strength associated with the $B^{(\mathrm{H})}_{\mu\nu}$ field:

$$\begin{aligned} H^{(\mathrm{H})}_{\mu\nu\rho} \quad = \quad & \partial_\mu B^{(\mathrm{H})}_{\nu\rho} - \frac{1}{2}\mathrm{Tr}\left(A^{(\mathrm{H})}_\mu F^{(\mathrm{H})}_{\nu\rho} - \frac{1}{3}\sqrt{\frac{2}{\alpha'_{\mathrm{H}}}}\,A^{(\mathrm{H})}_\mu[A^{(\mathrm{H})}_\nu, A^{(\mathrm{H})}_\rho]\right) \\ & +\text{cyclic permutations of } \mu,\nu,\rho\,. \end{aligned} \tag{4.5}$$

$2\pi\alpha'_{\mathrm{H}}$ and g_{H} are respectively the inverse string tension and the coupling constant of the heterotic string theory. The rescalings (2.1), (2.2) take the following form acting on the complete set of fields:

$$\begin{aligned} & g_{\mathrm{H}} \to \mathrm{e}^{C}g_{\mathrm{H}}, \qquad \Phi^{(\mathrm{H})} \to \Phi^{(\mathrm{H})} - 2C, \qquad g^{(\mathrm{H})}_{\mu\nu} \to \mathrm{e}^{C/2}g^{(\mathrm{H})}_{\mu\nu} \\ & B^{(\mathrm{H})}_{\mu\nu} \to B^{(\mathrm{H})}_{\mu\nu}, \qquad A^{(\mathrm{H})a}_\mu \to A^{(\mathrm{H})a}_\mu\,, \end{aligned} \tag{4.6}$$

$$\begin{aligned}&\alpha'_{\rm H}\to\lambda\alpha'_{\rm H}, \qquad \Phi^{(\rm H)}\to\Phi^{(\rm H)}, \qquad g^{(\rm H)}_{\mu\nu}\to\lambda g^{(\rm H)}_{\mu\nu}\\ &B^{(\rm H)}_{\mu\nu}\to\lambda B^{(\rm H)}_{\mu\nu}, \qquad A^{(\rm H)a}_{\mu}\to\lambda^{1/2}A^{(\rm H)a}_{\mu}\,,\end{aligned} \tag{4.7}$$

Since $g_{\rm H}$ and $\alpha'_{\rm H}$ can be changed by this rescaling, these parameters cannot have a universal significance. In particular, we can absorb $g_{\rm H}$ and $\alpha'_{\rm H}$ into the various fields by setting $\mathrm{e}^{-C}=g_{\rm H}$ and $\lambda=(\alpha'_{\rm H})^{-1}$ in (4.6), (4.7). This is equivalent to setting $g_{\rm H}=1$ and $\alpha'_{\rm H}=1$. In this notation the physical coupling constant is given by the vacuum expectation value of $\mathrm{e}^{\Phi^{(\rm H)}/2}$, and the ADM mass per unit length of an infinitely long straight string, measured in the metric $\mathrm{e}^{\langle\Phi^{(\rm H)}\rangle/4}g^{(\rm H)}_{\mu\nu}$ that approaches the string metric $G^{(\rm H)}_{\mu\nu}$ far away from the string, is equal to $1/2\pi$. By changing $\langle\Phi^{(\rm H)}\rangle$ we can get all possible values of string coupling, and using a metric that differs from the one used here by a constant multiplicative factor, we can get all possible values of the string tension.

For $\alpha'_{\rm H}=1$ and $g_{\rm H}=1$ equations (4.3)–(4.5) take the form:

$$\begin{aligned}S^{(\rm H)} &= \frac{1}{(2\pi)^7}\int \mathrm{d}^{10}x\sqrt{-g^{(\rm H)}}\Big[R^{(\rm H)}-\frac{1}{8}g^{(\rm H)\mu\nu}\partial_\mu\Phi^{(\rm H)}\partial_\nu\Phi^{(\rm H)}\\ &\quad -\frac{1}{4}g^{(\rm H)\mu\mu'}g^{(\rm H)\nu\nu'}\mathrm{e}^{-\Phi^{(\rm H)}/4}\mathrm{Tr}(F^{(\rm H)}_{\mu\nu}F^{(\rm H)}_{\mu'\nu'})\\ &\quad -\frac{1}{12}g^{(\rm H)\mu\mu'}g^{(\rm H)\nu\nu'}g^{(\rm H)\rho\rho'}\mathrm{e}^{-\Phi^{(\rm H)}/2}H^{(\rm H)}_{\mu\nu\rho}H^{(\rm H)}_{\mu'\nu'\rho'}\Big]\,,\end{aligned} \tag{4.8}$$

$$F^{(\rm H)}_{\mu\nu}=\partial_\mu A^{(\rm H)}_\nu-\partial_\nu A^{(\rm H)}_\mu+\sqrt{2}[A^{(\rm H)}_\mu,A^{(\rm H)}_\nu]\,, \tag{4.9}$$

$$\begin{aligned}H^{(\rm H)}_{\mu\nu\rho} &= \partial_\mu B^{(\rm H)}_{\nu\rho}-\frac{1}{2}\mathrm{Tr}\Big(A^{(\rm H)}_\mu F^{(\rm H)}_{\nu\rho}-\frac{\sqrt{2}}{3}A^{(\rm H)}_\mu[A^{(\rm H)}_\nu,A^{(\rm H)}_\rho]\Big)\\ &\quad +\text{cyclic permutations of }\mu,\nu,\rho\,.\end{aligned} \tag{4.10}$$

Let us now turn to the type I string theory. The massless bosonic states in type I theory come from three different sectors. The closed string Neveu-Schwarz – Neveu-Schwarz (NS) sector gives the metric $g^{(\rm I)}_{\mu\nu}$ and the dilaton $\Phi^{(\rm I)}$. The closed string Ramond-Ramond (RR) sector gives an anti-symmetric tensor field $B^{(\rm I)}_{\mu\nu}$. Besides these, there are bosonic fields coming from the NS sector of the open string. This sector gives rise to gauge fields $A^{(\rm I)a}_\mu$ ($a=1,\ldots 496$) in the adjoint representation of the group $SO(32)$. (The superscript (I) refers to the fact that these are the fields in the type I string theory.) The low energy dynamics is again described by the $N=1$ supergravity theory coupled to $SO(32)$ super Yang-Mills theory [115]. But it is instructive to rewrite the effective action in terms of the type I variables. For suitable choice of the string tension and the coupling constant, this is given by [102]

$$
\begin{aligned}
S^{(\mathrm{I})} \quad = \quad & \frac{1}{(2\pi)^7}\int \mathrm{d}^{10}x\sqrt{-g^{(\mathrm{I})}}\Big[R^{(\mathrm{I})} - \frac{1}{8}g^{(\mathrm{I})\mu\nu}\partial_\mu\Phi^{(\mathrm{I})}\partial_\nu\Phi^{(\mathrm{I})} \\
& -\frac{1}{4}g^{(\mathrm{I})\mu\mu'}g^{(\mathrm{I})\nu\nu'}\mathrm{e}^{\Phi^{(\mathrm{I})}/4}\mathrm{Tr}(F^{(\mathrm{I})}_{\mu\nu}F^{(\mathrm{I})}_{\mu'\nu'}) \\
& -\frac{1}{12}g^{(\mathrm{I})\mu\mu'}g^{(\mathrm{I})\nu\nu'}g^{(\mathrm{I})\rho\rho'}\mathrm{e}^{\Phi^{(\mathrm{I})}/2}H^{(\mathrm{I})}_{\mu\nu\rho}H^{(\mathrm{I})}_{\mu'\nu'\rho'}\Big]\,,
\end{aligned}
\tag{4.11}
$$

where $R^{(\mathrm{I})}$ is the Ricci scalar, $F^{(\mathrm{I})}_{\mu\nu}$ denotes the non-abelian gauge field strength,

$$
F^{(\mathrm{I})}_{\mu\nu} = \partial_\mu A^{(\mathrm{I})}_\nu - \partial_\nu A^{(\mathrm{I})}_\mu + \sqrt{2}[A^{(\mathrm{I})}_\mu, A^{(\mathrm{I})}_\nu]\,, \tag{4.12}
$$

and $H^{(\mathrm{I})}_{\mu\nu\rho}$ is the field strength associated with the $B^{(\mathrm{I})}_{\mu\nu}$ field:

$$
\begin{aligned}
H^{(\mathrm{I})}_{\mu\nu\rho} \quad = \quad & \partial_\mu B^{(\mathrm{I})}_{\nu\rho} - \frac{1}{2}\mathrm{Tr}\Big(A^{(\mathrm{I})}_\mu F^{(\mathrm{I})}_{\nu\rho} - \frac{\sqrt{2}}{3}A^{(\mathrm{I})}_\mu[A^{(\mathrm{I})}_\nu, A^{(\mathrm{I})}_\rho]\Big) \\
& +\text{cyclic permutations of } \mu, \nu, \rho\,.
\end{aligned}
\tag{4.13}
$$

For both, the type I and the $SO(32)$ heterotic string theory, the low energy effective action is derived from the string tree level analysis. However, to this order in the derivatives, the form of the effective action is determined completely by the requirement of supersymmetry for a given gauge group. Thus neither action can receive any quantum corrections.

It is straightforward to see that the actions (4.8) and (4.11) are identical provided we make the identification:

$$
\begin{aligned}
\Phi^{(\mathrm{H})} &= -\Phi^{(\mathrm{I})}, \qquad & g^{(\mathrm{H})}_{\mu\nu} &= g^{(\mathrm{I})}_{\mu\nu} \\
B^{(\mathrm{H})}_{\mu\nu} &= B^{(\mathrm{I})}_{\mu\nu}, \qquad & A^{(\mathrm{H})a}_{\mu} &= A^{(\mathrm{I})a}_{\mu}\,.
\end{aligned}
\tag{4.14}
$$

This led to the hypothesis that the type I and the $SO(32)$ heterotic string theories in ten dimensions are equivalent [19]. One can find stronger evidence for this hypothesis by analysing the spectrum of supersymmetris states, but the equivalence of the two effective actions was the reason for proposing this duality in the first place.

Note the $-$ sign in the relation between $\Phi^{(\mathrm{H})}$ and $\Phi^{(\mathrm{I})}$ in equation (4.14). Recalling that $\mathrm{e}^{\langle\Phi\rangle/2}$ is the string coupling, we see that the strong coupling limit of one theory is related to the weak coupling limit of the other theory and

From now on I shall use the unit $\alpha' = 1$ for writing down the effective action of all string theories. Physically this would mean that the ADM mass per unit length of a test string, measured in the metric $e^{2\langle\Phi\rangle/(d-1)} g_{\mu\nu}$ that agrees with the string metric $G_{\mu\nu}$ defined in (4.2) far away from the test string, is given by $1/2\pi$. In future we shall refer to the ADM mass of a particle measured in this metric as the mass measured in the string metric.

4.2 Self-duality of heterotic string theory on T^6

In the previous subsection we have described the massless bosonic field content of the ten dimensional $SO(32)$ heterotic string theory. When we compactify it on a six dimensional torus, we can get many other massless scalar fields from the internal components of the metric, the anti-symmetric tensor field and the gauge fields in the Cartan subalgebra of the gauge group[7]. This gives a total of $(21 + 15 + 96 = 132)$ scalar fields. It turns out that these scalars can be represented by a 28×28 matrix valued field M satisfying[8]

$$MLM^{\mathrm{T}} = L, \qquad M^{\mathrm{T}} = M\,, \tag{4.15}$$

where

$$L = \begin{pmatrix} & I_6 & \\ I_6 & & \\ & & -I_{16} \end{pmatrix}. \tag{4.16}$$

I_n denotes an $n \times n$ identity matrix. We shall choose a convention in which $M = I_{28}$ corresponds to a compactification on $(S^1)^6$ with each S^1 having radius $\sqrt{\alpha'} = 1$ measured in the string metric, and without any background gauge or antisymmetric tensor fields. We can get another scalar field a by dualizing the gauge invariant field strength H of the antisymmetrix tensor field through the relation:

$$H^{\mu\nu\rho} = -(\sqrt{-g})^{-1} e^{2\Phi} \epsilon^{\mu\nu\rho\sigma} \partial_\sigma a\,, \tag{4.17}$$

where Φ denotes the four dimensional dilaton and $g_{\mu\nu}$ denotes the $(3+1)$ dimensional canonical metric defined in equations (4.1), (4.2) respectively. It is convenient to combine the dilaton Φ and the axion field a into a single complex scalar λ:

$$\lambda = a + ie^{-\Phi} \equiv \lambda_1 + i\lambda_2\,. \tag{4.18}$$

[7]Only the sixteen gauge fields in the Cartan subalgebra of the gauge group can develop vacuum expectation value since such vacuum expectation values do not generate any field strength, and hence do not generate energy density.

[8]For a review of this construction, see [1].

At a generic point in the moduli space, where the scalars M take arbitrary vacuum expectation values, the non-abelian gauge symmetry of the ten dimensional theory is broken to its abelian subgroup $U(1)^{16}$. Besides these sixteen $U(1)$ gauge fields we get twelve other $U(1)$ gauge fields from components $G_{m\mu}$, $B_{m\mu}$ ($4 \le m \le 9$, $0 \le \mu \le 3$) of the metric and the anti-symmetric tensor field respectively. Let us denote these 28 $U(1)$ gauge fields (after suitable normalization) by A^a_μ ($1 \le a \le 28$). In terms of these fields, the low energy effective action of the theory is given by [1,30–32,34][9],

$$\begin{aligned} S &= \frac{1}{2\pi}\int \mathrm{d}^4x\sqrt{-g}\Big[R - g^{\mu\nu}\frac{\partial_\mu\lambda\partial_\nu\bar\lambda}{2(\lambda_2)^2} + \frac{1}{8}g^{\mu\nu}\mathrm{Tr}(\partial_\mu ML\partial_\nu ML) \\ &\quad -\frac{1}{4}\lambda_2 g^{\mu\mu'}g^{\nu\nu'}F^a_{\mu\nu}(LML)_{ab}F^b_{\mu'\nu'} + \frac{1}{4}\lambda_1 g^{\mu\rho}g^{\nu\sigma}F^a_{\mu\nu}L_{ab}\widetilde{F}^b_{\rho\sigma}\Big]\,, \end{aligned} \tag{4.19}$$

where $F^a_{\mu\nu}$ is the field strength associated with A^a_μ, R is the Ricci scalar. and

$$\widetilde{F}^{a\mu\nu} = \frac{1}{2}(\sqrt{-g})^{-1}\epsilon^{\mu\nu\rho\sigma}F^a_{\rho\sigma}\,. \tag{4.20}$$

This action is invariant under an O(6,22) transformation[10]:

$$M \to \Omega M \Omega^{\mathrm{T}}, \qquad A^a_\mu \to \Omega_{ab}A^b_\mu, \qquad g_{\mu\nu}\to g_{\mu\nu}, \qquad \lambda\to\lambda\,, \tag{4.21}$$

where Ω satisfies:

$$\Omega L \Omega^{\mathrm{T}} = L\,. \tag{4.22}$$

An $O(6{,}22;Z)$ subgroup of this can be shown to be a T-duality symmetry of the full string theory [27]. This $O(6{,}22;Z)$ subgroup can be described as follows. Let Λ_{28} denote a twenty eight dimensional lattice obtained by taking the direct sum of the twelve dimensional lattice of integers, and the sixteen dimensional root lattice of $SO(32)$[11]. $O(6{,}22;Z)$ is defined to be the subset of $O(6{,}22)$ transformations which leave Λ_{28} invariant,. acting on any vector in Λ_{28}, produces another vector in Λ_{28}. It will be useful

[9] The normalization of the gauge fields used here differ from that in reference [1] by a factor of two. Also there we used $\alpha' = 16$ whereas here we are using $\alpha' = 1$.

[10] $O(p,q)$ denotes the group of Lorentz transformations in p space-like and q time-like dimensions. (These have nothing to do with physical space-time, which always has only one time-like direction.) $O(p,q;Z)$ denotes a discrete subgroup of $O(p,q)$.

[11] More precisely we have to take the root lattice of $Spin(32)/Z_2$ which is obtained by adding to the $SO(32)$ root lattice the weight vectors of the spinor representations of $SO(32)$ with a definite chirality.

for our future reference to undertstand why only an $O(6{,}22;Z)$ subgroup of the full $O(6{,}22)$ group is a symmetry of the full string theory. Since $O(6{,}22;Z)$ is a T-duality symmetry, this question can be answered within the context of perturbative string theory. The point is that although at a generic point in the moduli space the massless string states do not carry any charge, there are massive charged states in the spectrum of full string theory. Since there are 28 charges associated with the 28 $U(1)$ gauge fields, a state can be characterized by a 28 dimensional charge vector. With appropriate normalization, this charge vector can be shown to lie in the lattice Λ_{28}, the charge vector of any state in the spectrum can be shown to be an element of the lattice Λ_{28}. Since the $O(6{,}22)$ transformation acts linearly on the $U(1)$ gauge fields, it also acts linearly on the charge vectors. As a result only those $O(6{,}22)$ elements can be genuine symmetries of string theory which preserve the lattice Λ_{28}. Any other $O(6{,}22)$ element, acting on a physical state in the spectrum, will take it to a state with charge vector outside the lattice Λ_{28}. Since such a state does not exist in the spectrum, such an $O(6{,}22)$ transformation cannot be a symmetry of the full string theory.

In order to see a specific example of a T-duality transformation, let us consider heterotic string theory compactified on $(S^1)^6$ with one of the circles having radius R measured in the string metric, and the rest having unit radius. Let us also assume that there is no background gauge or anti-symmetric tensor fields. Using the convention of reference [1] one can show that for this background

$$M^{(\mathrm{H})} = \begin{pmatrix} R^{-2} & & & & \\ & I_5 & & & \\ & & R^2 & & \\ & & & I_5 & \\ & & & & I_{16} \end{pmatrix}. \tag{4.23}$$

Consider now the $O(6{,}22;Z)$ transformation with the matrix:

$$\Omega = \begin{pmatrix} 0 & & 1 & \\ & I_5 & & \\ 1 & & 0 & \\ & & & I_{21} \end{pmatrix}. \tag{4.24}$$

Using equation (4.21) we see that this transforms $M^{(\mathrm{H})}$ to

$$M^{(\mathrm{H})} = \begin{pmatrix} R^{2} & & & & \\ & I_5 & & & \\ & & R^{-2} & & \\ & & & I_5 & \\ & & & & I_{16} \end{pmatrix}. \tag{4.25}$$

Thus the net effect of this transformation is $R \to R^{-1}$. It says that the heterotic string theory compactified on a circle of radius R is equivalent to the same theory compactified on a circle of radius R^{-1}. For this reason $R = 1$ ($R = \sqrt{\alpha'}$) is known as the self-dual radius. Other $O(6,22;Z)$ transformations acting on (4.23) will give rise to more complicated $M^{(\mathrm{H})}$ corresponding to a configuration with background gauge and/or anti-symmetric tensor fields.

Besides this symmetry, the equations of motion derived from this action can be shown to be invariant under an $SL(2,R)$ transformation of the form [30, 35, 36]

$$
\begin{aligned}
&F^a_{\mu\nu} \to (r\lambda_1 + s)F^a_{\mu\nu} + r\lambda_2 (ML)_{ab}\widetilde{F}^b_{\mu\nu}, \qquad \lambda \to \frac{p\lambda + q}{r\lambda + s}, \\
&g_{\mu\nu} \to g_{\mu\nu}, \qquad M \to M\,,
\end{aligned}
\tag{4.26}
$$

where p, q, r, s are real numbers satisfying $ps - qr = 1$. The existence of such symmetries (known as hidden non-compact symmetries) in this and in other supergravity theories were discovered in early days of supergravity theories and in fact played a crucial role in the construction of these theories in the first place [30, 113]. Since this $SL(2,R)$ transformation mixes the gauge field strength with its Poincare dual, it is an electric-magnetic duality transformation. This leads to the conjecture that a subgroup of this continuous symmetry group is an exact symmetry of string theory [1,36–42]. One might wonder why the conjecture refers to only a discrete subgroup of $SL(2,R)$ instead of the full $SL(2,R)$ group as the genuine symmetry group. This follows from the same logic that was responsible for breaking $O(6,22)$ to $O(6,22;Z)$; however since the $SL(2,R)$ transformation mixes electric field with magnetic field, we now need to take into account the quantization of magnetic charges. We have already described the quantization condition on the electric charges. Using the usual Dirac-Schwinger-Zwanziger rules one can show that in appropriate normalization, the 28 dimensional magnetic charge vectors also lie in the same lattice Λ_{28}. Also with this normalization convention the electric and magnetic charge vectors transform as doublet under the $SL(2,R)$ transformation; thus it is clear that the subgroup of $SL(2,R)$ that respects the charge quantization condition is $SL(2,Z)$. An arbitrary $SL(2,R)$ transformation acting on the quantized electric and magnetic charges will not give rise to electric and magnetic charges consistent with the quantization law. This is the reason behind the conjectured $SL(2,Z)$ symmetry of heterotic string theory on T^6. Note that since this duality acts non-trivially on the dilaton and hence the string coupling, this is a non-perturbative symmetry, and cannot be verified order by order in perturbation theory. Historically, this is the first example of a concrete duality conjecture in string theory. Later we shall review other tests of this duality conjecture.

4.3 Duality between heterotic on T^4 and type IIA on K3

The massless bosonic field content of heterotic string theory compactified on T^4 can be found in a manner identical to that in heterotic string theory on T^6. Besides the dilaton $\Phi^{(\mathrm{H})}$, we get many other massless scalar fields from the internal components of the metric, the anti-symmetric tensor field and the gauge fields. In this case these scalars can be represented by a 24×24 matrix valued field $M^{(\mathrm{H})}$ satisfying

$$M^{(\mathrm{H})} L M^{(\mathrm{H})\mathrm{T}} = L, \qquad M^{(\mathrm{H})\mathrm{T}} = M^{(\mathrm{H})}\,, \tag{4.27}$$

where

$$L = \begin{pmatrix} & I_4 & \\ I_4 & & \\ & & -I_{16}\,. \end{pmatrix} \tag{4.28}$$

We again use the convention that $M^{(\mathrm{H})} = I_{24}$ corresponds to compactification on $(S^1)^4$ with each S^1 having self-dual radius ($\sqrt{\alpha'} = 1$), without any background gauge field or anti-symmetric tensor field. At a generic point in the moduli space, where the scalars $M^{(\mathrm{H})}$ take arbitrary vacuum expectation values, we get a $U(1)^{24}$ gauge group, with 16 gauge fields coming from the Cartan subalgebra of the original gauge group in ten dimensions, and eight other gauge fields from components $G_{m\mu}$, $B_{m\mu}$ ($6 \le m \le 9$, $0 \le \mu \le 5$) of the metric and the anti-symmetric tensor field respectively. Here x^m denote the compact directions, and x^μ denote the non-compact directions. Let us denote these 24 $U(1)$ gauge fields by $A_\mu^{(\mathrm{H})a}$ ($1 \le a \le 24$). Finally, let $g_{\mu\nu}^{(\mathrm{H})}$ and $B_{\mu\nu}^{(\mathrm{H})}$ denote the canonical metric and the anti-symmetric tensor field respectively. In terms of these fields, the low energy effective action of the theory is given by,

$$\begin{aligned} S_{\mathrm{H}} \;=\; & \frac{1}{(2\pi)^3} \int \mathrm{d}^6 x \sqrt{-g^{(\mathrm{H})}} \Big[R^{(\mathrm{H})} - \frac{1}{2} g^{(\mathrm{H})\mu\nu} \partial_\mu \Phi^{(\mathrm{H})} \partial_\nu \Phi^{(\mathrm{H})} \\ & + \frac{1}{8} g^{\mu\nu} \mathrm{Tr}(\partial_\mu M^{(\mathrm{H})} L \partial_\nu M^{(\mathrm{H})} L) \\ & - \frac{1}{4} \mathrm{e}^{-\Phi^{(\mathrm{H})}/2} g^{(\mathrm{H})\mu\mu'} g^{(\mathrm{H})\nu\nu'} F_{\mu\nu}^{(\mathrm{H})a} (L M^{(\mathrm{H})} L)_{ab} F_{\mu'\nu'}^{(\mathrm{H})b} \\ & - \frac{1}{12} \mathrm{e}^{-\Phi^{(\mathrm{H})}} g^{(\mathrm{H})\mu\mu'} g^{(\mathrm{H})\nu\nu'} g^{(\mathrm{H})\rho\rho'} H_{\mu\nu\rho}^{(\mathrm{H})} H_{\mu'\nu'\rho'}^{(\mathrm{H})} \Big]\,, \end{aligned} \tag{4.29}$$

where $F_{\mu\nu}^{(\mathrm{H})a}$ is the field strength associated with $A_\mu^{(\mathrm{H})a}$, $R^{(\mathrm{H})}$ is the Ricci scalar, and $H_{\mu\nu\rho}^{(\mathrm{H})}$ is the field strength associated with $B_{\mu\nu}^{(\mathrm{H})}$:

$$H_{\mu\nu\rho}^{(\mathrm{H})} = \left(\partial_\mu B_{\nu\rho}^{(\mathrm{H})} + \frac{1}{2} A_\mu^{(\mathrm{H})a} L_{ab} F_{\nu\rho}^{(\mathrm{H})b}\right) + (\text{cyclic permutations of } \mu, \nu, \rho). \tag{4.30}$$

This action is invariant under an $O(4{,}20)$ transformation:

$$\begin{aligned} &M^{(\mathrm{H})} \to \Omega M^{(\mathrm{H})} \Omega^{\mathrm{T}}, \qquad A_\mu^{(\mathrm{H})a} \to \Omega_{ab} A_\mu^{(\mathrm{H})b}, \qquad g_{\mu\nu}^{(\mathrm{H})} \to g_{\mu\nu}^{(\mathrm{H})}, \\ &B_{\mu\nu}^{(\mathrm{H})} \to B_{\mu\nu}^{(\mathrm{H})}, \qquad \Phi^{(\mathrm{H})} \to \Phi^{(\mathrm{H})}, \end{aligned} \tag{4.31}$$

where Ω satisfies:

$$\Omega L \Omega^{\mathrm{T}} = L\,. \tag{4.32}$$

Again as in the case of T^6 compactification, only an $O(4{,}20;Z)$ subgroup of this which preserves the charge lattice Λ_{24} is an exact T-duality symmetry of this theory. The lattice Λ_{24} is obtained by taking the direct sum of the 8 dimensional lattice of integers and the root lattice of $Spin(32)/Z_2$.

Let us now turn to the spectrum of massless bosonic fields in type IIA string theory on K3. In ten dimensions the massless bosonic fields in type IIA string theory are the metric g_{MN}, the rank two anti-symmetric tensor B_{MN} and the scalar dilation Φ coming from the NS sector, and a gauge field A_{M} and a rank three antisymmetric tensor field C_{MNP} coming from the RR sector. The low energy effective action of this theory involving the massless bosonic fields is given by [90]

$$\begin{aligned} S_{\mathrm{IIA}} \;=\; & \frac{1}{(2\pi)^7}\int \mathrm{d}^{10}x\sqrt{-g}\Big[R - \frac{1}{8} g^{\mu\nu}\partial_\mu\Phi\partial_\nu\Phi \\ & -\frac{1}{12}\mathrm{e}^{-\Phi/2} g^{\mu\mu'} g^{\nu\nu'} g^{\rho\rho'} H_{\mu\nu\rho} H_{\mu'\nu'\rho'} - \frac{1}{4}\mathrm{e}^{3\Phi/4} g^{\mu\mu'} g^{\nu\nu'} F_{\mu\nu} F_{\mu'\nu'} \\ & -\frac{1}{48}\mathrm{e}^{\Phi/4} g^{\mu\mu'} g^{\nu\nu'} g^{\rho\rho'} g^{\sigma\sigma'} G_{\mu\nu\rho\sigma} G_{\mu'\nu'\rho'\sigma'} \\ & -\frac{1}{(48)^2}(\sqrt{-g})^{-1}\varepsilon^{\mu_0\cdots\mu_9} B_{\mu_0\mu_1} G_{\mu_2\cdots\mu_5} G_{\mu_6\cdots\mu_9}\Big]\,, \end{aligned} \tag{4.33}$$

where R is the Ricci scalar, and

$$\begin{aligned} F_{\mu\nu} &= \partial_\mu A_\nu - \partial_\nu A_\mu, \\ H_{\mu\nu\rho} &= \partial_\mu B_{\nu\rho} + \text{cyclic permutations of } \mu,\ \nu,\ \rho\,, \\ G_{\mu\nu\rho} &= \partial_\mu C_{\nu\rho\sigma} + A_\mu H_{\nu\rho\sigma} + (-1)^{\mathrm{P}} \cdot \text{cyclic permutations}\,, \end{aligned} \tag{4.34}$$

are the field strengths associated with A_μ, $B_{\mu\nu}$ and $C_{\mu\nu\rho}$ respectively. Upon compactification on K3 we get a new set of scalar fields from the Kahler and complex structure moduli of K3. These can be regarded as deformations of the metric and give a total of 58 real scalar fields. We get 22 more scalar fields $\phi^{(p)}$ by decomposing the antisymmetric tensor field B_{MN} along the twenty two harmonic two forms $\omega_{mn}^{(p)}$ in K3:

$$B_{mn}(x,y) \sim \sum_{p=1}^{22} \phi_p(x)\omega_{mn}^{(p)}(y) + \cdots . \tag{4.35}$$

Here $\{x^\mu\}$ and $\{y^m\}$ denote coordinates along the non-compact and K3 directions respectively. These eighty scalar fields together parametrize a coset $O(4,20)/O(4) \times O(20)$ and can be described by a matrix $M^{(\mathrm{A})}$ satisfying properties identical to those of $M^{(\mathrm{H})}$ described in (4.27). This theory also has twenty four $U(1)$ gauge fields. 22 of the gauge fields arise from the components of the three form field C_{MNP}:

$$C_{mn\mu}(x,y) = \sum_{p=1}^{22} \omega_{mn}^{(p)}(y)\mathcal{A}_\mu^{(p)}(x) + \dots . \tag{4.36}$$

$\mathcal{A}_\mu^{(p)}$ defined in (4.36) behaves as gauge fields in six dimensions. One more gauge field comes from the original RR gauge field A_μ. The last one $\mathcal{A}_\mu$ comes from dualizing $C_{\mu\nu\rho}$:

$$G \sim {}^*(d\mathcal{A}) , \tag{4.37}$$

where * denotes Poincare dual in six dimensions. Together we shall denote these gauge fields by $A_\mu^{(\mathrm{A})a}$ for $1 \le a \le 24$. Besides these fields, the theory contains the canonical metric and the anti-symmetric tensor field which we shall denote by $g_{\mu\nu}^{(\mathrm{A})}$ and $B_{\mu\nu}^{(\mathrm{A})}$ respectively. The action involving these fields is given by,

$$\begin{aligned} S_{\mathrm{A}} &= \frac{1}{(2\pi)^3}\int \mathrm{d}^6x\sqrt{-g^{(\mathrm{A})}}\Big[R^{(\mathrm{A})} - \frac{1}{2}g^{(A)\mu\nu}\partial_\mu\Phi^{(\mathrm{A})}\partial_\nu\Phi^{(\mathrm{A})} \\ &\quad + \frac{1}{8}g^{\mu\nu}\mathrm{Tr}(\partial_\mu M^{(\mathrm{A})}L\partial_\nu M^{(\mathrm{A})}L) \\ &\quad - \frac{1}{4}\mathrm{e}^{\Phi^{(\mathrm{A})}/2}g^{(\mathrm{A})\mu\mu'}g^{(\mathrm{A})\nu\nu'}F_{\mu\nu}^{(\mathrm{A})a}(LM^{(\mathrm{A})}L)_{ab}F_{\mu'\nu'}^{(\mathrm{A})b} \\ &\quad - \frac{1}{12}\mathrm{e}^{-\Phi^{(\mathrm{A})}}g^{(\mathrm{A})\mu\mu'}g^{(\mathrm{A})\nu\nu'}g^{(\mathrm{A})\rho\rho'}H_{\mu\nu\rho}^{(\mathrm{A})}H_{\mu'\nu'\rho'}^{(\mathrm{A})} \\ &\quad - \frac{1}{16}\varepsilon^{\mu\nu\rho\delta\epsilon\eta}(\sqrt{-g^{(\mathrm{A})}})^{-1}B_{\mu\nu}^{(\mathrm{A})}F_{\rho\delta}^{(\mathrm{A})a}L_{ab}F_{\epsilon\eta}^{(\mathrm{A})b}\Big] , \end{aligned} \tag{4.38}$$

where $F_{\mu\nu}^{(\mathrm{A})a}$ is the field strength associated with $A_\mu^{(\mathrm{A})a}$, $R^{(\mathrm{A})}$ is the Ricci scalar, and $H_{\mu\nu\rho}^{(\mathrm{A})}$ is the field strength associated with $B_{\mu\nu}^{(\mathrm{A})}$:

$$H_{\mu\nu\rho}^{(\mathrm{A})} = \partial_\mu B_{\nu\rho}^{(\mathrm{A})} + (\text{cyclic permutations of } \mu,\nu,\rho) . \tag{4.39}$$

In writing down the above action we have used the convention that $M^{(\mathrm{A})} = I_{24}$ corresponds to compactification on a specific reference K3, possibly with specific background B_{mn} fields. This action has an $O(4{,}20)$ symmetry of the form:

$$
\begin{aligned}
&M^{(\mathrm{A})} \to \Omega M^{(\mathrm{A})}\Omega^{\mathrm{T}}, \qquad A^{(\mathrm{A})a}_{\mu} \to \Omega_{ab} A^{(\mathrm{A})b}_{\mu}, \qquad g^{(\mathrm{A})}_{\mu\nu} \to g^{(\mathrm{A})}_{\mu\nu},\\
&B^{(\mathrm{A})}_{\mu\nu} \to B^{(\mathrm{A})}_{\mu\nu}, \qquad \Phi^{(\mathrm{A})} \to \Phi^{(\mathrm{A})}\,,
\end{aligned}
\tag{4.40}
$$

where Ω satisfies:

$$
\Omega L \Omega^{\mathrm{T}} = L\,. \tag{4.41}
$$

An $O(4{,}20;Z)$ subgroup of this can be shown to be an exact T-duality symmetry of string theory [126]. The lattice Λ'_{24} which is preserved by this $O(4{,}20;Z)$ subgroup of $O(4{,}20)$ is not the lattice Λ_{24} defined earlier, but is in general an $O(4{,}20)$ rotation of that lattice:

$$
\Lambda'_{24} = \Omega_0 \Lambda_{24}\,. \tag{4.42}
$$

Ω_0 depends on the choice of the special reference K3 mentioned earlier.

It is now a straightforward exercise to show that the equations of motion and the Bianchi identities derived from (4.29) and (4.38) are identical if we use the following map between the heterotic and the type II variables [23,43]:

$$
\begin{aligned}
&g^{(\mathrm{H})}_{\mu\nu} = g^{(\mathrm{A})}_{\mu\nu}, \qquad M^{(\mathrm{H})} = \widetilde{\Omega} M^{(\mathrm{A})} \widetilde{\Omega}^{\mathrm{T}},\\
&\Phi^{(\mathrm{H})} = -\Phi^{(\mathrm{A})}, \qquad A^{(\mathrm{H})a}_{\mu} = \widetilde{\Omega}_{ab} A^{(\mathrm{A})a}_{\mu},\\
&\sqrt{-g^{(\mathrm{H})}}\exp(-\Phi^{(\mathrm{H})}) H^{(\mathrm{H})\mu\nu\rho} = \frac{1}{6}\varepsilon^{\mu\nu\rho\delta\epsilon\eta} H^{(\mathrm{A})}_{\delta\epsilon\eta}\,.
\end{aligned}
\tag{4.43}
$$

where $\widetilde{\Omega}$ is an arbitrary $O(4{,}20)$ matrix. This leads to the conjectured equivalence between heterotic string theory compactified on T^4 and type IIA string theory compactified on K3 [23]. But clearly the two theories cannot be equivalent for all $\widetilde{\Omega}$ since in the individual theories the $O(4{,}20)$ symmetry is broken down to $O(4{,}20;Z)$. $\widetilde{\Omega}$ can be found (up to an $O(4{,}20;Z)$ transformation) by comparing the T-duality symmetry transformations in the two theories. To do this let us note that according to equation (4.43) a transformation $M^{(\mathrm{H})} \to \Omega M^{(\mathrm{H})}\Omega^{\mathrm{T}}$ will induce a transformation

$$
M^{(\mathrm{A})} \to (\widetilde{\Omega}^{-1}\Omega\widetilde{\Omega}) M^{(\mathrm{A})} (\widetilde{\Omega}^{-1}\Omega\widetilde{\Omega})^{\mathrm{T}}\,. \tag{4.44}
$$

Thus if Ω preserves the lattice Λ_{24}, $\widetilde{\Omega}^{-1}\Omega\widetilde{\Omega}$ should preserve the lattice $\Lambda'_{24} = \Omega_0\Lambda_{24}$. This happens if we choose:

$$
\widetilde{\Omega} = \Omega_0^{-1}\,. \tag{4.45}
$$

Note again that there is a relative minus sign that relates $\Phi^{(\mathrm{H})}$ and $\Phi^{(\mathrm{A})}$, showing that the strong coupling limit of one theory corresponds to the weak coupling limit of the other theory.

4.4 $SL(2,Z)$ self-duality of type IIB in $D = 10$

As described in Section 2.1, the massless bosonic fields in type IIB string theory come from two sectors, – Neveu-Schwarz–Neveu-Schwarz (NS) and Ramond-Ramond (RR). The NS sector gives the graviton described by the metric $g_{\mu\nu}$, an anti-symmetric tensor field $B_{\mu\nu}$, and a scalar field Φ known as the dilaton. The RR sector contributes a scalar field a sometimes called the axion, another rank two anti-symmetric tensor field $B'_{\mu\nu}$, and a rank four anti-symmetric tensor field $D_{\mu\nu\rho\sigma}$ whose field strength is self-dual.

It is often convenient to combine the axion and the dilaton into a complex scalar field λ as follows[12]:

$$\lambda = a + i\mathrm{e}^{-\Phi/2} \equiv \lambda_1 + i\lambda_2\,. \tag{4.46}$$

The low energy effective action in this theory can be determined either from the requirement of supersymmetry, or by explicit computation in string theory. Actually it turns out that there is no simple covariant action for this low energy theory, but there are covariant field equations [44], which are in fact just the equations of motion of type IIB supergravity. Although in string theory this low energy theory is derived from the tree level analysis, non-renormalization theorems tell us that this is exact to this order in the space-time derivatives. Basically supersymmetry determines the form of the equations of motion to this order in the derivatives completely, and so there is no scope for the quantum corrections to change the form of the action.

For the sake of brevity, we shall not explicitly write down the equations of motion. The main point is that these equations of motion are covariant (in the sense that they transform into each other) under an $SL(2,R)$ transformation [44]:

$$\begin{gathered}\lambda \to \frac{p\lambda + q}{r\lambda + s}, \qquad \begin{pmatrix} B_{\mu\nu} \\ B'_{\mu\nu} \end{pmatrix} \to \begin{pmatrix} p & q \\ r & s \end{pmatrix} \begin{pmatrix} B_{\mu\nu} \\ B'_{\mu\nu} \end{pmatrix}, \\ g_{\mu\nu} \to g_{\mu\nu}, \quad D_{\mu\nu\rho\sigma} \to D_{\mu\nu\rho\sigma}\,,\end{gathered} \tag{4.47}$$

where p, q, r, s are real numbers satisfying,

$$ps - qr = 1\,. \tag{4.48}$$

The existence of this $SL(2,R)$ symmetry in the type IIB supergravity theory led to the conjecture that an $SL(2,Z)$ subgroup of this $SL(2,R)$, obtained by restricting p, q, r, s to be integers instead of arbitrary real numbers, is

[12] Note that this field λ has no relation to the field λ defined in Section 4.2 for heterotic string theory on T^6, although both transform as modulus under the respective $SL(2,Z)$ duality transformations in the two theories.

a symmetry of the full string theory [23]. The breaking of $SL(2,R)$ to $SL(2,Z)$ can be seen as follows. An elementary string is known to carry $B_{\mu\nu}$ charge. In suitable normalization convention, it carries exactly one unit of $B_{\mu\nu}$ charge. This means that the $B_{\mu\nu}$ charge must be quantized in integer units, as the spectrum of string theory does not contain fractional strings carrying a fraction of the charge carried by the elementary string. From (4.47) we see that acting on an elementary string state carrying one unit of $B_{\mu\nu}$ charge, the $SL(2,R)$ transformation gives a state with p units of $B_{\mu\nu}$ charge and r units of $B'_{\mu\nu}$ charge. Thus p must be an integer. It is easy to see that the maximal subgroup of $SL(2,R)$ for which p is always an integer consists of matrices of the form

$$\begin{pmatrix} p & \alpha q \\ \alpha^{-1} r & s \end{pmatrix}, \tag{4.49}$$

with p, q, r, s integers satisfying $(ps - qr) = 1$, and α a fixed constant. Absorbing α into a redefinition of $B'_{\mu\nu}$ we see that the subgroup of $SL(2, R)$ matrices consistent with charge quantization are the $SL(2, Z)$ matrices $\begin{pmatrix} p & q \\ r & s \end{pmatrix}$ with p, q, r, s integers satisfying $ps - qr = 1$.

Note that this argument only shows that $SL(2,Z)$ is the maximal possible subgroup of $SL(2,R)$ that [illegible], but does not prove that $SL(2,Z)$ is a symmetry of string theory. In particular, since $SL(2,Z)$ acts non-trivially on the dilaton, whose vacuum expectation value represents the string coupling constant, it cannot be verified order by order in string perturbation theory. We shall see later how one can find non-trivial evidence for this symmetry.

Besides this non-perturbative $SL(2,Z)$ transformation, type IIB theory has two perturbatively verifiable discrete Z_2 symmetries. They are as follows:

- $(-1)^{F_{\rm L}}$: it changes the sign of all the Ramond sector states on the left moving sector of the world-sheet. In particular, acting on the massless bosonic sector fields, it changes the sign of a, $B'_{\mu\nu}$ and $D_{\mu\nu\rho\sigma}$, but leaves $g_{\mu\nu}$, $B_{\mu\nu}$ and Φ invariant;

- Ω: this is the world-sheet parity transformation mentioned in Section 2.1 that exchanges the left- and the right-moving sectors of the world-sheet. Acting on the massless bosonic sector fields, it changes the sign of $B_{\mu\nu}$, a and $D_{\mu\nu\rho\sigma}$, leaving the other fields invariant.

From this description, we see that the effect of $(-1)^{F_{\rm L}} \cdot \Omega$ is to change of sign of $B_{\mu\nu}$ and $B'_{\mu\nu}$, leaving the other massless bosonic fields invariant. Comparing this with the action of the $SL(2,Z)$ transformation laws of the massless bosonic sector fields, we see that $(-1)^{F_L} \cdot \Omega$ can be identified with the $SL(2,Z)$ transformation:

$$\begin{pmatrix} -1 & \\ & -1 \end{pmatrix} . \tag{4.50}$$

This information will be useful to us later.

Theories obtained by modding out (compactified) type IIB string theory by a discrete symmetry group, where some of the elements of the group involve Ω, are known as orientifolds [100, 101]. The simplest example of an orientifold is type IIB string theory modded out by Ω. This corresponds to type I string theory. The closed string sector of type I theory consists of the Ω invariant states of type IIB string theory. The open string states of type I string theory are the analogs of twisted sector states in an orbifold, which must be added to the theory in order to maintain finiteness.

4.5 Other examples

Following the same procedure, namely, studying symmetries of the effective action together with charge quantization rules, we are led to many other duality conjectures in theories with 16 or more supersymmetry generators. Here we shall list the main series of such duality conjectures. We begin with the self duality groups of type II string theories compactified on tori of different dimensions. As mentioned earlier, there is a T-duality that relates type IIA on a circle to type IIB on a circle of inverse radius. Thus for $n \geq 1$, the self-duality groups of type IIA and type IIB theories compactified on an n-dimensional torus T^n will be identical. We now list the conjectured self-duality groups of type IIA/IIB string theory compactified on T^n for different values of n [23]:

$D = (10 - n)$	Full Duality Group	T-duality Group
9	$SL(2, Z)$	$-$
8	$SL(2, Z) \times SL(3, Z)$	$SL(2, Z) \times SL(2, Z)$
7	$SL(5, Z)$	$SO(3, 3; Z)$
6	$SO(5, 5; Z)$	$SO(4, 4; Z)$
5	$E_{6(6)}(Z)$	$SO(5, 5; Z)$
4	$E_{7(7)}(Z)$	$SO(6, 6; Z)$
3	$E_{8(8)}(Z)$	$SO(7, 7; Z)$
2	$\widehat{E_{8(8)}}(Z)$	$SO(8, 8; Z)$

Note that besides the full duality group, we have also displayed the T-duality group of each theory which can be verified order by order in string perturbation theory. $E_{n(n)}$ denotes a non-compact version of the exceptional group E_n for $n = 6, 7, 8$, and $E_{n(n)}(Z)$ denotes a discrete subgroup of $E_{n(n)}$. $\widehat{G}$ for any group G denotes the loop group of G based on the corresponding affine algebra and $\widehat{G}(Z)$ denotes a discrete subgroup of this loop group. Note that we have stopped at $D = 2$. We could in principle continue this all the way to $D = 1$ where all space-like directions are compactified. In this case one expects a very large duality symmetry group based on hyperbolic Lie algebra [103], which is not well understood to this date.

In each of the cases mentioned, the low energy effective field theory is invariant under the full continuous group [45], but charge quantization breaks this symmetry to its discrete subgroup. As noted before, these symmetries were discovered in the early days of supergravity theories, and were known as hidden non-compact symmetries.

Next we turn to the self-duality conjectures involving compactified heterotic string theories. Although there are two distinct heterotic string theories in ten dimensions, upon compactification on a circle, the two heterotic string theories can be shown to be related by a T-duality transformation. As a result, upon compactification on T^n, both of them will have the same self-duality group. We now display this self-duality group in various dimensions:

$D = (10 - n)$	Full Duality Group	T-duality Group
9	$O(1, 17, Z)$	$O(1, 17; Z)$
8	$O(2, 18, Z)$	$O(2, 18; Z)$
7	$O(3, 19, Z)$	$O(3, 19; Z)$
6	$O(4, 20, Z)$	$O(4, 20; Z)$
5	$O(5, 21, Z)$	$O(5, 21; Z)$
4	$O(6, 22, Z) \times SL(2, Z)$	$O(6, 22; Z)$
3	$O(8, 24, Z)$	$O(7, 23; Z)$
2	$O(\widehat{8, 24}, Z)$	$O(8, 24; Z)$

Since type I and $SO(32)$ heterotic string theories are conjectured to be dual to each other in ten dimensions, the second column of the above table also represents the duality symmetry group of type I string theory on T^n. However, in the case of type I string theory, there is no perturbatively realised self-duality group (except trivial transformations which are part of the $SO(32)$ gauge group and the group of global diffeomorphisms of T^n).

The effective action of type IIB string theory compactified on K3 has an $SO(5,21)$ symmetry [43], which leads to the conjecture that an $SO(5,21;Z)$ subgroup of this is an exact self-duality symmetry of the type IIB string theory on K3. The conjectured duality between type IIA string theory compactified on K3 and heterotic string theory compactified on T^4 has already been discussed before. Due to the equivalence of type IIB on S^1 and type IIA on S^1, type IIA on K3 $\times$ T^n is equivalent to type IIB on K3 $\times$ T^n. Finally, due to the conjectured duality between type IIA on K3 and heterotic on T^4, type IIA/IIB on K3 $\times$ T^n are dual to heterotic string theory on T^{n+4} for $n \geq 1$. Thus the self-duality symmetry groups in these theories can be read out from the second column of the previous table displaying the self-duality groups of heterotic string theory on T^n.

Besides the theories discussed here, there are other theories with 16 or more supercharges obtained from non-geometric compactification of heterotic/type II string theories [46–48]. The duality symmetry groups of these theories can again be guessed from an analysis of the low energy effective field theory and the charge quantization conditions. Later we shall also describe a more systematic way of "deriving" various duality conjectures from some basic set of dualities.

Although in this section I have focussed on duality symmetries of the low energy effective action which satisfy a non-renormalization theorem as a consequence of space-time supersymmetry, this is not the only part of the full effective action which satisfy such a non-renormalization theorem. Quite often the effective action contains another set of terms satisfying non-renormalization theorems. They are required for anomaly cancellation, and are known as Green-Schwarz terms. Adler-Bardeen theorem guarantees that they are not renormalized beyond one loop. These terms have also been used effectively for testing various duality conjectures [127], but I shall not discuss it in this article.

5 Precision test of duality: Spectrum of BPS states

Analysis of the low energy effective action, as discussed in the last section, provides us with only a crude test of duality. Its value lies in its simplicity. Indeed, most of the duality conjectures in string theory were arrived at by analysing the symmetries of the low energy effective action.

But once we have arrived at a duality conjecture based on the analysis of the low energy effective action, we can perform a much more precise test by analysing the spectrum of BPS states in the theories. BPS states are states which are invariant under part of the supersymmetry transformation, and are characterized by two important properties:

- They belong to a supermultiplet which has typically less dimension than a non-BPS state. This has an analog in the theory of representations of the Lorentz group, where massless states form a shorter representation of the algebra than massive states. Thus for example a photon has only two polarizations but a massive vector particle has three polarizations;

- The mass of a BPS state is completely determined by its charge as a consequence of the supersymmetry algebra. This relation between the mass and the charge is known as the BPS mass formula. This statement also has an analog in the theory of representations of the Lorentz algebra, [illegible] a spin 1 representation of the Lorentz algebra containing only two states must be necessarily massless.

We shall now explain the origin of these two properties [29]. Suppose the theory has N real supersymmetry generators Q_α $(1 \le \alpha \le N)$. Acting on a single particle state [illegible], the supersymmetry algebra takes the form:

$$\{Q_\alpha, Q_\beta\} = f_{\alpha\beta}(m, \vec{Q}, \{y\})\,, \tag{5.1}$$

where $f_{\alpha\beta}$ is a real symmetric matrix which is a function of its arguments m, $\vec{Q}$ and $\{y\}$. Here m denotes the rest mass of the particle, $\vec{Q}$ denotes various gauge charges carried by the particle, and $\{y\}$ denotes the coordinates labelling the moduli space of the theory[13]. We shall now consider the following distinct cases:

1. $f_{\alpha\beta}$ has no zero eigenvalue. In this case by taking appropriate linear combinations of Q_α we can diagonalize f. By a further appropriate rescaling of Q_α, we can bring f into the identity matrix. Thus in this basis the supersymmetry algebra has the form:

$$\{Q_\alpha, Q_\beta\} = \delta_{\alpha\beta}. \tag{5.2}$$

   This is the N dimensional Clifford algebra. Thus the single particle states under consideration form a representation of this Clifford algebra, which is $2^{N/2}$ dimensional. (We are considering the case where N is even.) Such states would correspond to non-BPS states.

2. f has $(N - M)$ zero eigenvalues for some $M < N$. In this case, by taking linear combinations of the Q_α we can bring the algebra into the form:

$$\begin{aligned} \{Q_\alpha, Q_\beta\} &= \delta_{\alpha\beta}, \quad \text{for} \quad 1 \le \alpha, \beta \le M\,, \\ &= 0 \quad \text{for} \quad \alpha \text{ or } \beta > M\,. \end{aligned} \tag{5.3}$$

   We can form an irreducible representation of this algebra by taking all states to be annihilated by Q_α for $\alpha > M$. In that case the states will

[13] Only specific combinations of $\vec{Q}$ and $\{y\}$, known as central charges, appear in the algebra.

form a representation of an M dimensional Clifford algebra generated by Q_α for $1 \le \alpha \le M$. This representation is $2^{M/2}$ dimensional for M even. Since $M < N$, we see that these are lower dimensional representations compared to that of a generic non-BPS state. Furthermore, these states are invariant under part of the supersymmetry algebra generated by Q_α for $\alpha > M$. These are known as BPS states. We can get different kinds of BPS states depending on the value of M, depending on the number of supersymmetry generators that leave the state invariant.

From this discussion it is clear that in order to get a BPS state, the matrix f must have some zero eigenvalues. This in turn, gives a constraint involving mass m, charges $\vec{Q}$ and the moduli $\{y\}$, and is the origin of the BPS formula relating the mass and the charge of the particle.

Before we proceed, let us illustrate the preceeding discussion in the context of a string theory. Consider type IIB string theory compactified on a circle S^1. The total number of supersymmetry generators in this theory is 32. Thus a generic non-BPS supermultiplet is $2^{16} = (256)^2$ dimensional. These are known as long multiplets. This theory also has BPS states breaking half the space-time supersymmetry. For these states $M = 16$ and hence we have $2^8 = 256$ dimensional representation of the supersymmetry algebra. These states are known as ultra-short multiplets. We can also have BPS states breaking 3/4 of the space-time supersymmetry ($M = 24$). These will form a $2^{12} = 256 \times 16$ dimensional representation, and are known as short multiplets. In each case there is a specific relation between the mass and the various charges carried by the state. We shall discuss this relation as well as the origin of these BPS states in more detail later.

As another example, consider heterotic string theory compactified on an n-dimensional torus T^n. The original theory has 16 supercharges. Thus a generic non-BPS state will belong to a $2^8 = 256$ dimensional representation of the supersymmetry algebra. But if we consider states that are invariant under half of the supercharges, then they belong to a $2^4 = 16$ dimensional representation of the supersymmetry algebra. This is known as the short representation of this superalgebra. We can also have states that break 3/4 of the supersymmetries[14]. These belong to a 64 dimensional representation of the supersymmetry algebra known as intermediate states.

BPS states are further characterized by the property that the degeneracy of BPS states with a given set of charge quantum numbers is independent of the value of the moduli fields $\{y\}$. Since string coupling is also one of

[14]It turns out that these states can exist only for $n \ge 5$. This constraint arises due to the fact that the unbroken supersymmetry generators must form a representation of the little group $SO(9-n)$ of a massive particle in $(10-n)$ dimensional space-time.

the moduli of the theory, this implies that the degeneracy at any value of the string coupling is the same as that at weak coupling. This is the key property of the BPS states that makes them so useful in testing duality, so let us review the argument leading to this property [29]. We shall discuss this in the context of the specific example of type IIB string theory compactified on S^1, but it can be applied to any other theory. Suppose the theory has an ultra-short multiplet at some point in the moduli space. Now let us change the moduli. The question that we shall be asking is: can the ultra-short multiplet become a long (or any other) multiplet as we change the moduli? If we assume that the total number of states does not change discontinuously, then this is clearly not possible since other multiplets have different number of states. Thus as long as the spectrum varies smoothly with the moduli (which we shall assume), an ultra-short multiplet stays ultra-short as we move in the moduli space [88]. Furthermore, as long as it stays ultra-short, its mass is determined by the BPS formula. Thus we see that the degeneracy of ultra-short multiplets cannot change as we change the moduli of the theory. A similar argument can be given for other multiplets as well. Note that for this argument to be strictly valid, we require that the mass of the BPS state should stay away from the continuum, since otherwise the counting of states is not a well defined procedure. This requires that the mass of a BPS state should be strictly less than the total mass of any set of two or more particles carrying the same total charge as the BPS state.

Given this result, we can now adapt the following strategy to carry out tests of various duality conjectures using the spectrum of BPS states in the theory:

1. Identify BPS states in the spectrum of elementary string states. The spectrum of these BPS states can be trusted at all values of the coupling even though it is calculated at weak coupling;

2. Make a conjectured duality transformation. This typically takes a BPS state in the spectrum of elementary string states to another BPS state, but with quantum numbers that are not present in the spectrum of elementary string states. Thus these states must arise as solitons/composite states;

3. Try to explicitly verify the existence of these solitonic states with degeneracy as predicted by duality. This will provide a non-trivial test of the corresponding duality conjecture.

We shall now illustrate this procedure with the help of specific examples. We shall mainly follow [51, 62, 65].

5.1 $SL(2,Z)$ S-duality in heterotic on T^6 and multi-monopole moduli spaces

As discussed in Section 4.2, heterotic string theory compactified on T^6 is conjectured to have an $SL(2,Z)$ duality symmetry. In this subsection we shall see how one can test this conjecture by examining the spectrum of BPS states.

Since the BPS spectrum does not change as we change the moduli, we can analyse the spectrum near some particular point in the moduli space. As discussed in Section 4.2, at a generic point in the moduli space the unbroken gauge group is $U(1)^{28}$. But there are special points in this moduli space where we get enhanced non-abelian gauge group [107]. Thus for example, if we set the internal components of the original ten dimensional gauge fields to zero, we get unbroken $E_8 \times E_8$ or $SO(32)$ gauge symmetry. Let us consider a special point in the moduli space where an $SU(2)$ gauge symmetry is restored. This can be done for example by taking a particular S^1 in T^6 to be orthogional to all other circles, taking the components of the gauge fields along this S^1 to be zero, and taking the radius of this S^1 to be the self-dual radius. In that case the effective field theory at energies much below the string scale will be described by an $N = 4$ supersymmetric $SU(2)$ gauge theory, together with a set of decoupled $N = 4$ supersymmetric $U(1)$ gauge theories and $N = 4$ supergravity. The conjectured $SL(2,Z)$ duality of the heterotic string theory will require the $N = 4$ supersymmetric $SU(2)$ gauge theory to have this $SL(2,Z)$ symmetry[15]. Thus by testing the duality invariance of the spectrum of this $N = 4$ supersymmetric $SU(2)$ gauge theory we can test the conjectured $SL(2,Z)$ symmetry of heterotic string theory.

The $N = 4$ supersymmetric $SU(2)$ gauge theory has a vector, six massless scalars and four massless Majorana fermions in the adjoint representation of $SU(2)$ [50]. The form of the Lagrangian is fixed completely by the requirement of $N = 4$ supersymmetry up to two independent parameters − the coupling constant g that determines the strength of all interactions (gauge, Yukawa, scalar self-interaction etc.), and the vacuum angle θ that multiplies the topological term $\mathrm{Tr}(F\widetilde{F})$ involving the gauge field. With the choice of suitable normalization convention, g and θ are related to the vacuum expectation value of the field λ defined in (4.18) through the relation:

$$\langle\lambda\rangle = \frac{\theta}{2\pi} + i\frac{4\pi}{g^2} \; \cdot \tag{5.4}$$

[15]Independently of string theory, the existence of a strong-weak coupling duality in this theory was conjectured earlier [49, 50].

The potential involving the six adjoint representation scalar fields ϕ_m^α ($1 \leq \alpha \leq 3, 1 \leq m \leq 6$) is proportional to

$$\sum_{m<n} \sum_{\alpha} (\epsilon^{\alpha\beta\gamma} \phi_m^\beta \phi_n^\gamma)^2 \,. \tag{5.5}$$

This vanishes for

$$\phi_m^\alpha = a_m \delta_{\alpha 3} \,. \tag{5.6}$$

Vacuum expectation values of ϕ_m^α of the form (5.6) does not break supersymmetry, but breaks the gauge group $SU(2)$ to $U(1)$. The parameters $\{a_m\}$ correspond to the vacuum expectation values of a subset of the scalar moduli fields M in the full string theory. We shall work in a region in the moduli space where $a_m \neq 0$ for some m, but the scale of breaking of $SU(2)$ is small compared to the string scale ($|a_m| << (\sqrt{\alpha'})^{-1}$ for all m), so that gravity is still decoupled from this gauge theory. The BPS states in the spectrum of elementary particles in this theory are the heavy charged bosons $W^\pm$ and their superpartners. These break half of the 16 space-time supersymmetry generators and hence form a $2^{8/2} = 16$ dimensional representation of the supersymmetry algebra. These states can be found explicitly in the spectrum of elementary string states from the sector containing strings with one unit of winding and one unit of momentum along the special S^1 that is responsible for the enhanced $SU(2)$ gauge symmetry. As we approach the point in the moduli space where this special S^1 has self-dual radius, these states become massless and form part of the $SU(2)$ gauge multiplet.

When $SU(2)$ is broken to $U(1)$ by the vacuum expectation value of $\phi_{\rm m}$, the spectrum of solitons in this theory is characterized by two quantum numbers, the electric charge quantum number $n_{\rm e}$ and the magnetic charge quantum number $n_{\rm m}$, normalized so that $n_{\rm e}$ and $n_{\rm m}$ are both integers. We shall denote such a state by $\begin{pmatrix} n_{\rm e} \\ n_{\rm m} \end{pmatrix}$. In this notation the elementary W^+ boson corresponds to a $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ state. By studying the action of the $SL(2,Z)$ transformation (4.26) on the gauge fields, we can easily work out its action on the charge quantum numbers $\begin{pmatrix} n_{\rm e} \\ n_{\rm m} \end{pmatrix}$ [1]. The answer is

$$\begin{pmatrix} n_{\rm e} \\ n_{\rm m} \end{pmatrix} \to \begin{pmatrix} p & q \\ r & s \end{pmatrix} \begin{pmatrix} n_{\rm e} \\ n_{\rm m} \end{pmatrix} , \tag{5.7}$$

for appropriate choice of sign convention for $n_{\rm e}$ and $n_{\rm m}$. Thus acting on an $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ state it produces a $\begin{pmatrix} p \\ r \end{pmatrix}$ state. From the relation $ps - qr = 1$ satisfied by an $SL(2, Z)$ matrix, we can easily see that p and r are relatively prime. Furthermore for every p and r relatively prime, we can find integers q and s satisfying $ps - qr = 1$. Thus $SL(2, Z)$ duality predicts that [illegible] p [illegible] r [illegible] $\begin{pmatrix} p \\ r \end{pmatrix}$ [51].

We can now directly examine the solitonic sector of the theory to check this prediction. The theory contains classical monopole solutions which break half of the supersymmetries of the original theory. These solutions are non-singular everywhere, and in fact, for a given r, there is a $4r$ parameter non-singular solution with r units of total magnetic charge [52, 104]. These $4r$ parameters correspond to the bosonic collective excitations of this system [105]. In order to study the spectrum of BPS solitons, we need to quantize these collective excitations and look for supersymmetric ground states of the corresponding quantum mechanical system. Each solution also has infinite number of vibrational modes with non-zero frequency, but excitations of these modes are not relevant for finding supersymmetric ground states.

States with $r = 1$ come from one monopole solution. This has four bosonic collective coordinates, three of which correspond to the physical position of the monopole in the three dimensional space, and the fourth one is an angular variable describing the $U(1)$ phase of the monopole. The momenta conjugate to the first three coordinates correspond to the components of the physical momentum of the particle. These can be set to zero by working in the rest frame of the monopole. The fourth coordinate is periodically identified and hence its conjugate momentum is quantized in integer units. This integer p corresponds to the electric charge quantum number n_e. Thus the states obtained by quantizing the bosonic sector of the theory has charge quantum numbers $\begin{pmatrix} p \\ 1 \end{pmatrix}$ for all integer p.

The degeneracy comes from quantizing the fermionic sector. There are eight fermionic zero modes, which describe the result of applying the eight broken supersymmetry generators on the monopole solution. These form an eight dimensional Clifford algebra. Thus the ground state has $2^4 = 16$-fold degeneracy, exactly as predicted by $SL(2, Z)$ [50].

Let us now turn to the analysis of states with $r > 1$ [51]. As has already been said, this system has $4r$ bosonic collective coordinates, which, when the monopoles are far away from each other, correspond to the spatial location and the $U(1)$ phase of each of the r monopoles. The total number of fermionic collective coordinates can be computed from an index theorem and is equal to $8r$ [106]. We can divide this set into the "center of mass" coordinates containing four bosonic and eight fermionic coordinates, and the "relative coordinates" containing $4(r - 1)$ bosonic and $8(r - 1)$

fermionic coordinates. The quantization of the center of mass system gives states carrying charge quantum numbers $\begin{pmatrix} p \\ r \end{pmatrix}$ with 16-fold degeneracy, p being the momentum conjugate to the overall $U(1)$ phase. This shows that the degeneracy is always a multiple of 16, consistent with the fact that a short multiplet is 16-fold degenerate. At this stage p can be any integer, not necessarily prime relative to r. However, since the total wave-function is a product of the wave-function of the center of mass system and the relative system, in order to determine the number of short multiplets for a given value of p, we need to turn to the quantum mechanics of the relative coordinates.

It turns out that the bosonic coordinates in the relative coordinate system describe a non-trivial $4(r-1)$ dimensional manifold, known as the relative moduli space of r monopoles [52, 53, 105]. The quantum mechanics of the bosonic and fermionic relative coordinates can be regarded as that of a supersymmetric particle moving in this moduli space. There are several subtleties with this system. They are listed below:

- First of all, the center of mass and the relative coordinates do not completely decouple, although they decouple locally. The full moduli space has the structure [52]:

$$(R^3 \times S^1 \times \mathcal{M}_{\mathrm{r}})/Z_{\mathrm{r}}\,, \tag{5.8}$$

  where R^3 is parametrized by the center of mass location, S^1 by the overall $U(1)$ phase, and $\mathcal{M}_{\mathrm{r}}$ by the relative coordinates. There is an identification of points in the product space $R^3 \times S^1 \times \mathcal{M}_{\mathrm{r}}$ by a Z_{r} transformation that acts as a shift by $2\pi/r$ on S^1 and as a diffeomorphism on $\mathcal{M}_{\mathrm{r}}$ without any fixed point [52, 53]. Due to this identification, the total wave-function must be invariant under this Z_{r} transformation. Since the part of the wave-function involving the coordinate of S^1 picks up a phase $\exp(2\pi i p/r)$ under this Z_{r}, we see that the wave-function involving the relative coordinates must pick up a phase of $\exp(-2\pi i p/r)$ under this Z_{r} transformation;

- Normally the part of the wave-function involving the relative coordinates will be a function on $\mathcal{M}_{\mathrm{r}}$. But it turns out that the effect of the $8(r-1)$ fermionic degrees of freedom in the quantum mechanical system makes the wave-function a differential form of arbitrary rank on $\mathcal{M}_{\mathrm{r}}$ [54, 55];

- Finally, among all the possible states, the ones saturating Bogomol'nyi bound correspond to harmonic differential forms on $\mathcal{M}_{\mathrm{r}}$. This can be understood as follows. It can be shown that the Hamiltonian of the

> relative coordinates correspond to the Laplacian on $\mathcal{M}$. Also it turns out that the BPS mass formula is saturated by contribution from the center of mass coordinates. Hence in order to get a BPS state, the part of the wave-function involving the relative coordinates must be an eigenstate of the corresponding Hamiltonian with zero eigenvalue [illegible] it must be a harmonic form on $\mathcal{M}_{\mathrm{r}}$. Thus for every harmonic differential form we get a short multiplet, since the fermionic degrees of freedom associated with the center of mass coordinates supply the necessary 16-fold degeneracy.

Thus the existence of a short multiplet of charge quantum numbers $\begin{pmatrix} p \\ r \end{pmatrix}$ would require the existence of a harmonic form on $\mathcal{M}_{\mathrm{r}}$ that picks up a phase of $\exp(2\pi i p/r)$ under the action of Z_{r}. According to the prediction of $SL(2, Z)$ [illegible] p [illegible] r [illegible] p [illegible]

For $r = 2$ the relevant harmonic form can be constructed explicitly [51, 56,57], therby verifying the existence of the states predicted by $SL(2, Z)$ duality. For $r > 2$ the analysis is more complicated since the metric in the multimonopole moduli space is not known. However general arguments showing the existence of the necessary harmonic forms has been given [58, 59].

Besides the BPS states discussed here, the spectrum of elementary string states in the heterotic string theory on T^6 contains many other BPS states. In the world-sheet theory, a generic state is created by applying oscillators from the left- and the right-moving sector on the Fock vacuum. The Fock vacuum, in turn, is characterized by a pair of vectors $(\vec{k}_{\mathrm{L}}, \vec{k}_{\mathrm{R}})$ specifying the charges (momenta) associated with the six right-handed and twenty two left-handed currents on the world-sheet. From the viewpoint of the space-time theory, these 28 components of $(\vec{k}_{\mathrm{L}}, \vec{k}_{\mathrm{R}})$ are just appropriate linear combinations of the charges carried by the state under the 28 $U(1)$ gauge fields. The tree level mass formula for an elementary string state in the NS sector is given by[16],

$$m^2 = \frac{4}{\lambda_2}\left[\frac{\vec{k}_{\mathrm{R}}^2}{2} + N_{\mathrm{R}} - \frac{1}{2}\right] = \frac{4}{\lambda_2}\left[\frac{\vec{k}_{\mathrm{L}}^2}{2} + N_{\mathrm{L}} - 1\right], \tag{5.9}$$

where N_{R} and N_{L} denote respectively the oscillator levels of the state in the right- and the left-moving sectors of the world-sheet. In the above equation the terms in the square bracket denote the total contribution to L_0 and $\bar{L}_0$ from the oscillators, the internal momenta, and the vacua in the right- and

[16]In this and all subsequent mass formula λ_2 should really be interpreted as the vacuum expectation value of λ_2.

the left-moving sectors respectively. Normally we do not have the factor of λ_2^{-1} in the mass formula since the formula refers to the ADM mass measured in the string metric $G_{\mu\nu} = \lambda_2^{-1} g_{\mu\nu}$. But here (and in the rest of the article) we quote the ADM mass measured in the canonical metric $g_{\mu\nu}$. This is more convenient for discussing duality invariance of the spectrum, since it is $g_{\mu\nu}$ and not $G_{\mu\nu}$ that remains invariant under a duality transformation. The additive factor of $-1/2$ and -1 can be interpreted as the contributions to L_0 and $\bar{L}_0$ from the vacuum. (In the covariant formulation these can be traced to the contributions from the world-sheet ghost fields).

It turns out that of the full set of elementary string states, only those states which satisfy the constraint [64]

$$N_{\rm R} = \frac{1}{2}\,, \tag{5.10}$$

correspond to BPS states (short multiplets). From equations (5.9) we see that for these states

$$N_{\rm L} = \frac{1}{2}(\vec{k}_{\rm R}^2 - \vec{k}_{\rm L}^2) + 1\,. \tag{5.11}$$

The degeneracy $\mathrm{d}(N_{\rm L})$ of short multiplets for a given set of $\vec{k}_{\rm L}$, $\vec{k}_{\rm R}$ is determined by the number of ways a level $N_{\rm L}$ state can be created out of the Fock vacuum by the 24 left-moving bosonic oscillators (in the light-cone gauge) $-$ 8 from the transverse bosonic coordinates of the string and 16 from the bosonization of the 32 left-moving fermions on the world-sheet $-$ and is given by the formula:

$$\sum_{N_{\rm L}=0}^{\infty} \mathrm{d}(N_{\rm L}) q^{N_{\rm L}} = \prod_{n=1}^{\infty} \frac{1}{(1-q^n)^{24}}\ \cdot \tag{5.12}$$

The BPS states discussed earlier $-$ the ones which can be regarded as the massive gauge bosons of a spontaneously broken non-abelian gauge theory $-$ correspond to the $N_{\rm L} = 0$ states in this classification. From equation (5.12) we see that we have only one short multiplet for states with this quantum number; this is consistent with their description as heavy gauge bosons in an $N = 4$ supersymmetric gauge theory. The next interesting class of states are the ones with $N_{\rm L} = 1$. From (5.12) we see that they have degeneracy 24[17]. An $SL(2, Z)$ transformation relates these states to appropriate magnetically charged states with r units of magnetic charge and p units of electric charge for p and r relatively prime. Thus the $SL(2, Z)$ self-duality symmetry of the

[17] In counting degeneracy we are only counting the number of short multiplets, and ignoring the trivial factor of 16 that represents the degeneracy within each short multiplet.

heterotic string theory predicts the existence of 24-fold degenerate solitonic states with these charge quantum numbers.

Verifying the existence of these solitonic states turns out to be quite difficult [60]. The main problem is that unlike the $N_{\rm L} = 0$ states, the solitonic states (known as H-monopoles) which are related to the $N_{\rm L} = 1$ states by $SL(2, Z)$ duality turn out to be singular objects, and hence we cannot unambiguously determine the dynamics of collective coordinates of these solitons just from the low energy effective field theory. Nevertheless, the problem has now been solved for $r = 1$ [70, 72–74], and one finds that these solitons have exactly the correct degeneracy 24.

Similar analysis based on soliton solutions of low energy supergravity theory has been used to test many other duality conjectures [20, 21, 23, 25, 26, 61]. One of the main problems with this approach has been that unlike the example discussed in this section, most of these other solutions are either singular, or has strong curvature at the core where the low energy approximation breaks down. As a result, analysis based on these solutions has been of limited use. The situation changed after the advent of D-branes, to which we now turn.

5.2 $SL(2, Z)$ duality in type IIB on S^1 and D-branes

As discussed earlier, type IIB string theory in ten dimensions has a conjectured $SL(2, Z)$ duality symmetry group. In this section I shall discuss the consequence of this conjectured symmetry for the spectrum of BPS states in type IIB string theory compactified on a circle S^1. For details, see [62, 63].

The spectrum of elementary string states in this theory are characterized by two charges $k_{\rm L}$ and $k_{\rm R}$ defined as:

$$k_{\rm L} = (k\lambda_2^{1/4}/R - wR/\lambda_2^{1/4})/\sqrt{2}, \qquad k_{\rm R} = (k\lambda_2^{1/4}/R + wR/\lambda_2^{1/4})/\sqrt{2}\,, \tag{5.13}$$

where R denotes the radius of S^1 measured in the ten dimensional canonical metric, k/R denotes the momentum along S^1 with k being an integer, and w, also an integer, denotes the number of times the elementary string is wound along S^1. As usual we have set $\alpha' = 1$. In the world-sheet theory describing first quantized string theory, $k_{\rm L}$ and $k_{\rm R}$ denote the left and the right-moving momenta respectively. There are infinite tower of states with this quantum number, obtained by applied appropriate oscillators, both from the left- and the right-moving sector of the world-sheet, on the Fock vacuum of the world-sheet theory carrying these quantum numbers. The mass formula for any state in this tower, measured in the ten dimensional canonical metric, is given by:

$$m^2 = \frac{2}{\sqrt{\lambda_2}}(k_{\rm L}^2 + 2N_{\rm L}) = \frac{2}{\sqrt{\lambda_2}}(k_{\rm R}^2 + 2N_{\rm R})\,, \tag{5.14}$$

where $N_{\rm L}, N_{\rm R}$ denote oscillator levels on the left- and the right- moving sector of the world-sheet respectively[18]. In normal convention, one does not have the factors of (λ_2) in the mass formula, but here it comes due to the fact that we are using the [illegible] canonical metric instead of the string metric to define the mass of a state. (Note that if we had used the nine dimensional canonical metric as defined in equations (4.1), (4.2), there will be an additional multiplicative factor of $R^{-2/9}$ in the expression for m^2.)

Most of these states are not BPS states as they are not invariant under any part of the supersymmetry transformation. It turns out that in order to be invariant under half of the space-time supersymmetry coming from the left- (right-) moving sector of the world-sheet, $N_{\rm L}$ ($N_{\rm R}$) must vanish [64]. Thus a state with $N_{\rm L} = N_{\rm R} = 0$ will preserve half of the total number of supersymmetries and will correspond to ultra-short multiplets. From equation (5.14) we see that mass formula for these states takes the form:

$$m^2 = \frac{2k_{\rm L}^2}{\sqrt{\lambda_2}} = \frac{2k_{\rm R}^2}{\sqrt{\lambda_2}}\ \cdot \tag{5.15}$$

This is the BPS mass formula for these ultra-short multiplets. This requires $k_{\rm L} = \pm k_{\rm R}$ or, equivalently, $k = 0$ or $w = 0$. On the other hand, a state with either $N_{\rm L} = 0$ or $N_{\rm R} = 0$ will break (3/4)th of the total number of supersymmetries in the theory, and will correspond to short multiplets. If, for definiteness, we consider states with $N_{\rm R} = 0$, then the BPS mass formula takes the form:

$$m^2 = \frac{2k_{\rm R}^2}{\sqrt{\lambda_2}}\ \cdot \tag{5.16}$$

$N_{\rm L}$ is determined in terms of $k_{\rm L}$ and $k_{\rm R}$ through the relation:

$$N_{\rm L} = \frac{1}{2}(k_{\rm R}^2 - k_{\rm L}^2) = wk\,. \tag{5.17}$$

There is no further constraint on w and k. Although we have derived these mass formulae by directly analysing the spectrum of elementary string states, they can also be derived by analyzing the supersymmetry algebra, as indicated earlier.

[18]We have stated the formula in the RR sector, but due to space-time supersymmetry we get identical spectrum from the NS and the R sectors.

One can easily calculate the degeneracy of these states by analyzing the spectrum of elementary string states in detail. For example, for the states with $N_{\mathrm{L}} = N_{\mathrm{R}} = 0$, there is a 16-fold degeneracy of states in each (left- and right-) sector of the world-sheet, − 8 from the NS sector and 8 from the R sector. Thus the net degeneracy of such a state is $16\times 16 = 256$, showing that there is a unique ultra-short multiplet carrying given charges $(k_{\mathrm{L}}, k_{\mathrm{R}})$. The degeneracy of short multiplets can be found in a similar manner. Consider for example states with $N_{\mathrm{R}} = 0$, $N_{\mathrm{L}} = 1$. In this case there is a 16-fold degeneracy coming from the right-moving sector of the world-sheet. There is an 8-fold degeneracy from the Ramond sector Fock vacuum of the left-moving sector. There is also an extra degeneracy factor in the left-moving Ramond sector due to the fact that there are many oscillators that can act on the Fock vacuum of the world-sheet theory to give a state at oscillator level $N_{\mathrm{L}} = 1$. For example we get eight states by acting with the transverse bosonic oscillators α^i_{-1} $(1 \le i \le 8)$, and eight states by acting with the transverse fermionic oscillators ψ^i_{-1}[19]. This gives total degeneracy factor of 8×16 in the left-moving Ramond sector. Due to supersymmetry, we get an identical factor from the left-moving NS sector as well. Thus we get a state with total degeneracy $16 \times 16 \times 16$, − 16 from the right moving sector, and 16×16 from the left-moving sector − which is the correct degeneracy of a single short multiplet. Similar counting can be done for higher values of N_{L} as well. It turns out that the total number of short multiplets $\mathrm{d}(N_{\mathrm{L}})$ with $N_{\mathrm{R}} = 0$ for some given value of $N_{\mathrm{L}} \ge 1$ is given by the formula:

$$\sum_{N_{\mathrm{L}}} \mathrm{d}(N_L) q^{N_{\mathrm{L}}} = \frac{1}{16} \prod_{n=1}^{\infty} \left(\frac{1+q^n}{1-q^n} \right)^8 \cdot \qquad (5.18)$$

The $(1+q^n)^8$ and $(1-q^n)^8$ factors in the numerator and the denominator are related respectively to the fact that in the light-cone gauge there are 8 left-moving fermionic fields and 8 left-moving bosonic fields on the world-sheet. The overall factor of (1/16) is due to the fact that the lowest level state is only 256-fold degenerate but a single short multiplet requires 16×256 states.

Let us first consider the ultra-short multiplet with $k = 0$, $w = 1$. These states have mass

$$m^2 = \frac{R^2}{\lambda_2} \cdot \qquad (5.19)$$

[19]Since ψ^i_{-1} has fermion number one, it has to act on the Fock vacua with odd fermion number in order that the states obtained after acting with ψ^i_{-1} on the vacua satisfy GSO projection.

It is well known that an elementary string acts as a source of the $B_{\mu\nu}$ field (see , Ref. [64]). Thus in the $(8+1)$ dimensional theory obtained by compactifying type IIB on S^1, the $w = 1$ state will carry one unit of $B_{9\mu}$ gauge field charge. Now, under $SL(2,Z)$

$$\begin{pmatrix} B_{9\mu} \\ B'_{9\mu} \end{pmatrix} \to \begin{pmatrix} p & q \\ r & s \end{pmatrix} \begin{pmatrix} B_{9\mu} \\ B'_{9\mu} \end{pmatrix} . \tag{5.20}$$

This converts the $w = 1$ state, which we shall denote by $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ reflecting the $\begin{pmatrix} B_{9\mu} \\ B'_{9\mu} \end{pmatrix}$ charge carried by the state, to a $\begin{pmatrix} p \\ r \end{pmatrix}$ state, a state carrying p units of $B_{9\mu}$ charge and r units of $B'_{9\mu}$ charge. The condition $ps - qr = 1$ implies that the pair of integers (p, r) are relatively prime. Thus $SL(2,Z)$ duality of type IIB string theory predicts that $\forall(p,r)$ relatively prime, the theory must have a unique ultra-short multiplet with p units of $B_{9\mu}$ charge and r units of $B'_{9\mu}$ charge [61]. The BPS mass formula for these states can be derived by analysing the supersymmetry algebra, as indicated earlier, and is given by,

$$m^2 = \frac{R^2}{\lambda_2} |r\lambda - p|^2 . \tag{5.21}$$

Note that this formula is invariant under the $SL(2,Z)$ transformation:

$$\lambda \to \frac{a\lambda + b}{c\lambda + d}, \qquad \begin{pmatrix} p \\ r \end{pmatrix} \to \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} p \\ r \end{pmatrix} , \tag{5.22}$$

where $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is an $SL(2,Z)$ matrix.

A similar prediction for the spectrum of BPS states can be made for short multiplets as well. In this case the state is characterized by three integers p, r and k reflecting the $B_{9\mu}$, $B'_{9\mu}$ and $G_{9\mu}$ charge (momentum along S^1) respectively. Let us denote by $\mathrm{d}(k,p,r)$ the degeneracy of such short multiplets. For (p,r) relatively prime, an $SL(2,Z)$ transformation relates these to elementary string states with one unit of winding and k units of momentum along S^1. Such states have degeneracy $\mathrm{d}(k)$ given in equation (5.18). Then by following the same logic as before, we see that the $SL(2,Z)$ duality predicts that for (p,r) relatively prime, $\mathrm{d}(k,p,r)$ is independent of p and r and depends on k according to the relation:

$$\sum_k \mathrm{d}(k,p,r) q^k = \frac{1}{16} \prod_{n=1}^{\infty} \left(\frac{1+q^n}{1-q^n}\right)^8 \cdot \tag{5.23}$$

In other words, there should be a Hagedorn spectrum of short multiplets with charge $\begin{pmatrix} p \\ r \end{pmatrix}$.

A test of $SL(2, Z)$ symmetry involves explicitly verifying the existence of these states. To see what such a test involves, recall that $B'_{\mu\nu}$ arises in the RR sector of string theory. In type II theory, all elementary string states are neutral under RR gauge fields as can be seen by computing a three point function involving any two elementary string states and an RR sector gauge field. Thus a state carrying $B'_{9\mu}$ charge must arise as a soliton. The naive approach will involve constructing such a soliton solution as a solution to the low energy supergravity equations of motion, quantizing its zero modes, and seeing if we recover the correct spectrum of BPS states. However, in actual practice, when one constructs the solution carrying $B'_{\mu\nu}$ charge, it turns out to be singular. Due to this fact it is difficult to proceed further along this line, as identifying the zero modes of a singular solution is not a well defined procedure. In particular we need to determine what boundary condition the modes must satisfy at the singularity. Fortunately, in this theory, there is a novel way of constructing a soliton solution that avoids this problem. This construction uses Dirichlet (D-) branes [65,66]. In order to compute the degeneracy of these solitonic states, we must understand the definition and some of the the properties of these D-branes. This is the subject to which we now turn.

Normally type IIA/IIB string theory contains closed string states only. But we can postulate existence of solitonic extended objects in these theories such that in the presence of these solitons, there can be open string states whose ends lie on these extended objects (see Fig. 10). This can in fact be taken to be the defining relation for these solitons, with the open string states with ends lying on the soliton corresponding to the (infinite number of) vibrational modes of the soliton. Of course, one needs to ensure that the soliton defined this way satisfy all the properties expected of a soliton solution in this theory , partially unbroken supersymmetry, existence of static multi-soliton solutions etc. Since open strings satisfy Dirichlet boundary condition in directions transverse to these solitons, these solitons are called D-branes. In particular, we shall call a D-brane with Neumann boundary condition in $(p+1)$ directions (including time) and Dirichlet boundary condition in $(9-p)$ directions a Dirichlet p-brane, since it can be regarded as a soliton extending along p space-like directions in which we have put Neumann boundary condition. (Thus a 0-brane represents a particle like object, a 1-brane a string like object, and a 2-brane a membrane like object.) To be more explicit, let us consider the following boundary condition on the open string:

$$\begin{aligned} X^m(\sigma=0,\pi) &= x_0^m \quad \text{for} \quad (p+1)\le m\le 9\,, \\ \partial_\sigma X^\mu(\sigma=0,\pi) &= 0 \quad \text{for} \quad 0\le \mu\le p\,, \end{aligned} \tag{5.24}$$

where σ denotes the spatial direction on the string world-sheet. The boundary conditions on the world-sheet fermion fields are determined from (5.24) using various consistency requirements including world-sheet supersymmetry that relates the world-sheet bosons and fermions. Note that these boundary conditions break translational invariance along x^m. Since we want the full theory to be translationally invariant, the only possible interpretation of such a boundary condition is that there is a p dimensional extended object situated at $x^m = x_0^m$ that is responsible for breaking this translational invariance. We call this a Dirichlet p-brane located at $x^m = x_0^m$ $(p+1\le m\le 9)$, and extended along $x^1,\ldots x^p$.

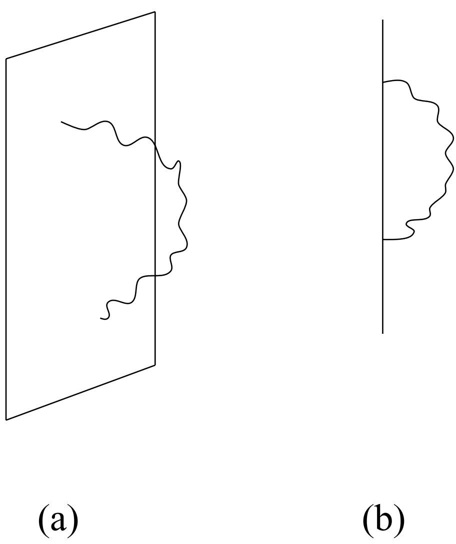


Fig. 10. Open string states with ends attached to a **a)** Dirichlet membrane, **b)** Dirichlet string.

Let us now summarize some of the important properties of D-branes that will be relevant for understanding the test of $SL(2,Z)$ duality in type IIB string theory:

- The Dirichlet p-brane in IIB is invariant under half of the space-time supersymmetry transformations for odd p. To see how this property arises, let us denote by $\epsilon_{\rm L}$ and $\epsilon_{\rm R}$ the space-time supersymmetry transformation parameters in type IIB string theory, originating in the left- and the right-moving sector of the world-sheet theory respectively. $\epsilon_{\rm L}$ and $\epsilon_{\rm R}$ satisfy the chirality constraint:

$$\Gamma^0 \cdots \Gamma^9 \epsilon_{\rm L} = \epsilon_{\rm L}, \qquad \Gamma^0 \cdots \Gamma^9 \epsilon_{\rm R} = \epsilon_{\rm R}\,, \tag{5.25}$$

where Γ^μ are the ten dimensional gamma matrices. The open string boundary conditions (5.24) together with the corresponding boundary conditions on the world-sheet fermions give further restriction on $\epsilon_{\rm L}$ and $\epsilon_{\rm R}$ of the form [65]:

$$\epsilon_{\rm L} = \Gamma^{p+1} \ldots \Gamma^9 \epsilon_{\rm R}\,. \tag{5.26}$$

It is easy to see that the two equations (5.25) and (5.26) are compatible only for odd p. Thus in type IIB string theory Dirichlet p-branes are invariant under half of the space-time supersymmetry transformations for odd p. An identical argument shows that in type IIA string theory we have supersymmetric Dirichlet p-branes only for even p since in this theory equation (5.25) is replaced by,

$$\Gamma^0 \cdots \Gamma^9 \epsilon_{\rm L} = \epsilon_{\rm L}, \qquad \Gamma^0 \cdots \Gamma^9 \epsilon_{\rm R} = -\epsilon_{\rm R}\,. \tag{5.27}$$

- Type IIB (IIA) string theory contains a p-form gauge field for even (odd) p. For example, in type IIB string theory these p-form gauge fields correspond to the scalar a, the rank two anti-symmetric tensor field $B'_{\mu\nu}$ and the rank four anti-symmetric tensor field $D_{\mu\nu\rho\sigma}$. It can be shown that a Dirichlet p-brane carries one unit of charge under the RR $(p+1)$-form gauge field [65]. More precisely, if we denote by $C_{\mu_1\cdots\mu_q}$ the q-form gauge potential, then a Dirichlet p-brane extending along $1\cdots p$ direction acts as a source of $C_{01\cdots p}$. (For $p=5$ and 7 these correspond to magnetic dual potentials of $B'_{\mu\nu}$ and a respectively.) This result can be obtained by computing the one point function of the vertex operator for the field C in the presence of a D-brane. The relevant string world-sheet diagram has been indicated in Figure 11. We shall not discuss the details of this computation here.

From this discussion it follows that a Dirichlet 1-brane (D-string) in type IIB theory carries one unit of charge under the RR 2-form field $B'_{\mu\nu}$. This means that in type IIB on S^1 (labelled by the coordinate x^9) a D-string wrapped around the S^1 describes a particle charged under $B'_{9\mu}$. This then is a candidate soliton carrying charge quantum numbers $\begin{pmatrix}0\\1\end{pmatrix}$ that is related to the $\begin{pmatrix}1\\0\end{pmatrix}$ state.. $SL(2,Z)$ duality. As we had seen earlier, $SL(2,Z)$ duality

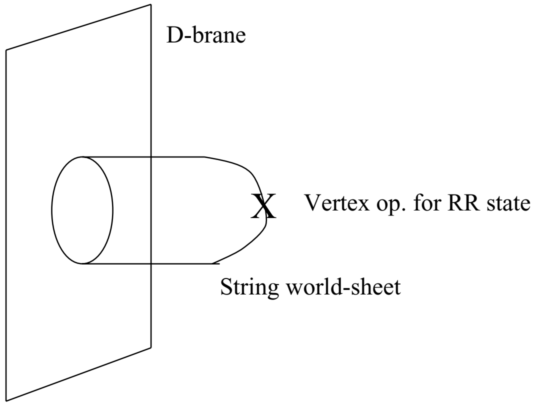


Fig. 11. The string world-sheet diagram relevant for computing the coupling of the RR gauge field to the D-brane. It corresponds to a surface of the topology of a hemisphere with its boundary glued to the D-brane. The vertex operator of the RR-field is inserted at a point on the hemisphere.

predicts that there should be a unique ultra-short multiplet carrying charge quantum numbers $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$. Thus our task now is as follows:

- Quantize the collective coordinates of this soliton;
- Verify if we get an ultra-short multiplet in this quantum theory.

Since the D-string is a one dimensional object, the dynamics of its collective coordinates should be described by a $(1+1)$ dimensional field theory. As we had discussed earlier, all the vibrational modes of the D-string are given by the open string states with ends attached to the D-string. In particular, the zero frequency modes (collective modes) of the D-string that are relevant for analyzing the spectrum of BPS states correspond to [illegible] open string states propagating on the D-string. By analyzing the spectrum of these open string states one finds that the collective coordinates in this case correspond to

- 8 bosonic fields y^m denoting the location of this string in eight transverse directions;
- A $U(1)$ gauge field;
- 8 Majorana fermions.

It can be shown that the dynamics of these collective coordinates is described by a $(1+1)$ dimensional supersymmetric quantum field theory which is the dimensional reduction of the $N = 1$ supersymmetric $U(1)$ gauge theory from $(9+1)$ to $(1+1)$ dimensions. Normally in $(1+1)$ dimension gauge fields have no dynamics. But here since the space direction is compact, $y \equiv \oint A_1 \mathrm{d}l$ is a physical variable. Furthermore, the compactness of $U(1)$ makes y to be periodically identified $((y \equiv y + a)$ for some $a)$. Thus the momentum p_y conjugate to y is quantized ($p_y = 2\pi k/a$ with k integer.) It can be shown that [62] this momentum, which represents electric flux along the D-string, is actually a source of $B_{9\mu}$ charge! Thus if we restrict to the $p_y = 0$ sector then these states carry $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ charge quantum numbers as discussed earlier, but by taking $p_y = 2\pi k/a$, we can get states carrying charge quantum numbers $\begin{pmatrix} k \\ 1 \end{pmatrix}$ as well.

Due to the compactness of the space direction, we can actually regard this as a quantum mechanical system instead of a $(1+1)$ dimensional quantum field theory. It turns out that in looking for ultra-short multiplets, we can ignore all modes carrying momentum along S^1. This corresponds to dimensionally reducing the theory to $(0 + 1)$ dimensions. The degrees of freedom of this quantum mechanical system are:

- 8 bosonic coordinates y^m;
- 1 compact bosonic coordinate y;
- 16 fermionic coordinates.

A quantum state is labelled by the momenta conjugate to y^m (ordinary momenta) and an integer labelling momentum conjugate to y which can be identified with the quantum number p labelling $B_{9\mu}$ charge. The fermionic coordinates satisfy the sixteen dimensional Clifford algebra. Thus quantization of the fermionic coordinates gives $2^8 = 256$ -fold degeneracy, which is precisely the correct degeneracy for a ultra-short multiplet. This establishes the existence of all the required states of charge $\begin{pmatrix} p \\ 1 \end{pmatrix}$ predicted by $SL(2, Z)$ symmetry.

What about $\begin{pmatrix} p \\ r \end{pmatrix}$ states with $r > 1$? These carry r units of $B'_{9\mu}$ charge and hence must arise as a bound state of r D-strings wrapped along S^1. Thus the first question we need to ask is: what is the $(1 + 1)$ dimensional quantum field theory governing the dynamics of this system? In order to answer this question we need to study the dynamics of r D-strings. This system can be described as easily as a single D-string: instead of allowing open strings to end on a single D-string, we allow it to end on any of the r

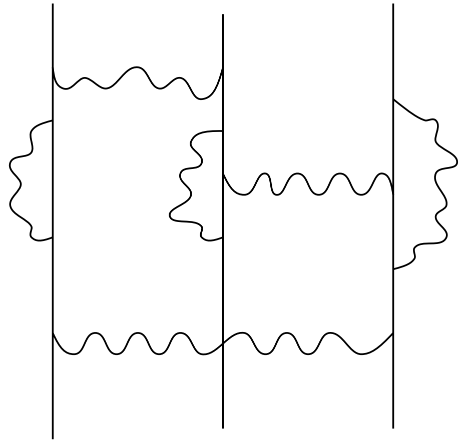


Fig. 12. Possible open string states in the presence of three parallel D-strings.

D-strings situated at

$$x^m = x^m_{(i)}\,, \qquad 2 \le m \le 9, \qquad 1 \le i \le r\,, \tag{5.28}$$

where $\vec{x}_{(i)}$ denotes the location of the i-th D-string. The situation is illustrated in Figure 12. Thus the dynamics of this system will now be described not only by the open strings starting and ending on the same D-string, but also by open strings whose two ends lie on two different D-strings.

For studying the spectrum of BPS states we need to focus our attention on the massless open string states. First of all, for each of the r D-strings we get a $U(1)$ gauge field, eight scalar fields and eight Majorana fermions from open strings with both ends lying on that D-string. But we can get extra massless states from open strings whose two ends lie on two different D-strings when these two D-strings coincide. It turns out that for r coincident D-strings the dynamics of massless strings on the D-string world-sheet is given by the dimensional reduction to $(1+1)$ dimension of $N=1$ supersymmetric $U(r)$ gauge theory in ten dimensions, or equivalently, $N=4$ supersymmetric $U(r)$ gauge theory in four dimensions [62]. Following a logic similar to that in the case of a single D-string, one can show that the problem of computing the degeneracy of $\begin{pmatrix} p \\ r \end{pmatrix}$ states reduces to the computation of certain Witten index in this quantum theory. We shall not go through the details of this analysis, but just state the final result. It turns out that [illegible] (p, r) [illegible] $SL(2, Z)$ [62]!

A similar analysis can be carried out for the short multiplets that carry momentum k along S^1 besides carrying the B and B' charges p and r [62, 63]. In order to get these states from the D-brane spectrum, we can no longer dimensionally reduce the $(1+1)$ dimensional theory to $(0+1)$ dimensions. Instead we need to take into account the modes of the various fields of the $(1+1)$ dimensional field theory carrying momentum along the internal S^1. The BPS states come from configurations where only the left- (or right-) moving modes on S^1 are excited. The calculation of the degeneracy $\mathrm{d}(k,p,r)$ of BPS states carrying given charge quantum numbers (p,r,k) is done by determining in how many ways the total momentum k can be divided among the various left-moving bosonic and fermionic modes. This counting problem turns out to be identical to the one used to get the Hagedorn spectrum of BPS states in the elementary string spectrum, except that the elementary string is replaced here by the solitonic D-string. Naturally, we get back the Hagedorn spectrum for $\mathrm{d}(k,p,r)$ as well. Thus the answer agrees exactly with that predicted by $SL(2,Z)$ duality. This provides us with a test of the conjectured $SL(2,Z)$ symmetry of type IIB on S^1.

The method of using D-branes to derive the dynamics of collective coordinates has been used to verify the predictions of other duality conjectures involving various string compactifications. Among them are self-duality of type II string theory on T^4 [67–70], the duality between heterotic on T^4 and type IIA on K3 [71], the duality between type I and $SO(32)$ heterotic string theory [22], etc.

5.3 Massless solitons and tensionless strings

An interesting aspect of the conjectured duality between the heterotic string theory on T^4 and type IIA string theory on K3 is that at special points in the moduli space the heterotic string theory has enhanced non-abelian gauge symmetry , $E_8 \times E_8$ or $SO(32)$ in the absence of vacuum expectation value of the internal components of the gauge fields, $SU(2)$ at the self-dual radius etc. Perturbative type IIA string theory on K3 does not have any such gauge symmetry enhancement, since the spectrum of elementary string states does not contain any state charged under the $U(1)$ gauge fields arising in the RR sector. Thus, for example, we do not have the $W^{\pm}$ bosons that are required for enhancing a $U(1)$ gauge group to $SU(2)$. At first sight this seems to lead to a contradiction. However upon closer examination one realises that this cannot really be a problem [116]. To see this let us consider a point in the moduli space of heterotic string theory on T^4 where the non-abelian gauge symmetry is broken. At this point the would be massless gauge bosons of the non-abelian gauge theory acquire mass by Higgs mechanism, and appear as BPS states in the abelian theory.

As we approach the point of enhanced gauge symmetry, the masses of these states vanish. Since the masses of BPS states are determined by the BPS formula, the vanishing of the masses must be a consequence of the BPS formula. Thus if we are able to find the images of these BPS states on the type IIA side as appropriate D-brane states, then the masses of these D-brane states must also vanish as we approach the point in the moduli space where the heterotic theory has enhanced gauge symmetry. These massless D-brane solitons will then provide the states necessary for enhancing the gauge symmetry.

To see this more explicitly, let us examine the BPS formula. It can be shown that in the variables defined in Section 4.3 the BPS formula is given by,

$$m^2 = \mathrm{e}^{-\Phi^{(A)}/2}\alpha^{\mathrm{T}}(LM^{(\mathrm{A})}L + L)\alpha\,, \tag{5.29}$$

where α is a 24 dimensional vector belonging to the lattice Λ'_{24}, and represents the $U(1)$ charges carried by this particular state. For each $\vec{\alpha}$ we can assign an occupation number $n(\vec{\alpha})$ which gives the number of BPS multiplets carrying this specific set of charges. Since $M^{(\mathrm{A})}$ is a symmetric $O(4,20)$ matrix, we can express this as $\Omega^{(\mathrm{A})T}\Omega^{(\mathrm{A})}$ for some $O(4,20)$ matrix $\Omega^{(\mathrm{A})}$, and rewrite equation (5.29) as

$$m^2 = \mathrm{e}^{-\Phi^{(\mathrm{A})}/2}\alpha^{\mathrm{T}}L\Omega^{(\mathrm{A})\mathrm{T}}(I_{24} + L)\Omega^{(\mathrm{A})}L\alpha\,. \tag{5.30}$$

As can be seen from equation (4.28), $(I_{24} + L)$ has 20 zero eigenvalues. As we vary $M^{(\mathrm{A})}$ and hence $\Omega^{(\mathrm{A})}$, the vector $\Omega^{(A)}L\alpha$ rotates in the twenty four dimensional space. If for some $\Omega^{(\mathrm{A})}$ it is aligned along one of the eigenvectors of $(I_{24}+L)$ with zero eigenvalue, we shall get massless solitons provided the occupation number $n(\vec{\alpha})$ for this specific $\vec{\alpha}$ is non-zero.

Although this argument resolves the problem at an abstract level, one would like to understand this mechanism directly by analysing the type IIA string theory, since, after all, we do not encounter massless solitons very often in physics. This has been possible through the work of [19, 71, 117]. For simplicity let us focus on the case of enhanced $SU(2)$ gauge symmetry. First of all, one finds that at a generic point in the moduli space where $SU(2)$ is broken, the images of the $W^{\pm}$ bosons in the type IIA theory are given by a D-2 brane wrapped around a certain 2-cycle (topologically non-trivial two dimensional surface) inside K3, the + and the − sign of the charge being obtained from two different orientations of the D-2 brane. Since the two tangential directions on the D-2 brane are directed along the two internal directions of K3 tangential to the 2-cycle, this object has no extension in any of the five non-compact spatial directions, and hence behaves like a particle[20]. It turns out that as we approach the point in the moduli space where the theory on the heterotic side develops enhanced $SU(2)$ gauge symmetry, the K3 on which type IIA theory is compactified becomes singular. At this singularity the area of the topologicaly non-trivial 2-cycle mentioned above goes to zero. As a result, the mass of the wrapped D-2 brane, obtained by multiplying the tension of the D-2 brane by the area of the two cycle, vanishes. This gives us the massless solitons that are required for the gauge symmetry enhancement. A similar mechanism works for getting other gauge groups as well. In fact it turns out that there is a one to one correspondence between the enhanced gauge groups, which are classified by A-D-E dynkin diagram, and the singularity type of K3, which are also classified by the A-D-E dynkin diagram [19]. This establishes an explicit physical relationship between A-D-E singularities and A-D-E lie algebras.

The appearance of enhanced gauge symmetry in type IIA on K3 poses another puzzle. Let us compactify this theory on one more circle. Since such a compactification does not destroy gauge symmetry, this theory also has enhanced gauge symmetry when the K3 becomes singular. But type IIA on K3 $\times S^1$ is T-dual to type IIB on K3 $\times S^1$; thus type IIB on K3 $\times S^1$ must also develop enhanced gauge symmetry when K3 develops singularities. Does this imply that type IIB on K3 also develops enhanced gauge symmetry at these special points in the K3 moduli space? This does not seem possible, since type IIB string theory does not have any D-2 brane solitons which can be wrapped around the collapsed two cycles of K3. It turns out that instead of acquiring enhanced gauge symmetry, type IIB string theory acquires tensionless strings at these special points in the K3 moduli space [119]. These arise from taking a D-3 brane of type IIB string theory, and wrapping it on a two cycle of K3. Thus two of the tangential directions of the three brane are directed along the internal directions of K3, and the third direction of the three brane is along one of the non-compact spatial directions. Thus from the point of view of the $(5+1)$ dimensional theory such a configuration will appear as a string. The tension of this string is given by the product of the tension (energy per unit three volume) of the three brane and the area of the two cycle on which the three brane is wrapped. Thus as we approach the singular point on the K3 moduli space where the area of the two cycle vanishes, the tension of the string goes to zero. In other words, we get tensionless strings. Upon further compactification on a circle we get massless particles from configurations where this tensionless string is wound around the circle. These are precisely the massless gauge bosons required for the gauge symmetry enhancement in type IIB on K3 $\times S^1$.

[20] These states were analyzed in detail in [118].

6 Interrelation between different duality conjectures

In the last three sections we have seen many different duality conjectures and have learned how to test these conjectures. We shall now see that many of these conjectures are not independent, but can be "derived" from each other. There are several different ways in which dualities can be related to each other. We shall discuss them one by one. The material covered in this section is taken mainly from [75, 77, 78].

6.1 Combining non-perturbative and T-dualities

Suppose a string theory A compactified on a manifold $K_{\rm A}$ has a conjectured duality symmetry group G. Now further compactify this theory on some manifold $\mathcal{M}$. Then the theory A on $K_{\rm A} \times \mathcal{M}$ is expected to have the following set of duality symmetries:

- It inherits the original duality symmetry group G of A on $K_{\rm A}$;
- It also has a perturbatively verifiable T-duality group. Let us call it H.

Quite often G and H do not commute and together generate a much bigger group [23, 76]. In that case, the existence of this bigger group of symmetries can be regarded as a consequence of the duality symmetry of A on $K_{\rm A}$ and T-dualities.

We shall illustrate this with a specific example [23]. We have seen that in ten dimensions type IIB string theory has a conjectured duality group $SL(2, Z)$ that acts non-trivially on the coupling constant. From the table given in Section 4.5 we see that type IIB on T^n also has a T-duality group $SO(n, n; Z)$, whose existence can be verified order by order in string perturbation theory. It turns out that typically these two duality groups do not commute, and in fact generate the full duality symmetry group of type IIB on T^n as given in the table of Section 4.5. Thus we see that the existence of the full duality symmetry group of type IIB on T^n can be infered from the $SL(2, Z)$ duality symmetry of the ten dimensional type IIB string theory, and the perturbatively verifiable T-duality symmetries of type IIB on T^n.

6.2 Duality of dualities

Suppose two theories are conjectured to be dual to each other, and each theory in turn has a conjectured self-duality group. Typically part of this self duality group is T-duality, and the rest involves non-trivial transformation of the coupling constant. But quite often the non-perturbative duality transformations in one theory correspond to T-duality in the dual theory

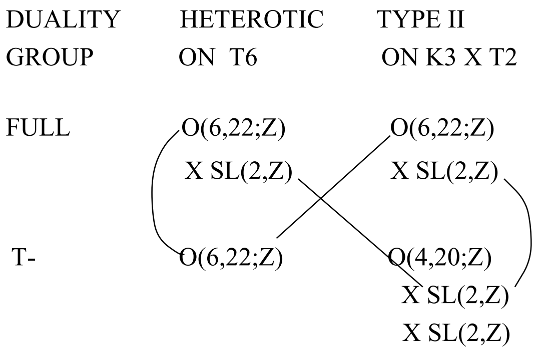


Fig. 13. The embedding of the T-duality groups in the full duality group in heterotic on T^6 and type IIA on $K3 \times T^2$.

and ,-, , ... As a result, the full self duality group in both theories follows from the conjectured duality between the two theories.

Again we shall illustrate this with an example [19, 24]. Let us start with the conjectured duality between heterotic on T^4 and type IIA on K3. Now let us compactify both theories further on a two dimensional torus T^2. This produces a dual pair of theories: type IIA on K3 $\times$ T^2 and heterotic on T^6. Now, heterotic on T^6 has a T-duality group $O(6, 22; Z)$ that can be verified using heterotic perturbation theory. On the other hand , type IIA on K3$\times T^2$ has a T-duality group $O(4, 20; Z) \times SL(2, Z) \times SL(2, Z)'$ that can be verified using type II perturbation theory. The full conjectured duality group in both theories is $O(6, 22, Z) \times SL(2, Z)$.

Now the question we would like to address is, how are the T-duality symmetry groups in the two theories embedded in the full conjectured $O(6, 22; Z) \times SL(2, Z)$ duality group? This has been illustrated in Figure 13. In particular we find that the $SL(2, Z)$ factor of the full duality group is a subgroup of the T-duality group in type IIA on K3 $\times$ T^2, and hence can be verified in this theory order by order in perturbation theory. On the other hand, the O(6,22;Z) factor of the duality group appears as a T-duality symmetry of the heterotic string theory, and hence can be verified order by order in perturbation theory in this theory. Thus assuming that T-duality in either theory is a valid symmetry, and the duality between the heterotic on T^4 and type IIA on K3, we can establish the existence of the self-duality group $O(6, 22; Z) \times SL(2, Z)$ in heterotic on T^6 and type IIA on K3$\times T^2$.

Using the results of this and the previous subsection, we see that so far among all the conjectured non-perturbative duality symmetries, the independent ones are:

1. $SL(2, Z)$ of type IIB in $D = 10$,
2. type I $\leftrightarrow$ $SO(32)$ heterotic in $D = 10$, and
3. IIA on K3 $\leftrightarrow$ heterotic on T^4.

We shall now show how to "derive" 3) from 1) and 2).

6.3 Fiberwise duality transformation

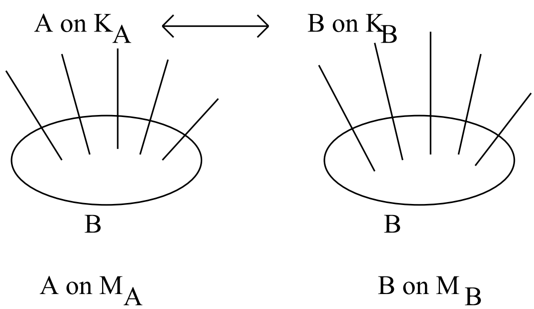


Fig. 14. Application of fiberwise duality transformation. In each local neighbourhood of the base manifold $\mathcal{B}$, the two theories are equivalent due to the equivalence of the theories living on the fiber ($\times$ any manifold). Thus we would expect the theories A on $\mathcal{M}_{\rm A}$ and B on $\mathcal{M}_{\rm B}$ to be equivalent.

In this subsection we shall describe the idea of constructing dual pairs of theories using fiberwise duality transformation [77]. Suppose (Theory A on $\mathcal{K}_{\rm A}$) has been conjectured to be dual to (Theory B on $\mathcal{K}_{\rm B}$). Here A and B are two of the five different string theories in $D = 10$, and $\mathcal{K}_{\rm A}$, $\mathcal{K}_{\rm B}$ are two different manifolds (in general). This duality involves a precise map between the moduli spaces of the two theories. Now construct a pair of new manifolds $\mathcal{M}_{\rm A}$, $\mathcal{M}_{\rm B}$ by starting from some other manifold $\mathcal{B}$, and erecting at every point on $\mathcal{B}$ a copy of $\mathcal{K}_{\rm A}$, $\mathcal{K}_{\rm B}$. The moduli of $\mathcal{K}_{\rm A}$, $\mathcal{K}_{\rm B}$ vary slowly over $\mathcal{B}$ and are related to each other by the duality map that relates (A on $\mathcal{K}_{\rm A}$) to (B on $\mathcal{K}_{\rm B}$). Then we would expect a duality

$$\text{Theory A on } \mathcal{M}_{\rm A} \leftrightarrow \text{Theory B on } \mathcal{M}_{\rm B}$$

by applying the duality transformation fiberwise. This then gives rise to a new duality conjecture. This situation has been illustrated in Figure 14.

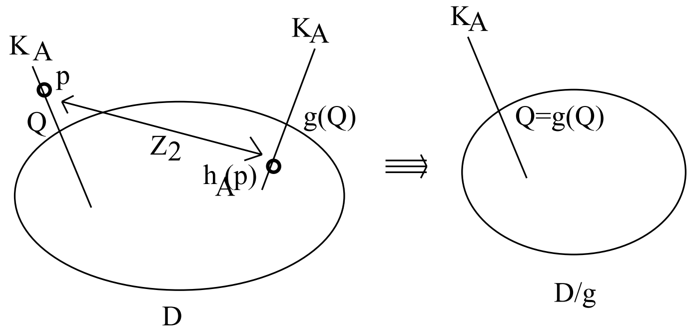


Fig. 15. Representation of a Z_2 orbifold as a fibered space. The Z_2 transformation relates the point (Q, p) on $\mathcal{D} \times \mathcal{K}_{\rm A}$ to the point $(g(Q), h_{\rm A}(p))$.

Now suppose that at some isolated points (or subspaces of codimension ≥ 1) on $\mathcal{B}$ the fibers $\mathcal{K}_{\rm A}$ and $\mathcal{K}_{\rm B}$ degenerate. (We shall see some explicit examples of this later.) Is the duality between (A on $\mathcal{M}_{\rm A}$) and (B on $\mathcal{M}_{\rm B}$) still valid? We might expect that even in this case the duality between the two theories holds since the singularities occur on subspaces of "measure zero". Although there is no rigorous argument as to why this should be so, this appears to be the case in all known examples. Conversely, assuming that this is the case, we can derive the existence of many new duality symmetries from a given duality symmetry.

A special case of this construction involves Z_2 orbifolds. Suppose we have a dual pair (A on $\mathcal{K}_{\rm A}$) $\leftrightarrow$ (B on $\mathcal{K}_{\rm B}$). Further suppose that (A on $\mathcal{K}_{\rm A}$) has a Z_2 symmetry generated by $h_{\rm A}$. Then the dual theory must also have a Z_2 symmetry generated by $h_{\rm B}$. $h_{\rm A}$ and $h_{\rm B}$ are mapped to each other under duality. Now compactify both theories on another manifold $\mathcal{D}$ with a Z_2 isometry generated by g, and compare the two quotient theories

$$(\text{A on } K_{\rm A} \times \mathcal{D}/h_{\rm A} \cdot g) \text{ and } (\text{B on } K_{\rm B} \times \mathcal{D}/h_{\rm B} \cdot g)$$

$(\mathcal{K}_{\rm A} \times \mathcal{D}/h_{\rm A} \cdot g)$ is obtained from the product manifold $\mathcal{K}_{\rm A} \times \mathcal{D}$ by identifying points that are related by the Z_2 transformation $h_{\rm A} \cdot g$. This situation is illustrated in Figure 15. As shown in this figure, $(\mathcal{K}_{\rm A} \times \mathcal{D}/h_{\rm A} \cdot g)$ admits a fibration with base $\mathcal{D}/g$ and fiber $\mathcal{K}_{\rm A}$. In particular, note that since $h_{\rm A} \cdot g$ takes a point $(p \in \mathcal{K}_{\rm A}, Q \in \mathcal{D})$ to $(h_{\rm A}(p), g(Q))$, if we focus our attention on a definite point Q on $\mathcal{D}$, then there is no identification of the points in the copy of $\mathcal{K}_{\rm A}$ that is sitting at Q. This shows that the fiber is $\mathcal{K}_{\rm A}$ and . . . $\mathcal{K}_{\rm A}/h_{\rm A}$. As we go from Q to $g(Q)$, which is a closed cycle on $\mathcal{D}/g$, the fiber gets twisted by the transformation $h_{\rm A}$.

The second theory, B on $(\mathcal{K}_{\rm B} \times \mathcal{D})/(h_{\rm B} \cdot g)$ has an identical structure. Thus we can now apply fiberwise duality transformation to derive a new duality:

$$(\text{A on } K_\text{A} \times \mathcal{D}/h_\text{A} \cdot g) \leftrightarrow (\text{B on } K_\text{B} \times \mathcal{D}/h_\text{B} \cdot g)$$

Note that if $P_0 \in \mathcal{D}$ is a fixed point of g (if $g(P_0) = P_0$) then in $\mathcal{K}_\text{A} \times \mathcal{D}/h_\text{A} \cdot g$ there is an identification of points (p, P_0) and $(h_\text{A}(p), P_0)$. Similarly in $\mathcal{K}_\text{B} \times \mathcal{D}/h_\text{B} \cdot g$ there is an identification of points (p', P_0) and $(h_\text{B}(p'), P_0)$. Thus at P_0 the fibers degenerate to $\mathcal{K}_\text{A}/h_\text{A}$ and $\mathcal{K}_\text{B}/h_\text{B}$ respectively. At these points the argument in support of duality between the two theories breaks down. However, as we have discuused earlier, since these are points of "measure zero" on $\mathcal{D}$, we would expect that the two quotient theories are still dual to each other [78]. We shall now illustrate this construction in the context of a specific example.

We start with type IIB string theory in ten dimensions. This has a conjectured $SL(2,Z)$ symmetry. Let S denote the $SL(2,Z)$ element

$$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} . \tag{6.1}$$

Recall that this theory also has two global discrete symmetries $(-1)^{F_\text{L}}$ and Ω. The action of S, $(-1)^{F_\text{L}}$ and Ω on the massless bosonic fields in this theory were described in Section 4.4. From this one can explicitly compute the action of $S(-1)^{F_\text{L}}S^{-1}$ on these massless fields. This action turns out to be identical to that of Ω. A similar result holds for their action on the massless fermionic fields as well. Finally, since the action of S on the massive fields is not known, one can define this action in such a way that the actions of $S(-1)^{F_\text{L}}S^{-1}$ and Ω are identical on all states. This gives:

$$S(-1)^{F_\text{L}}S^{-1} = \Omega \,. \tag{6.2}$$

We are now ready to apply our formalism. We take (A on $\mathcal{K}_\text{A}$) to be type IIB in $D = 10$, (B on $\mathcal{K}_\text{B}$) to be type IIB in $D = 10$ transformed by S, h_A to be $(-1)^{F_\text{L}}$, h_B to be Ω, $\mathcal{D}$ to be T^4, and g to be the transformation $\mathcal{I}_4$ that changes the sign of all the coordinates on T^4. This gives the duality:

$$(\text{IIB on } T^4/(-1)^{F_\text{L}} \cdot \mathcal{I}_4) \leftrightarrow (\text{IIB on } T^4/\Omega \cdot \mathcal{I}_4).$$

Note that in this case the fibers $\mathcal{K}_\text{A}$ and $\mathcal{K}_\text{B}$ are points, but this does not prevent us from applying our method of constructing dual pairs. Also there are sixteen fixed points on T^4 under $\mathcal{I}_4$ where the application of fiberwise duality transformation breaks down, but as has been argued before, we still expect the duality to hold since these are points of measure zero on T^4.

We shall now bring this duality into a more familiar form . T-duality transformation. Let us make $R \to (1/R)$ duality transformation on one of the circles of T^4 in the theory on the left hand side. This converts type IIB theory to type IIA. This also transforms $(-1)^{F_L} \cdot \mathcal{I}_4$ to $\mathcal{I}_4$, which can be checked by explicitly studying the action of these transformations on the various massless fields. Thus the theory on the left hand side is T-dual to type IIA on $T^4/\mathcal{I}_4$. This of course is just a special case of type IIA on K3.

Let us now take the theory on the right hand side and make $R \to (1/R)$ duality transformation on all four circles. This takes type IIB theory to type IIB theory. But this transforms $\Omega \cdot \mathcal{I}_4$ to Ω, which can again be seen by studying the action of these transformations on the massless fields. Thus the theory on the right is T-dual to type IIB on T^4/Ω. Since type I string theory can be regarded as type IIB string theory modded out by Ω, we see that the theory on the right hand side is type I on T^4. But by (heterotic – type I) duality in ten dimensions this is dual to heterotic on T^4. Thus we have "derived" the duality

$$(\text{Type IIA on K3}) \leftrightarrow (\text{Heterotic on } T^4)$$

from other conjectured dualities in $D = 10$. Although this way the duality has been established only at a particular point in the moduli space (the orbifold limit of K3), the argument can be generalized to establish this duality at a generic point in the moduli space as well [78].

There are many other applications of fiberwise duality transformation. Some of them will be discussed later in this review.

6.4 Recovering higher dimensional dualities from lower dimensional ones

So far we have discussed methods of deriving dualities involving compactified string theories by starting with the duality symmetries of string theories in higher dimensions. But we can also proceed in the reverse direction. Suppose a string theory compactified on a manifold $\mathcal{M}_1 \times \mathcal{M}_2$ has a self-duality symmetry group G. Now consider the limit when the size of $\mathcal{M}_2$ goes to infinity. A generic element of G, acting on this configuration, will convert this configuration to one where $\mathcal{M}_2$ has small or finite size. However, there may be a subgroup H of G that commutes with this limit, any element of this subgroup, acting on a configuration where $\mathcal{M}_2$ is big, gives us back a configuration where $\mathcal{M}_2$ is big. Thus we would expect that H is the duality symmetry group of the theory in the decompactification limit, of the original string theory compactified on $\mathcal{M}_1$. The same argument can be extended to the case of a pair of dual theories.

this procedure does not appear to be very useful, since one normally likes to derive more complicated duality transformations of lower dimensional theories from the simpler ones in the higher dimensional theory. But we shall now show how this procedure can be used to derive the $SL(2, Z)$ duality symmetry of type IIB string theory from the conjectured duality between type I and $SO(32)$ heterotic string theories, and

T-duality symmetries of the heterotic string theory. We shall describe the main steps in this argument, for details, see [75]. We start with the duality between type I on T^2 and heterotic on T^2 that follows from the duality between these theories in ten dimensions. Now heterotic string theory on T^2 has a T-duality group O(2,18;Z). We shall focus our attention on an $SL(2,Z) \times SL(2,Z)$ subgroup of this T-duality group. One of these two $SL(2,Z)$ factors is associated with the global diffeomorphism of T^2, and the other one is associated with the $R \to (1/R)$ duality symmetries on the two circles. By the "duality of dualities" argument, this must also be a symmetry of type I on T^2. Since type I string theory can be regarded as type IIB string theory modded out by the world-sheet parity transformation Ω discussed in Section 4.4, we conclude that $SL(2,Z) \times SL(2,Z)$ is a subgroup of the self-duality group of type IIB on T^2/Ω. Let us now make an $R \to (1/R)$ duality transformation on both the circles of this T^2. This converts type IIB on T^2 to type IIB theory compactified on a dual T^2, and Ω to $(-1)^{F_L} \cdot \Omega \cdot \mathcal{I}_2$, where $\mathcal{I}_2$ denotes the reversal of orientation of both the circles of T^2. (This can be seen by studying the action of various transformations on the massless fields.) Geometrically, this model describes type IIB string theory compactified on the surface of a tetrahedron (which is geometrically $T^2/\mathcal{I}_2$), with an added twist of $(-1)^{F_L} \cdot \Omega$ as we go around any of the four vertices of the tetrahedron (the fixed points of $\mathcal{I}_2$). Thus we conclude that $SL(2,Z) \times SL(2,Z)$ is a subgroup of the self-duality group of type IIB on a tetrahedron. Now take the limit where the size of the tetrahedron goes to infinity. It turns out that both the $SL(2,Z)$ factors commute with this limit. One of these $SL(2,Z)$ groups becomes part of the diffeomorphism group of type IIB string theory and does not correspond to anything new, but the other $SL(2,Z)$ factor represents the S-duality transformation discussed in Section 4.4. Since this limit gives us back the decompactified type IIB string theory, we conclude that type IIB string theory in ten dimensions has a self-duality group $SL(2,Z)$.

Thus we see that all the dualities discussed so far can be "derived" from a single duality conjecture, – the one between type I and $SO(32)$ heterotic string theories in ten dimensions. In the next section we shall see more examples of dualities which can be derived from the ones that we have already discussed.

7 Duality in theories with less than sixteen supersymmetry generators

So far our discussion has been focussed on theories with 16 or more supersymmetry charges. As was pointed out in Section 4, for these theories the non-renormalization theorems for the low energy effective action and the spectrum of BPS states are particularly powerful. This makes it easy to test duality conjectures involving these theories. In this section we shall extend our discussion to theories with eight supercharges. Examples of such theories are provided by $N = 2$ supersymmetric theories in four dimensions. We shall see that these theories have a very rich structure, and although the non-renormalization theorems are less powerful here, they are still powerful enough to provide us with some of the most striking tests of duality conjectures involving these theories. The material covered in this section is based mainly on refs. [77, 79, 80].

7.1 Construction of a dual pair of theories with eight supercharges

For definiteness we shall focus our attention on $N = 2$ supersymmetric theories in four dimensions. There are several ways to get theories with $N = 2$ supersymmetry in four dimensions. Two of them are:

1. Type IIA/IIB on Calabi-Yau 3-folds: in our convention an n-fold describes an n complex or $2n$ real dimensional manifold. In ten dimensions type II theories have 32 supersymmetry generators. Compactification on a Calabi-Yau 3-fold breaks 3/4 of the supersymmetry. Thus we are left with 8 supersymmetry generators in $D = 4$, giving rise to $N = 2$ supersymmetry;

2. Heterotic string theory on $K3 \times T^2$: in ten dimensions heterotic string theory has sixteen supersymmetry generators. Compactification on $\mathrm{K3} \times T^2$ breaks half of the supersymmetry. Thus we have a theory with eight supersymmetry generators, again giving $N = 2$ supersymmetry in four dimensions. It is also possible to construct more general class of four dimensional heterotic string theories with the same number of supersymmetries where the background does not have the product structure $\mathrm{K3} \times T^2$ [79, 108].

The question we would like to ask is: is it possible to construct pairs of $N = 2$ supersymmetric type II and heterotic string compactifications in four dimensions which will be non-perturbatively dual to each other? Historically such dual pairs were first constructed by trial and error [79] and then a more systematic approach was developed [77, 80–83]. However we shall begin by describing the systematic approach, and then describe how one tests these dualities. The systematic construction of such dual pairs can be carried out by application of fiberwise duality transformation as described in the last section. The steps involved in this construction are as follows:

- Start from the conjectured duality (Type IIA on K3) $\leftrightarrow$ (Heterotic on T^4);

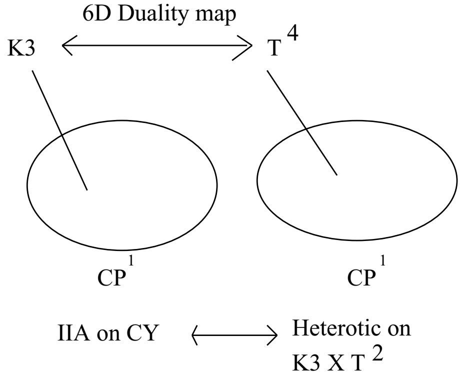


Fig. 16. Construction of dual pair of $N = 2$ supersymmetric string theories in four dimensions from the dual pair of theories in six dimensions.

- Choose a CP^1 base;
- Construct a Calabi-Yau 3-fold by fibering K3 over the base CP^1; One can construct a whole class of Calabi-Yau manifolds this way by choosing different ways of varying K3 over CP^1;
- For type IIA on each such Calabi-Yau 3-fold we can get a dual heterotic compactification by replacing the type IIA on $K3$ by heterotic on T^4 on each fiber according to the duality map. This gives heterotic string theory on a manifold obtained by varying T^4 on CP^1 according to the duality map. Typically this manifold turns out to be K3$\times T^2$ or some variant of this. This model is expected to be dual to the type IIA string theory on the Calabi-Yau manifold that we started with. Thus we get a duality map

$$\text{(Type IIA on CY)} \leftrightarrow \text{(Heterotic on K3}\times T^2\text{)}.$$

This construction has been illustrated in Figure 16. Note that the original duality map gives a precise relationship between the moduli of type IIA on K3 and heterotic on T^4. On the heterotic side the moduli involve background gauge fields on T^4 besides the shape and size of T^4. Thus for a specific Calabi-Yau, knowing how K3 varies over CP^1, we can find out how on the heterotic side the background gauge fields on T^4 vary as we move along CP^1. This gives the gauge field configuration on K3 $\times T^2$. Different Calabi-Yau manifolds will give rise to different gauge fields on K3 $\times T^2$.

We shall illustrate this procedure with the example of a pair of Z_2 orbifolds of the form [80]:

$$(\text{IIA on K3} \times T^2/h_{\text{A}} \cdot g) \leftrightarrow (\text{Heterotic on } T^4 \times T^2/h_{\text{B}} \cdot g)$$

where g acts on T^2 by changing the sign of both its coordinates, h_{A} is a specific involution of K3 known as the Enriques involution, and h_{B} is the image of this transformation on the heterotic side. By our previous argument relating orbifolds to fibered spaces, these two theories are expected to be dual to each other [illegible] fiberwise duality transformation. K3$\times T^2/(h_{\text{A}} \cdot g)$ can be shown to describe a Calabi-Yau manifold. Thus the theory on the left-hand side corresponds to type IIA string theory compactified on this Calabi-Yau manifold. In order to determine the theory on the heterotic side, we need to determine h_{B}. We shall now describe this procedure in some detail.

In order to determine h_{B}, we need to study the relationship between the fields appearing in type IIA on K3 and heterotic on T^4. The low energy effective action of both the theories and the origin of the various massless fields in these theories were discussed in Section 4.3. We shall focus our attention on the gauge fields. As discussed there, in the type IIA on K3, 22 of the gauge fields come from decomposing the three form field along the harmonic two forms on K3. Now, h_{A}, being a geometric transformation on K3, has known action on the harmonic forms $\omega^{(p)}$. For this particular example, h_{A} corresponds to

- exchanging ten of the $\omega^{(p)}$ with ten others and
- changing the sign of two more $\omega^{(p)}$.

This translates into a similar action on the fields $\mathcal{A}_\mu^{(p)}$ defined in equation (4.36). Furthermore, h_{A} leaves the other two gauge fields, coming from the ten dimensional gauge field A_μ and the dual of $C_{\mu\nu\rho}$ invariant. We can now translate this into an action on the gauge fields in heterotic on T^4. It turns out that the action on the heterotic side is given by:

- exchanging the gauge fields in the two E_8 factors,
- exchanging $(G_{9\mu}, B_{9\mu})$ with $(G_{8\mu}, B_{8\mu})$, and,
- changing the sign of $(G_{7\mu}$ and $B_{7\mu})$.

This translates into the following geometric action in heterotic string theory on T^4:[21]

- exchange of two E_8 factors in the gauge group,
- $x^8 \leftrightarrow x^9$,
- $x^7 \to -x^7$.

This is $h_{\rm B}$[22]. It turns out that modding out heterotic string theory on T^6 by the transformation $h_{\rm B} \cdot g$ produces an $N = 2$ supersymmetric theory. Thus this construction gives a type II – heterotic dual pair with $N = 2$ supersymmetry.

Using the idea of fiberwise duality transformation we can construct many more examples of heterotic – type IIA dual pairs in four dimensions with $N = 2$ supersymmetry. Quite often using mirror symmetry [109] we can also relate this to IIB string theory on a mirror Calabi-Yau manifold.

7.2 Test of duality conjectures involving theories with eight supercharges

Given such a dual pair of theories constructed by application of fiberwise duality transformation, the next question will be: how do we test if these theories are really dual to each other? After all, as we have seen, there is no rigorous proof that fiberwise duality transformation always produces a correct dual pair of theories, particularly when the fiber degenerates at some points/regions in the base. Unlike in the case of theories with sixteen supercharges, one cannot directly compare the tree level low energy effective action in the two theories, as they undergo quantum corrections in general. Furthermore, in this theory the spectrum of BPS saturated states can change discontinuously as we move in the moduli space [88]. Hence the spectrum computed at weak coupling cannot always be trusted at strong coupling. Nevertheless there are some non-renormalization theorems which allow us to test these proposed dualities, as we shall now describe.

Matter multiplets in $N = 2$ supersymmetric theories in four dimensions are of two types. (For a review, see [88].) They are

- vector multiplet containing one vector, one complex scalar, and two Majorana fermions, and

[21] Here we are regarding this theory as the $E_8 \times E_8$ heterotic string theory compactified on T^4. By the duality between the two heterotic string theories upon compactification on a circle, this is equivalent to $SO(32)$ heterotic string theory compactified on T^4.

[22] We need to add to this a non-geometrical shift involving half of a lattice vector in Λ_{24} in order to get a modular invariant theory on the heterotic side. This transformation is not visible in perturbative type IIA theory.

- hypermultiplet containing two complex scalars and two Majorana fermions.

Let us consider a theory at a generic point in the moduli space where the massless matter fields include only abelian gauge fields and neutral hypermultiplets. Let $\vec{\phi}$ denote the complex scalars in the vector multiplet, and $\vec{\psi}$ denote the complex scalars in the hypermultiplet. The $N = 2$ supersymmetry requires that there is no coupling between the vector and the hypermultiplets in those terms in the low energy effective action S_{eff} which contain at most two space-time derivatives [84]. Thus the scalar kinetic terms appearing in the Lagrangian density associated with S_{eff} must be of the form:

$$G^{V}_{m\bar{n}}(\vec{\phi})\partial_\mu\phi^m\partial^\mu\bar{\phi}^n + G^{H}_{\alpha\bar{\beta}}(\vec{\psi})\partial_\mu\psi^\alpha\partial^\mu\bar{\psi}^\beta\,, \tag{7.1}$$

where G^{V} and G^{H} are appropriate metrics in the vector and the hypermultiplet moduli spaces. The kinetic terms of the vectors and the fermionic fields are related to these scalar kinetic terms by the requirement of $N = 2$ supersymmetry.

This decoupling between the hyper- and the vector- multiplet moduli spaces by itself is not of much help, since each term may be independently modified by quantum corrections[23]. But in string theory we have some extra ingredient [77, 79]. Recall that the coupling constant in string theory involves the dilaton. Thus quantum corrections to a given term must involve a coupling to the dilaton. Now consider the following two special cases.

1. The dilaton belongs to a hypermultiplet. Then there can be no correction to the vector multiplet kinetic term since such corrections will give a coupling between the dilaton and the vector multiplet;

2. The dilaton belongs to a vector multiplet. In this case the same argument shows that there can be no correction to the hypermultiplet kinetic term.

In type IIA/IIB string theory on Calabi-Yau manifold the dilaton belongs to a hypermultiplet. Thus in these theories the vector multiplet kinetic term, calculated at the tree level, is exact. On the other hand in heterotic on $\text{K3} \times T^2$, the dilaton is in the vector multiplet. Thus the hypermultiplet

[23] There are however strong restrictions on what kind of metric G_{V} and G_{H} should describe. In particular G_{V} must describe a special Kahler geometry [84, 110], whereas G_{H} must describe a quaternionic geometry [111]. However, these restrictions do not fix G_{V} and G_{H} completely.

kinetic term, calculated at the tree level, is exact. Using this information we can adopt the following strategy for testing duality[24].

1. Take a type II – heterotic dual pair and calculate the vector multiplet kinetic term exactly from the tree level analysis on the type II side;

2. Using the map between the fields in the type II and the heterotic theory, we can rewrite the exact vector multiplet kinetic term in terms of the heterotic variables;

3. In particular the heterotic variables include the heterotic dilaton $\Phi_{\rm H}$ which is in the vector multiplet. So we can now expand the exact answer in powers of $e^{\Phi_{\rm H}}$ and compare this answer with the explicit calculations in heterotic string perturbation theory. Typically the expansion involves tree, one loop, and non-perturbative terms. (There is no perturbative contribution in the heterotic theory beyond one loop due to some Adler-Bardeen type non-renormalization theorems.) Thus one can compare the expected tree and one loop terms, calculated explicitly in the heterotic string theory, with the expansion of the exact answer.

The results of the above calculation in heterotic and type II string theories agree in all the cases tested [79,85,86]! This agreement is quite remarkable, since the one loop calculation is highly non-trivial on the heterotic side, and involves integrals over the moduli space of the torus. Indeed, the agreement between the two answers is a consequence of highly non-trivial mathematical identities.

Given that the tree and one loop results in the heterotic string theory agree with the expansion of the exact result on the type II side, one might ask if a similar agreement can be found for the non-perturbative contribution from the heterotic string theory as well. From the exact answer calculated from the type II side we know what this contribution should be. But we cannot calculate it directly on the heterotic side, since there is no non-perturbative formulation of string theory. However, one can take an appropriate limit in which the stringy effects on the heterotic side disappear and the theory reduces to some appropriate $N = 2$ supersymmetric quantum field theory[25]. Thus now the calculation of these non-perturbative effects on the heterotic side reduces to a calculation in the $N = 2$ supersymmetric field theory. This can be carried out using the method developed by

[24] Here we describe the test using the vector multiplet kinetic term, but a similar analysis should be possible with the hypermultiplet kinetic term as well.

[25] This is in the same spirit as in the case of toroidal compactification of heterotic string theory, where, by going near a special point in the moduli space, we can effectively get an $N = 4$ supersymmetric Yang-Mills theory.

Seiberg and Witten [88]. Again there is perfect agreement with the results from the type II side [87]. Besides providing a non-trivial test of string duality, this also shows that the complete Seiberg-Witten [88] results (and more) are contained in the classical geometry of Calabi-Yau spaces!

8 M-theory

So far we have discussed dualities that relate known string theories. However, sometime analysis similar to those that lead to various duality conjectures can also lead to the discovery of new theories. One such theory is a conjectured theory living in eleven dimensions. This theory is now known as M-theory. In this section we shall give a brief description of this theory following references [19, 93–98].

8.1 M-theory in eleven dimensions

The arguments leading to the existence of M-theory goes as follows [19, 89]. Take type IIA string theory in ten dimensions. The low energy effective action of this theory is non-chiral $N = 2$ supergravity in ten dimensions. It is well known that this can be obtained from the dimensional reduction of $N = 1$ supergravity in eleven dimensions [90]. More specifically, the relationship between the two theories is as follows. The bosonic fields in $N = 1$ supergravity theory in eleven dimensions consist of the metric $g^{(\mathrm{S})}_{\mathrm{MN}}$ and a rank three anti-symmetric tensor field $C^{(\mathrm{S})}_{\mathrm{MNP}}$ $(0 \leq M, N \leq 10)$. The bosonic part of the action of this theory is given by [112]

$$\begin{aligned} S_{\mathrm{SG}} &= \frac{1}{(2\pi)^8} \int \mathrm{d}^{11}x \Bigg[\sqrt{-g^{(\mathrm{S})}} \Big(R^{(\mathrm{S})} - \frac{1}{48} G^{(\mathrm{S})2} \Big) \\ &\quad - \frac{1}{(12)^4} \varepsilon^{\mu_0 \cdots \mu_{10}} C^{(\mathrm{S})}_{\mu_0\mu_1\mu_2} G^{(\mathrm{S})}_{\mu_3\cdots\mu_8} G^{(\mathrm{S})}_{\mu_7\cdots\mu_{10}} \Bigg], \end{aligned} \tag{8.1}$$

where $G^{(\mathrm{S})} \sim \mathrm{d}C^{(\mathrm{S})}$ is the four form field strength associated with the three form field $C^{(\mathrm{S})}$. In writing down the above equation we have set the eleven dimensional Planck mass to unity (or equivalently we can say that we have absorbed it into a redefinition of the metric.) Let us now compactify this supergravity theory on a circle of radius $R(\sim \sqrt{g^{(\mathrm{S})}_{10,10}})$ measured in the supergravity metric $g^{(\mathrm{S})}_{\mathrm{MN}}$ and ignore (for the time being) the Kaluza-Klein modes carrying momentum in the internal direction. Then the effective action in the dimensionally reduced theory agrees with that of type IIA string theory given in (4.33) under the identification [90]:

$$\sqrt{g^{(\mathrm{S})}_{10,10}} = \mathrm{e}^{\Phi/3}, \qquad g^{(\mathrm{S})}_{\mu\nu} \simeq \mathrm{e}^{-\Phi/12} g_{\mu\nu} \qquad g^{(\mathrm{S})}_{10\mu} \simeq \mathrm{e}^{2\Phi/3} A_\mu,$$
$$C^{(\mathrm{S})}_{\mu\nu\rho} \simeq C_{\mu\nu\rho}, \qquad C^{(\mathrm{S})}_{10\mu\nu} \simeq B_{\mu\nu}, \qquad (0 \le \mu, \nu \le 9)\,. \tag{8.2}$$

Here $\simeq$ denotes equality up to additive terms involving second and higher

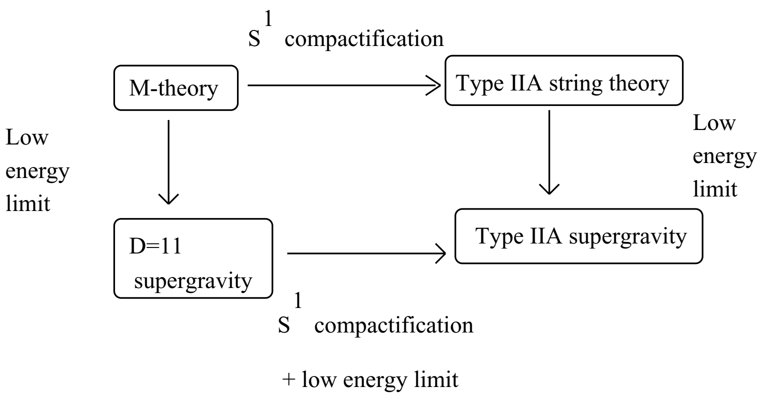


Fig. 17. The relationship between M-theory and various other supergravity/string theories.

powers in fields. We are using the convention that $\Phi = 0$ corresponds to compactification on a circle of unit radius. Note that as the radius $R(\sim \sqrt{g^{(\mathrm{S})}_{10,10}})$ approaches ∞, $\Phi \to \infty$. This corresponds to strong coupling limit of the type IIA string theory. This leads one to the conjecture [19, 89] that [illegible] $N = 1$ [illegible] This theory has been called M-theory. The situation is illustrated in Figure 17. Part of the conjecture is just the definition of M-theory as the strong coupling limit of type IIA string theory. The non-trivial part of the conjecture is that it describes a Lorentz invariant theory in eleven dimensions.

The evidence for the existence of an eleven dimensional theory, as discussed so far, has been analogous to the evidence for various duality conjectures based on the comparison of their low energy effective action. One might ask if there are more precise tests involving the spectrum of BPS states. There are indeed such tests. M theory on S^1 will have Kaluza-Klein modes representing states in the eleven dimensional $N = 1$ supergravity multiplet carrying momentum along the compact x^{10} direction. These are BPS states, and can be shown to belong to the 256 dimensional ultra-short representation of the supersymmetry algebra. The charge quantum number characterizing such a state is the momentum (k/R) along S^1. Thus for every integer k we should find such BPS states in type IIA string theory in ten dimensions. In M-theory these states carry k units of $g^{(\mathrm{S})}_{10\mu}$ charge. Since $g^{(\mathrm{S})}_{10\mu}$ gets mapped to A_μ under the M-theory - IIA duality, these states must carry k units of A_μ charge in type IIA string theory. If we now recall that in type IIA string theory A_μ arises in the RR sector, we see that these states cannot come from elementary string states, as elementary string excitations are neutral under RR sector gauge fields. However Dirichlet 0-branes in this theory do carry A_μ charge. In particular the state with $k = 1$ corresponds to a single Dirichlet zero brane. As usual, the collective coordinate dynamics of the 0-branes is determined from the dynamics of massless open string states with ends lying on the D0-brane, and in this case is described by the dimensional reduction of $N = 1$ super-Maxwell theory from $(9 + 1)$ to $(0 + 1)$ dimensions. This theory has sixteen fermion zero modes whose quantization leads to a $2^8 = 256$ fold degenerate state. Thus we see that we indeed have an ultra-short multiplet with unit A_μ charge, as predicted by the M-theory - IIA duality conjecture.

What about states with $k > 1$? In type IIA string theory these must arise as bound states of k D0-branes. Dynamics of collective coordinates of k D0 branes is given by the dimensional reduction of $N = 1$ supersymmetric $U(k)$ gauge theory from $(9 + 1)$ to $(0 + 1)$ dimensions. Thus the number of ultra-short multiplets with k-units of A_μ charge is determined in terms of the number of normalizable supersymmetric ground states of this quantum mechanical system. Finding these bound states is much more difficult than the bound state problems discussed earlier. The main obstacle to this analysis is that a charge k state has the same energy as k charge 1 states at rest. Thus the bound states we are looking for sit at the bottom of a continuum. Such states are difficult to study. For $k = 2$ such a bound state with the correct degeneracy has been found [91]. The analysis for higher k still remains to be done.

The analysis can be simplified by compactifying M-theory on T^2 and considering the Kaluza-Klein modes carrying (k_1, k_2) units of momenta along the two S^1's. Assuming that the two S^1's are orthogonal, and have radii R_1 and R_2 respectively, the mass of such a state, up to a proportionality factor, is

$$\sqrt{\left(\frac{k_1}{R_1}\right)^2 + \left(\frac{k_2}{R_2}\right)^2}\ . \tag{8.3}$$

For (k_1, k_2) relatively prime, such a state has strictly less energy than the sum of the masses of any other set of states with the same total charge [92]. Thus one should be able to find these states in type IIA string theory on

S^1 (which, according to the conjecture, is equivalent to M-theory on T^2) without encountering the difficulties mentioned earlier. By following the same kind of argument, these states can be shown to be in one to one correspondence to a class of supersymmetric vacua in a $(1+1)$ dimensional supersymmetric gauge theory compactified on a circle[26]. All such states have been found with degeneracy as predicted by the M-theory – IIA duality.

There are also other consistency checks on the proposed M-theory – IIA duality. Consider M-theory on T^2. According to M-theory – type IIA duality, it is dual to IIA on S^1. But we know that IIA on S^1 is related by T-duality to IIB on S^1. Thus we have a duality between M-theory on T^2 and IIB on S^1. Now IIB on S^1 has an $SL(2,Z)$ strong-weak coupling duality inherited from ten dimensional type IIB string theory. Thus one might ask, what does it correspond to in M-theory on T^2? One can find the answer by using the known map between the massless fields in the two theories, and the action of $SL(2,Z)$ in type IIB string theory. It turns out that this $SL(2,Z)$ symmetry in M-theory is simply the group of global diffeomorphisms of T^2 [61, 93, 94]. Thus we again have an example of "duality of dualities". The $SL(2,Z)$ of IIB is a non-perturbative symmetry. But in M-theory on T^2 it is simply a consequence of the diffeomorphism invariance of the 11-dimensional theory.

Turning this analysis around we see that this also supports the ansatz that M-theory, defined as the strong coupling limit of IIA, is a fully Lorentz invariant theory in eleven dimension. The argument goes as follows:

- First of all, from Lorentz invariance of type IIA string theory we know that we have Lorentz invariance in coordinates $x^0, \dots x^9$ when all the coordinates $x^0, \dots x^9$ are non-compact;

- Then from the conjectured $SL(2,Z)$ duality symmetry of type IIB string theory we know that we have an exchange symmetry between the 9th and the 10th coordinate of M-theory when these coordinates are compact. In the limit when the radius of both the compact circles are taken to be large, this would mean that we should have Lorentz invariance in coordinates $x^0, \dots x^{10}$.

8.2 Compactification of M-theory

Given the existence of M-theory, we can now construct new vacua of the theory by compactifying M-theory on various manifolds. (For a review of

[26]In fact, these states are related *via* an $R \to (1/R)$ duality transformation to the ultra-short multiplets in type IIB on S^1 discussed in Section 5.2.

compactification of eleven dimensional supergravity, see [138].) For example, we can consider M-theory compactified on K3, Calabi-Yau, and various orbifolds. These can all be regarded as appropriate strong coupling limits of type IIA compactification on the same manifold. But in general these cannot be regarded as perturbative string vacua. The essential feature of this strong coupling limit is the emergence of Lorentz invariance in one higher dimension. For example, M-theory on a Calabi-Yau manifold gives a five dimensional theory with $N = 1$ supersymmetry [137]. Such a theory cannot be constructed by conventional compactification of type IIA string theory at weak coupling.

Of course in many cases these non-perturbative vacua are related to perturbative string vacua by conjectured duality relations. These duality conjectures can be arrived at by using arguments very similar to those used in arriving at string duality conjectures. Some examples of such conjectured dualities are given below [95–98]:

$$
\begin{array}{rcl}
\text{M-theory on} & & \\
S^1/Z_2 & \leftrightarrow & (E_8 \times E_8) \text{ heterotic in } D = 10 \\
K3 & \leftrightarrow & \text{Heterotic/Type I on } T^3 \\
T^5/Z_2 & \leftrightarrow & \text{IIB on K3} \\
T^8/Z_2 & \leftrightarrow & \text{Type I/Heterotic on } T^7 \\
T^9/Z_2 & \leftrightarrow & \text{Type IIB on } T^8/Z_2
\end{array}
$$

In each case Z_2 acts by reversing the sign of all the coordinates of T^n; for odd n this is also accompanied by a reversal of sign of $C^{(\mathrm{S})}_{\mathrm{MNP}}$. Each of these duality conjectures satisfy the consistency condition that the theory on the right hand side, upon further compactification on a circle, is dual to type IIA string theory compactified on the manifold on the left hand side.

The duality between M-theory on S^1/Z_2 and the $E_8 \times E_8$ heterotic string theory is particularly amusing. Here the Z_2 transformation acts by reversing the orientation of S^1, together with a change of sign of the three form field $C^{(\mathrm{S})}_{\mathrm{MNP}}$. S^1/Z_2 denotes a real line segment bounded by the two fixed points on S^1. It turns out that the two E_8 gauge multiplets arise from "twisted sector" of the theory and sit at the two ends of this line segment. The supergravity sector, on the other hand, sits in the bulk. Now in the conventional heterotic string compactification on Calabi-Yau spaces, all the observed gauge bosons and charged particles come from one E_8 and are neutral under the second E_8 [128]. The second E_8, known as the hidden sector or the shadow world, is expected to be responsible for supersymmetry breaking. In the M-theory picture these two sectors are physically separated in space. In other words, the real world and the shadow world live at two ends of the line and interact only [illegible] the exchange of supergravity multiplets propagating in the bulk [136]! It has been suggested that this physical separation could be as large as a millimeter [99]! This limit comes from the analysis of the fifth force experiment, since if this dimension is too large, we should have inverse cube law for the gravitational force instead of inverse square law. No such direct limit comes from the inverse square law of gauge interaction, since gauge fields live on the boundary of S^1/Z_2 and hence do not get affected by the existence of this extra dimension.

Many of the listed duality conjectures involving M-theory (in fact all except the first one) can be derived by fiberwise duality transformation [98][27]. Let us for example consider the duality

$$(\text{M theory on } T^5/Z_2) \leftrightarrow (\text{type IIB on K3}).$$

The Z_2 generator is $\mathcal{I}_5 \cdot \sigma$ where $\mathcal{I}_5$ changes the sign of all five coordinates $(x^6, \ldots x^{10})$ on T^5, and σ denotes the transformation $C^{(\mathrm{S})}_{\mathrm{MNP}} \to -C^{(\mathrm{S})}_{\mathrm{MNP}}$. Let us express this as $(\mathcal{I}_1 \cdot \sigma) \cdot \mathcal{I}_4$ where $\mathcal{I}_1$ changes the sign of x^{10}, and $\mathcal{I}_4$ changes the sign of $(x^6, \ldots x^9)$. We now use the result of fiberwise duality transformation:

$$(\text{A on } K_{\mathrm{A}} \times \mathcal{D}/(h_{\mathrm{A}} \cdot g)) \equiv (\text{B on } K_{\mathrm{B}} \times \mathcal{D}/(h_{\mathrm{B}} \cdot g))$$

by choosing A on K_{A} to be M-theory on S^1, B on K_{B} to be type IIA string theory, h_{A} to be $\mathcal{I}_1 \cdot \sigma$, h_{B} to be $(-1)^{F_{\mathrm{L}}}$ (this can be shown to be the image of h_{A} in the type IIA string theory), $\mathcal{D}$ to be T^4 spanned by $x^6, \ldots x^9$, and g to be $\mathcal{I}_4$. Thus we get the duality

$$\text{M-theory on } (S^1 \times T^4/\mathcal{I}_1 \cdot \sigma \cdot \mathcal{I}_4) \leftrightarrow \text{IIA on } T^4/(-1)^{F_L} \cdot \mathcal{I}_4.$$

The theory on the left hand side is M-theory on T^5/Z_2. On the other hand, if we take the theory on the right hand side and make an $R \to (1/R)$ duality transformation on one of the circles, it converts

- type IIA theory to type IIB theory, and
- $(-1)^{F_{\mathrm{L}}} \cdot \mathcal{I}_4$ into $\mathcal{I}_4$.

Thus the theory on the right is dual to type IIB on $T^4/\mathcal{I}_4$, which is a special case of type IIB on K3. Thus we get the duality:

$$(\text{M-theory on } T^5/Z_2) \leftrightarrow (\text{IIB on K3}).$$

[27] The duality between $E_8 \times E_8$ heterotic string theory and M-theory on S^1 can be "derived" from other known duality conjectures by taking the infinite radius limit of a lower dimensional duality relation [75].

This duality was first conjectured in [96, 97].

As in the case of type IIA string theory, we can get non-perturbative enhancement of gauge symmetries in M-theory when the compact manifold develops singularities [19, 134, 135]. M-theory contains classical membrane and five-brane soliton solutions carrying electric and magnetic charges of $C^{(\mathrm{S})}_{\mathrm{MNP}}$ respectively [8]. The extra massless states required for this symmetry enhancement come from membranes wrapped around the collapsed two cycles of the singular manifold.

References

[1] A. Sen, *Int. J. Mod. Phys. A* **9** (1994) 3707 [hep-th/9402002].

[2] M. Duff, R. Khuri and J. Lu, *Phys. Rep.* **259** (1995) 213 [hep-th/9412184].

[3] J. Schwarz, *Nucl. Phys. Proc. Suppl. B* **55** (1) [hep-th/9607201].

[4] S. Chaudhuri, C. Johnson and J. Polchinski [hep-th/9602052].

[5] J. Polchinski, *Rev. Mod. Phys.* **68** (1996) 1245 [hep-th/9607050].

[6] J. Polchinski [hep-th/9611050].

[7] P. Townsend [hep-th/9612121; gr-qc/9707012; hep-th/9712004].

[8] M. Duff [hep-th/9611203].

[9] M. Douglas [hep-th/9610041].

[10] W. Lerche [hep-th/9611190; hep-th/9710246].

[11] S. Forste and J. Louis [hep-th/9612192].

[12] C. Vafa [hep-th/9702201].

[13] A. Klemm [hep-th/9705131].

[14] E. Kiritsis [hep-th/9708130].

[15] B. de Wit and J. Louis [hep-th/9801132].

[16] M. Trigiante [hep-th/9801144].

[17] S. Mukhi [hep-ph/9710470].

[18] M. Green, J. Schwarz and E. Witten, Superstring Theory, Vols. 1 and 2 (Cambridge University Press, 1986); edited by D. Lust and S. Theisen, Lectures on String Theory (Springer, 1989); J. Polchinski [hep-th/9411028].

[19] E. Witten, *Nucl. Phys. B* **443** (1995) 85 [hep-th/9503124].

[20] A. Dabholkar, *Phys. Lett. B* **357** (1995) 307 [hep-th/9506160].

[21] C. Hull, *Phys. Lett. B* **357** (1995) 545 [hep-th/9506194].

[22] J. Polchinski and E. Witten, *Nucl. Phys. B* **460** (1996) 525 [hep-th/9510169].

[23] C. Hull and P. Townsend, *Nucl. Phys. B* **438** (1995) 109 [hep-th/9410167].

[24] M. Duff, *Nucl. Phys. B* **442** (1995) 47 [hep-th/9501030]. M. Duff and R. Khuri, *Nucl. Phys. B* **411** (1994) 473 [hep-th/9305142].

[25] A. Sen, *Nucl. Phys. B* **450** (1995) 103 [hep-th/9504027].

[26] J. Harvey and A. Strominger, *Nucl. Phys. B* **449** (1995) 535 [hep-th/9504047].

[27] A. Giveon, M. Porrati and E. Rabinovici, *Phys. Rep.* **244** (1994) 77 [hep-th/9401139].

[28] P. Fre, *Nucl. Phys. B* **[Proc. Sup.] 45B,C** (1996) 59 [hep-th/9512043] and references therein.

[29] E. Witten and D. Olive, *Phys. Lett. B* **78** (1978) 97.

[30] M. de Roo, *Nucl. Phys. B* **255** (1985) 515.

[31] S. Ferrara, C. Kounnas and M. Porrati, *Phys. Lett. B* **181** (1986) 263.

[32] M. Terentev, *Sov. J. Nucl. Phys.* **49** (1989) 713.

[33] S.F. Hassan and A. Sen, *Nucl. Phys. B* **375** (1992) 103 [hep-th/9109038].

[34] J. Maharana and J. Schwarz, *Nucl. Phys. B* **390** (1993) 3 [hep-th/9207016].

[35] A. Shapere, S. Trivedi and F. Wilczek, *Mod. Phys. Lett. A* **6** (1991) 2677.

[36] A. Sen, *Nucl. Phys. B* **404** (1993) 109 [hep-th/9207053].

[37] A. Font, L. Ibanez, D. Lust and F. Quevedo, *Phys. Lett. B* **249** (1990) 35.

[38] S.J. Rey, *Phys. Rev. D* **43** (1991) 526.

[39] J. Schwarz [hep-th/9209125].

[40] A. Sen, *Phys. Lett. B* **303** (1993) 22 [hep-th/9209016].

[41] A. Sen, *Mod. Phys. Lett. A* **8** (1993) 2023 [hep-th/9303057].

[42] J. Schwarz and A. Sen, *Nucl. Phys. B* **411** (1994) 35 [hep-th/9304154]; *Phys. Lett. B* **312** (1993) 105 [hep-th/9305185].

[43] N. Seiberg, *Nucl. Phys. B* **303** (1988) 286.

[44] M. Green and J. Schwarz, *Phys. Lett. B* **122** (1983) 143; J. Schwarz and P. West, *Phys. Lett. B* **126** (1983) 301; J. Schwarz, *Nucl. Phys. B* **226** (1983) 269; P. Howe and P. West, *Nucl. Phys. B* **238** (1984) 181.

[45] E. Cremmer, in "Unification of Fundamental Particle Interactions" (Plenum, 1980); B. Julia, in "Superspace and Supergravity" (Cambridge Univ. Press, 1981).

[46] S. Ferrara and C. Kounnas, *Nucl. Phys. B* **328** (1989) 406.

[47] S. Chaudhuri, G. Hockney and J. Lykken, *Phys. Rev. Lett.* **75** (1995) 2264 [hep-th/9505054].

[48] S. Chaudhuri and J. Polchinski, *Phys. Rev. D* **52** (1995) 7168 [hep-th/9506048].

[49] C. Montonen and D. Olive, *Phys. Lett. B* **72** (1977) 117.

[50] H. Osborn, *Phys. Lett. B* **83** (1979) 321.

[51] A. Sen, *Phys. Lett. B* **329** (1994) 217 [hep-th/9402032].

[52] M. Atiyah and N. Hitchin, *Phys. Lett. A* **107** (1985) 21; *Phil. Trans. Roy. Soc. Lond. A* **315** (1985) 459; Geometry and Dynamics of Magnetic Monopoles (Cambridge Univ. Press).

[53] G. Gibbons and N. Manton, *Nucl. Phys. B* **274** (1986) 183.

[54] J. Gauntlett, *Nucl. Phys. B* **400** (1993) 103 [hep-th/9205008]; *Nucl. Phys. B* **411** (1994) 443 [hep-th/9305068].

[55] J. Blum, *Phys. Lett. B* **333** (1994) 92 [hep-th/9401133].

[56] N. Manton and B. Schroers, *Annals Phys.* **225** (1993) 290.

[57] G. Gibbons and P. Ruback, *Comm. Math. Phys.* **115** (1988) 267.

[58] G. Segal and A. Selby, *Comm. Math. Phys.* **177** (1996) 775.

[59] M. Porrati, *Phys. Lett. B* **377** (1996) 67 [hep-th/9505187].

[60] J. Gauntlett and J. Harvey [hep-th/9407111].

[61] J. Schwarz, *Phys. Lett. B* **360** (1995) 13 [hep-th/9508143].

[62] E. Witten, *Nucl. Phys. B* **460** (1996) 335 [hep-th/9510135].

[63] S. Das and S. Mathur, *Phys. Lett. B* **375** (1996) 103 [hep-th/9601152].

[64] A. Dabholkar and J. Harvey, *Phys. Rev. Lett.* **63** (1989) 478; A. Dabholkar, G. Gibbons, J. Harvey and F. Ruiz, *Nucl. Phys. B* **340** (1990) 33.

[65] J. Polchinski, *Phys. Rev. Lett.* **75** (1995) 4724 [hep-th/9510017].

[66] J. Dai, R. Leigh and J. Polchinski, *Mod. Phys. Lett. A* **4** (1989) 2073; R. Leigh, *Mod. Phys. Lett. A* **4** (1989) 2767; J. Polchinski, *Phys. Rev. D* **50** (1994) 6041 [hep-th/9407031].

[67] M. Bershadsky, V. Sadov and C. Vafa, *Nucl. Phys. B* **463** (1996) 398 [hep-th/9510225].

[68] A. Sen, *Phys. Rev. D* **53** (1996) 2874 [hep-th/9711026].

[69] C. Vafa, *Nucl. Phys. B* **469** (1996) 415 [hep-th/9511088]; *Nucl. Phys. B* **463** (1996) 435 [hep-th/9512078].

[70] S. Sethi and M. Stern, *Phys. Lett. B* **398** (1997) 47 [hep-th/9607145].

[71] M. Bershadsky, V. Sadov and C. Vafa, *Nucl. Phys. B* **463** (1996) 420 [hep-th/9511222].

[72] E. Witten, *Nucl. Phys. B* **460** (1996) 541 [hep-th/9511030].

[73] M. Porrati, *Phys. Lett. B* **387** (1996) 492 [hep-th/9607082].

[74] J. Blum, *Nucl. Phys. B* **506** (1997) 223 [hep-th/9705030].

[75] A. Sen [hep-th/9609176].

[76] A. Sen, *Nucl. Phys. B* **434** (1995) 179 [hep-th/9408083]; *Nucl. Phys. B* **447** (1995) 62 [hep-th/9503057].

[77] C. Vafa and E. Witten [hep-th/9507050].

[78] A. Sen, *Nucl. Phys. B* **474** (1996) 361 [hep-th/9604070].

[79] S. Kachru and C. Vafa, *Nucl. Phys. B* **450** ((1995) 69 [hep-th/9605105].

[80] S. Ferrara, J. Harvey, A. Strominger and C. Vafa, *Phys. Lett. B* **361** (1995) 59 [hep-th/9505162].

[81] A. Klemm, W. Lerche and P. Mayr, *Phys. Lett. B* **357** (1995) 313 [hep-th/9506112].

[82] G. Aldazabal, A. Font, L. Ibanez and F. Quevedo, *Nucl. Phys. B* **461** (1996) 537 [hep-th/9510093].

[83] B. Hunt and R. Schimmrigk, *Phys. Lett. B* **381** (1996) 427 [hep-th/9512138]; B. Hunt, M. Lynker and R. Schimmtigk [hep-th/9609082].

[84] B. de Wit, P. Lauwers and A. van Proeyen, *Nucl. Phys. B* **255** (1985) 569.

[85] V. Kaplunovsky, J. Louis and S. Theisen, *Phys. Lett. B* **357** (1995) 71 [hep-th/9506110].

[86] I. Antoniadis, E. Gava, K. Narain and T. Taylor, *Nucl. Phys. B* **455** (1995) 109 [hep-th/9507115].

[87] S. Kachru, A. Klemm, W. Lerche, P. Mayr and C. Vafa, *Nucl. Phys. B* **459** (1996) 537 [hep-th/9508155]; A. Klemm, W. Lerche, P. Mayr, C. Vafa and N. Warner, *Nucl. Phys. B* **477** (1996) 746 [hep-th/9604034].

[88] N. Seiberg and E. Witten, *Nucl. Phys. B* **426** (1994) 19 [hep-th/9407087]; *Nucl. Phys. B* **431** (1994) 484 [hep-th/9408099].

[89] P. Townsend, *Phys. Lett. B* **350** (1995) 184 [hep-th/9501068].

[90] F. Giani and M. Pernici, *Phys. Rev. D* **30** (1984) 325; I. Campbell and P. West, *Nucl. Phys. B* **243** (1984) 112; M. Huq and M. Namazie, *Class. Quant. Grav.* **2** (1985) 293.

[91] S. Sethi and M. Stern [hep-th/9705046].

[92] A. Sen, *Phys. Rev. D* **54** (1996) 2964 [hep-th/9510229].

[93] J. Schwarz, *Phys. Lett. B* **367** (1996) 97 [hep-th/9510086].

[94] P. Aspinwall, *Nucl. Phys. Proc. Suppl.* **46** (1996) 30 [hep-th/9508154].

[95] P. Horava and E. Witten, *Nucl. Phys. B* **460** (1996) 506 [hep-th/9510209]; *Nucl. Phys. B* **475** (1996) 94 [hep-th/9603142].

[96] K. Dasgupta and S. Mukhi, *Nucl. Phys. B* **465** (1996) 399 [hep-th/9512196].

[97] E. Witten, *Nucl. Phys. B* **463** (1996) 383 [hep-th/9512219].

[98] A. Sen, Mod. *Phys. Lett. A* **11** (1996) 1339 [hep-th/9603113].

[99] E. Caceres, V. Kaplunovsky and M. Mandelberg, *Nucl. Phys. B* **493** (1997) 73 [hep-th/9606036].

[100] A. Sagnotti, *Open Strings and their Symmetry Groups* (Talk at Cargese Summer Inst., 1987); G. Pradisi and A. Sagnotti, *Phys. Lett. B* **216** (1989) 59; M. Bianchi, G. Pradisi and A. Sagnotti, *Nucl. Phys. B* **376** (1992) 365; P. Horava, *Nucl. Phys. B* **327** (1989) 461; *Phys. Lett. B* **231** (1989) 251.

[101] E. Gimon and J. Polchinski, *Phys. Rev. D* **54** (1996) 1667 [hep-th/9601038].

[102] A. Chamseddine, *Phys. Rev. D* **24** (1981) 3065; E. Bergshoeff, M. de Roo, B. de Wit and P. van Niewenhuizen, *Nucl. Phys. B* **195** (1982) 97; E. Bergshoeff, M. de Roo and B. de Wit, *Nucl. Phys. B* **217** (1983) 143; G. Chapline and N. Manton, *Phys. Lett. B* **120** (1983) 105.

[103] H. Nicolai, *Phys. Lett. B* **276** (1992) 333.

[104] E. Weinberg, *Phys. Rev. D* **20** (1979) 936; E. Corrigan and P. Goddard, *Comm. Math. Phys.* **80** (1981) 575; C. Taubes, *Comm. Math. Phys.* **91** (1983) 235.

[105] N. Manton, *Phys. Lett. B* **110** (1982) 54.

[106] C. Callias, *Comm. Math. Phys.* **62** (1978) 213.

[107] K. Narain, *Phys. Lett. B* **169** (1986) 41; K. Narain, H. Sarmadi and E. Witten, *Nucl. Phys. B* **279** (1987) 369.

[108] T. Banks, L. Dixon, D. Friedan and E. Martinec, *Nucl. Phys. B* **299** (1988) 613; T. Banks and L. Dixon, *Nucl. Phys. B* **307** (1988) 93.

[109] B. Greene [hep-th/9702155] and references therein.

[110] B. de Wit and A. van Proeyen, *Nucl. Phys. B* **245** (1984) 89; E. Cremmer, C. Kounnas, A. van Proeyen, J. Derendinger, S. Ferrara, B. de Wit and L. Girardello, *Nucl. Phys. B* **250** (1985) 385; S. Ferrara and A. Strominger, in "Strings 89", World Scientific (1989); A. Strominger, *Comm. Math. Phys.* **133** (1990) 163.

[111] J. Bagger and E. Witten, *Nucl. Phys. B* **222** (1983) 1; K. Galicki, *Comm. Math. Phys.* **108** (1987) 117.

[112] E. Cremmer and B. Julia, *Nucl. Phys. B* **159** (1979) 141.

[113] M. Gaillard and B. Zumino, *Nucl. Phys. B* **193** (1981) 221; S. Ferrara, J. Scherk and B. Zumino, *Nucl. Phys. B* **121** (1977) 393; E. Cremmer, J. Scherk and S. Ferrara, *Phys. Lett. B* **74** (1978) 61; B. de Wit, *Nucl. Phys. B* **158** (1979) 189; B. de Wit and H. Nicolai, *Nucl. Phys. B* **208** (1982) 323; E. Cremmer and B. Julia, *Phys. Lett. B* **80** (1978) 48; B. de Wit and A. van Proeyen, *Nucl. Phys. B* **245** (1984) 89.

[114] D. Gross, J. Harvey, E. Martinec and R. Rohm, *Phys. Rev. Lett.* **54** (1985) 502; *Nucl. Phys. B* **256** (1985) 253; *Nucl. Phys. B* **267** (1986) 75.

[115] M. Green and J. Schwarz, *Phys. Lett. B* **109** (1982) 444; *Phys. Lett. B* **149** (1984) 117; *Phys. Lett. B* **151** (1985) 21; *Nucl. Phys. B* **255** (1985) 93.

[116] C. Hull and P. Townsend, *Nucl. Phys. B* **451** (1995) 525 [hep-th/9505073].

[117] A. Strominger, *Nucl. Phys. B* **451** (1995) 96 [hep-th/9504090]; B. Greene, D. Morrison and A. Strominger, *Nucl. Phys. B* **451** (1995) 109 [hep-th/9504145].

[118] M. Douglas and G. Moore [hep-th/9603167]; J. Polchinski, *Phys. Rev. B* **55** (1997) 6423 [hep-th/9606165]; C. Johnson and R. Myers, *Phys. Rev. D* **55** (1997) 6382 [hep-th/9610140]; M. Douglas [hep-th/9612126] published in JHEP electronics journal; D. Diaconescu, M. Douglas and J. Gomis [hep-th/9712230].

[119] E. Witten [hep-th/9507121].

[120] M. Douglas [hep-th/9512077; hep-th/9604198].

[121] L. Alvarez-Gaume, D. Freedman and S. Mukhi, *Ann. Phys.* (NY) **134** (1981) 85; L. Alvarez-Gaume and D. Freedman, *Comm. Math. Phys.* **80** (1981) 443.

[122] E. Bergshoeff, E. Sezgin and P. Townsend, *Phys. Lett. B* **189** (1987) 75; *Ann. Phys.* (NY) **185** (1988) 330; M. Duff, *Class. Quant. Grav.* **5** (1988) 189; B. de Wit, M. Luscher and H. Nicolai, *Nucl. Phys. B* **320** (1989) 135; B. de Wit, J. Hoppe and H. Nicolai, *Nucl. Phys. B* **305** (1988) 545.

[123] P. Ramond, *Phys. Rev. D* **3** (1971) 2415.

[124] A. Neveu and J. Schwarz, *Nucl. Phys. B* **31** (1971) 86; *Phys. Rev. D* **4** (1971) 1109.

[125] F. Gliozzi, J. Scherk and D. Olive, *Phys. Lett. B* **65** (1976) 282; *Nucl. Phys. B* **122** (1977) 253.

[126] P. Aspinwall and D. Morrison [hep-th/9404151]; P. Aspinwall [hep-th/9611137] and references therein.

[127] M. Duff and R. Minasian, *Nucl. Phys. B* **436** (1995) 507 [hep-th/9406198]; C. Vafa and E. Witten, *Nucl. Phys. B* **447** (1995) 261 [hep-th/9505053]; M. Duff, J. Liu and R. Minasian, *Nucl. Phys. B* **452** (1995) 261 [hep-th/9506126]; M. Duff, R. Minasian and E. Witten, *Nucl. Phys. B* **463** (1996) 435 [hep-th/9601036]; M. Berkooz, R. Leigh, J. Polchinski, J. Schwarz, N. Seiberg and E. Witten, *Nucl. Phys. B* **475** (1996) 115 [hep-th/9605184].

[128] P. Candelas, G. Horowitz, A. Strominger and E. Witten, *Nucl. Phys. B* **258** (1985) 46.

[129] E. Gimon and C. Johnson, *Nucl. Phys. B* **479** (1996) 285 [hep-th/9606176]; J. Blum and A. Zaffaroni, *Phys. Lett. B* **387** (1996) 71 [hep-th/9607019]; A. Dabholkar and J. Park, *Phys. Lett. B* **394** (1997) 302 [hep-th/9607041]; R. Gopakumar and S. Mukhi, *Nucl. Phys. B* **479** (1996) 260 [hep-th/9607057]; J. Park [hep-th/9611119]; A. Sen, *Nucl. Phys. B* **489** (1997) 139 [hep-th/9611186]; *Nucl. Phys. B* **498** (1997) 135 [hep-th/9702061]; *Phys. Rev. D* **55** (1997) 7345 [hep-th/9702165]; O. Aharony, J. Sonneschein, S. Yankielowicz and S. Theisen, *Nucl. Phys. B* **493** (1997) 177 [hep-th/9611222].

[130] M. Bershadsky, K. Intrilligator, S. Kachru, D. Morrison, V. Sadov and C. Vafa, *Nucl. Phys. B* **481** (1996) 215 [hep-th/9605200]; P. Aspinwall and M. Gross, *Phys. Lett. B* **387** (1996) 735 [hep-th/9605131].

[131] K. Dasgupta and S. Mukhi, *Phys. Lett. B* **385** (1996) 125 [hep-th/9606044].

[132] A. Johansen, *Phys. Lett. B* **395** (1997) 36 [hep-th/9608186].

[133] M. Gaberdiel and B. Zwiebach [hep-th/9709013]; M. Gaberdiel, T. Hauer and B. Zwiebach [hep-th/9801205].

[134] E. Witten, *Nucl. Phys. B* **500** (1997) 3 [hep-th/9703166].

[135] A. Sen [hep-th/9707123] published in JHEP electronic journal.

[136] E. Witten, *Nucl. Phys. B* **471** (1996) 135 [hep-th/9602070].

[137] A. Cadavid, A. Ceresole, R. D'Auria and S. Ferrara [hep-th/9506144]; I. Antoniadis, S. Ferrara and T. Taylor, *Nucl. Phys. B* **460** (1996) 489; S. Ferrara, R. Khuri and R. Minasian [hep-th/9602102]; E. Witten, *Nucl. Phys. B* **471** (1996) 195 [hep-th/9603150].

[138] M. Duff, B. Nilsson and C. Pope, *Phys. Rep.* **130** (1986) 1.

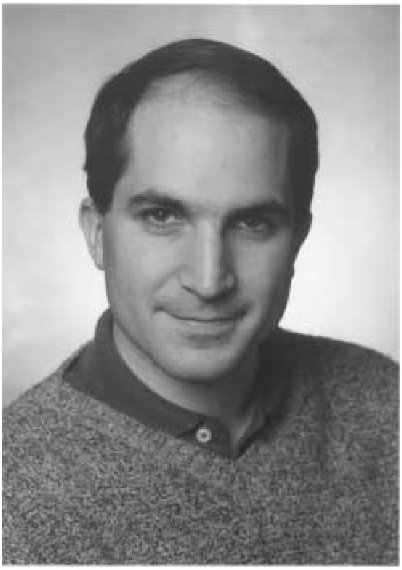
[139] J. Cardy, *Nucl. Phys. B* **270** (1986) 186.

LECTURE 4

LES HOUCHES LECTURES ON LARGE N FIELD THEORIES AND GRAVITY

J. MALDACENA

Institute for Advanced Study,
Princeton, NJ, U.S.A.



Contents

LES HOUCHES LECTURES ON LARGE N FIELD THEORIES AND GRAVITY

J. Maldacena

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 Aharony *et al.* [1].

1 General introduction

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

Even though though string theory is normally used as a theory of quantum gravity, it is not how string theory was originally discovered. String theory was discovered in an attempt to describe the large number of mesons and hadrons that were experimentally discovered in the 1960's. The idea was to view all these particles as different oscillation modes of a string. The string idea described well some features of the hadron spectrum. For example, the mass of the lightest hadron with a given spin obeys a relation like $m^2 \sim TJ^2 + \text{const}$. 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.

This work is supported by the Packard foundation and the DOE grant DE-FGO2-91ER40654.

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

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

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

The relationship we described between gauge theories and string theory on Anti-de-Sitter spaces was motivated by studies of D-branes and black holes in strings theory. D-branes are solitons in string theory [10]. They come in various dimensionalities. If they have zero spatial dimensions they are like ordinary localized, particle-type soliton solutions, analogous to the 't Hooft-Polyakov [11,12] monopole in gauge theories. These are called D-zero-branes. If they have one extended dimension they are called D-one-branes or D-strings. They are much heavier than ordinary fundamental strings when the string coupling is small. In fact, the tension of all D-branes is proportional to $1/g_{\mathrm{s}}$, 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 [14–16]. 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 [20–22]. 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 [1, 32–35].

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 $1/l_{\mathrm{s}}$, 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

$$S = S_{\mathrm{bulk}} + S_{\mathrm{brane}} + S_{\mathrm{int}}. \tag{2.1}$$

S_{bulk} is the action of ten dimensional supergravity, plus some higher derivative corrections. Note that the Lagrangian (2.1) 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, we integrate out all massive degrees of freedom but we do not integrate out the massless ones. The brane action $S_{\rm brane}$ 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 $\alpha'^2 {\rm Tr}(F^4)$. Finally, $S_{\rm int}$ 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

$$S_{\rm bulk} \sim \frac{1}{2\kappa^2}\int \sqrt{g}\mathcal{R} \sim \int (\partial h)^2 + \kappa(\partial h)^2 h + \cdots , \tag{2.2}$$

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

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

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

$$\begin{aligned} ds^2 &= f^{-1/2}(-dt^2 + dx_1^2 + dx_2^2 + dx_3^2) + f^{1/2}(dr^2 + r^2 d\Omega_5^2) \ , \\ F_5 &= (1 + *) dt dx_1 dx_2 dx_3 df^{-1} \ , \\ f &= 1 + \frac{R^4}{r^4} \ , \qquad R^4 \equiv 4\pi g_s \alpha'^2 N \ . \end{aligned} \tag{2.3}$$

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

$$E = f^{-1/4} E_p. \tag{2.4}$$

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

$$ds^2 = \frac{r^2}{R^2}(-dt^2 + dx_1^2 + dx_2^2 + dx_3^2) + R^2 \frac{dr^2}{r^2} + R^2 d\Omega_5^2, \tag{2.5}$$

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

We see that both from the point of view of a field theory of open strings living on the brane, and from the point of view of the supergravity description, we have two decoupled theories in the low-energy limit. In both cases one of the decoupled systems is supergravity in flat space. So, it is natural to identify the second system which appears in both descriptions. Thus, we are led to the conjecture that $\mathcal{N} = 4$ $U(N)$

$3+1$ [illegible] ([illegible]) [illegible] $AdS_5 \times S^5$ [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 $\sqrt{\alpha'}E_{\rm p} \sim$ fixed. For small α' (2.4) reduces to $E \sim E_{\rm p} r/\sqrt{\alpha'}$. Since we want to keep fixed the energy measured from infinity, which is the way energies are measured in the field theory, we need to take $r \to 0$ keeping r/α' fixed. It is then convenient to define a new variable $U \equiv r/\alpha'$, so that the metric becomes

$$\mathrm{d}s^2 = \alpha' \left[\frac{U^2}{\sqrt{4\pi g_{\rm s} N}} (-\mathrm{d}t^2 + \mathrm{d}x_1^2 + \mathrm{d}x_2^2 + \mathrm{d}x_3^2) + \sqrt{4\pi g_{\rm s} N} \left(\frac{\mathrm{d}U^2}{U^2} + \mathrm{d}\Omega_5^2 \right) \right]. \tag{2.6}$$

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

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

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 (2.3) 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 $SU(4)_{\rm R}$ R-symmetry group of the field theory. In fact, the whole supergroup is the same for the $\mathcal{N}=4$ field theory and the $AdS_5 \times S^5$ geometry, so both sides of the conjecture have the same spacetime symmetries. We will discuss in more detail the matching between the two sides of the correspondence in Section 3.

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

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

$$\int_{S^5} F_5 = N. \tag{2.7}$$

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

$$\tau \equiv \frac{4\pi i}{g_{\rm YM}^2} + \frac{\theta}{2\pi} = \frac{i}{g_{\rm s}} + \frac{\chi}{2\pi} \ , \tag{2.8}$$

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

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

$$g_{\mathrm{YM}}^2 N \sim g_\mathrm{s} N \sim \frac{R^4}{l_\mathrm{s}^4} \ll 1. \tag{2.9}$$

Note that we need $g_{\mathrm{YM}}^2 N$ 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,

$$\frac{R^4}{l_\mathrm{s}^4} \sim g_\mathrm{s} N \sim g_{\mathrm{YM}}^2 N \gg 1. \tag{2.10}$$

We see that the gravity regime (2.10) and the perturbative field theory regime (2.9) 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 [illegible]. 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 (2.9, 2.10) we implicitly assumed that $g_\mathrm{s} < 1$. If $g_\mathrm{s} > 1$ we can perform an $SL(2,\mathbb{Z})$

duality transformation and get conditions similar to (2.9, 2.10) but with $g_\mathrm{s} \to 1/g_\mathrm{s}$. So, we cannot get into the gravity regime (2.10) by taking N small ($N = 1, 2, ...$) and g_s very large, since in that case the D-string becomes light and renders the gravity approximation invalid. Another way to see this is to note that the radius of curvature in Planck units is $R^4/l_\mathrm{p}^4 \sim N$. So, it is always necessary, but not sufficient, to have large N in order to have a weakly coupled supergravity description.

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

2.1 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 (2.8) to changing the coupling constant in the string theory, which is then related to the expectation value of the dilaton. The expectation value of the dilaton is set by the boundary condition for the dilaton at infinity. So, changing the gauge theory coupling constant corresponds to changing the boundary value of the dilaton. More precisely, let us denote by $\mathcal{O}$ the corresponding operator. We can consider adding the term $\int \mathrm{d}^4x \phi_0(\vec{x})\mathcal{O}(\vec{x})$ to the Lagrangian (for simplicity we assume that such a term was not present in the original Lagrangian, otherwise we consider $\phi_0(\vec{x})$ to be the total coefficient of $\mathcal{O}(\vec{x})$ in the Lagrangian). According to the discussion above, it is natural to assume that this will change the boundary condition of the dilaton at the boundary of AdS to $\phi(\vec{x}, z)|_{z=0} = \phi_0(\vec{x})$, in the coordinate system

$$\mathrm{d}s^2 = R^2_{AdS} \frac{-\mathrm{d}t^2 + \mathrm{d}x_1^2 + \cdots + \mathrm{d}x_3^3 + \mathrm{d}z^2}{z^2}.$$

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

$$\left\langle \mathrm{e}^{\int \mathrm{d}^4 x \phi_0(\vec{x}) \mathcal{O}(\vec{x})} \right\rangle_{\mathrm{CFT}} = \mathcal{Z}_{\mathrm{string}} \left[\phi(\vec{x}, z) \Big|_{z=0} = \phi_0(\vec{x}) \right], \tag{2.11}$$

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

A formula like (2.11) is valid in general, for any field ϕ. Therefore, each field propagating on AdS space is in a one to one correspondence with an operator in the field theory. There is a relation between the mass of the field ϕ and the scaling dimension of the operator in the conformal field theory. Let us describe this more generally in AdS_{d+1}. The wave equation in Euclidean space for a field of mass m has two independent solutions, which behave like $z^{d-\Delta}$ and z^{Δ} for small z (close to the boundary of AdS), where

$$\Delta = \frac{\mathrm{d}}{2} + \sqrt{\frac{\mathrm{d}^2}{4} + R^2 m^2}. \tag{2.12}$$

Therefore, in order to get consistent behavior for a massive field, the boundary condition on the field in the right hand side of (2.11) should in general be changed to

$$\phi(\vec{x}, \epsilon) = \epsilon^{d-\Delta} \phi_0(\vec{x}), \tag{2.13}$$

and eventually we would take the limit where $\epsilon \to 0$. Since ϕ is dimensionless, we see that ϕ_0 has dimensions of $[\mathrm{length}]^{\Delta - d}$ which implies, through the left hand side of (2.11), that the associated operator $\mathcal{O}$ has dimension Δ (2.12). A more detailed derivation of this relation will be given in Section 4, where we will verify that the two-point correlation function of the operator $\mathcal{O}$ behaves as that of an operator of dimension Δ [23,24]. A similar relation between fields on AdS and operators in the field theory exists also for non-scalar fields, including fermions and tensors on AdS space.

Correlation functions in the gauge theory can be computed from (2.11) by differentiating with respect to ϕ_0. Each differentiation brings down an insertion $\mathcal{O}$, which sends a ϕ particle (a closed string state) into the bulk. Feynman diagrams can be used to compute the interactions of particles in the bulk. In the limit where classical supergravity is applicable, the only diagrams that contribute are the tree-level diagrams of the gravity theory (see for instance Fig. 1).

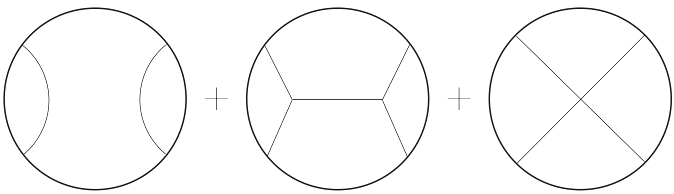


Fig. 1. Correlation functions can be calculated (in the large g_sN limit) in terms of supergravity Feynman diagrams. Here we see the leading contribution coming from a disconnected diagram plus connected pieces involving interactions of the supergravity fields in the bulk of AdS. At tree level, these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

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

2.2 Holography

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

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

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

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

$$\mathrm{d}s^2 = R^2\left[-\left(\frac{1+r^2}{1-r^2}\right)^2\mathrm{d}t^2 + \frac{4}{(1-r^2)^2}(\mathrm{d}r^2 + r^2\mathrm{d}\Omega^2)\right]. \tag{2.14}$$

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

$$\mathrm{d}s^2 = R^2\frac{-\mathrm{d}t^2 + \mathrm{d}\vec{x}^2 + \mathrm{d}z^2}{z^2}, \tag{2.15}$$

in which the boundary is at $z = 0$. If we consider a particle or wave propagating in (2.15) or (2.14) we see that its motion is independent of R in the supergravity approximation. Furthermore, if we are in Euclidean space and we have a wave that has some spatial extent λ in the $\vec{x}$ directions, it will also have an extent λ in the z direction. This can be seen from (2.15) by eliminating λ through the change of variables $x \to \lambda x$, $z \to \lambda z$. This implies that a cutoff at

$$z \sim \delta \tag{2.16}$$

corresponds to a UV cutoff in the field theory at distances δ, with no factors of R (δ here is dimensionless, in the field theory it is measured in terms of the radius of the S^4 or S^3 that the theory lives on). Equation (2.16) is called the UV-IR relation [66].

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

$$S \sim N^2 \delta^{-3}, \tag{2.17}$$

since the number of cells into which we divide the three-sphere is of order $1/\delta^3$. In the gravity solution (2.14) the area in Planck units of the surface at $r = 1 - \delta$, for $\delta \ll 1$, is

$$\frac{\text{Area}}{4G_N} = \frac{V_{S^5} R^3 \delta^{-3}}{4G_N} \sim N^2 \delta^{-3}. \tag{2.18}$$

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

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

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

- some correlation functions, which are usually related to anomalies, are protected from any quantum corrections and do not depend on λ. The matching of these correlation functions will be described in Section 3.2 below;
- the spectrum of chiral operators does not change as the coupling varies, and it will be compared in Section 3.1 below;
- the moduli space of the theory also does not depend on the coupling. In the $SU(N)$ field theory the moduli space is $\mathbb{R}^{6(N-1)}/S_N$, parametrized by the eigenvalues of six commuting traceless $N \times N$ matrices. On the AdS side it is not clear exactly how to define the moduli space. There are however multicenter solutions so that one might think that 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, 61, 69].

3.1 *The spectrum of chiral primary operators*

3.1.1 The field theory spectrum

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

$$\begin{aligned} \{Q^A_\alpha, \bar{Q}_{\dot{\alpha}B}\} &= 2(\sigma^\mu)_{\alpha\dot{\alpha}} P_\mu \delta^A_B, \\ \{Q^A_\alpha, Q^B_\beta\} &= \{\bar{Q}_{\dot{\alpha}A}, \bar{Q}_{\dot{\beta}B}\} = 0, \end{aligned} \tag{3.1}$$

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

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

$$\begin{aligned}
[Q_\alpha^A, \phi^I] &\sim \lambda_{\alpha B}, \\
\{Q_\alpha^A, \lambda_{\beta B}\} &\sim (\sigma^{\mu\nu})_{\alpha\beta} F_{\mu\nu} + \epsilon_{\alpha\beta}[\phi^I, \phi^J], \\
\{Q_\alpha^A, \bar{\lambda}_{\dot\beta}^B\} &\sim (\sigma^\mu)_{\alpha\dot\beta} \mathcal{D}_\mu \phi^I, \\
[Q_\alpha^A, A_\mu] &\sim (\sigma_\mu)_{\alpha\dot\alpha} \bar{\lambda}_{\dot\beta}^A \epsilon^{\dot\alpha\dot\beta},
\end{aligned} \tag{3.2}$$

with similar expressions for the action of the $\bar{Q}$'s, where $\sigma^{\mu\nu}$ are the generators of the Lorentz group in the spinor representation, $\mathcal{D}_\mu$ is the covariant derivative, the field strength $F_{\mu\nu} \equiv [\mathcal{D}_\mu, \mathcal{D}_\nu]$, and we have suppressed the $SU(4)$ Clebsch-Gordan coefficients corresponding to the products $\mathbf{4} \times \mathbf{6} \rightarrow \bar{\mathbf{4}}$, $\mathbf{4} \times \bar{\mathbf{4}} \rightarrow \mathbf{1} + \mathbf{15}$ and $\mathbf{4} \times \mathbf{4} \rightarrow \mathbf{6}$ in the first three lines of (3.2).

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

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

It is natural to classify the operators in a conformal theory into primary operators and their descendants. In a superconformal theory it is also natural to distinguish between chiral primary operators, which are in short representations of the superconformal algebra and are annihilated by some of the supercharges, and non-chiral primary operators. Representations of the superconformal algebra are formed by starting with some state of lowest dimension, which is annihilated by the operators S and K_μ, and acting on it with the operators Q and P_μ. The $\mathcal{N} = 4$ supersymmetry algebra involves 16 real supercharges. A generic primary representation of the superconformal algebra will thus include 2^{16} primaries of the conformal algebra, generated by acting on the lowest state with products of different supercharges; acting with additional supercharges always leads to descendants of the conformal algebra (derivatives). Since the supercharges have helicities $\pm 1/2$, the primary fields in such representations will have a range of helicities between $\lambda - 4$ (if the lowest dimension operator ψ has helicity λ) and $\lambda + 4$ (acting with more than 8 supercharges of the same helicity either annihilates the state or leads to a conformal descendant). In non-generic representations of the superconformal algebra a product of less than 16 different Q's annihilates the lowest dimension operator, and the range of helicities appearing is smaller. In particular, in the small representations of the $\mathcal{N} = 4$ superconformal algebra only up to 4 Q's of the same helicity acting on the lowest dimension operator give a non-zero result, and the range of helicities is between $\lambda - 2$ and $\lambda + 2$. For the $\mathcal{N} = 4$ supersymmetry algebra (not including the conformal algebra) it is known that medium representations, whose range of helicities is 6, can also exist (they arise, for instance, on the moduli space of the $SU(N)$ $\mathcal{N} = 4$ SYM theory [71–78]); 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 [79–85] and references therein.

In the $U(1)$ $\mathcal{N} = 4$ SYM theory (which is a free theory), the only gauge-invariant "single trace" operators are the fields of the vector multiplet itself (which are $\phi^I, \lambda_A, \bar{\lambda}^A$ and $F_{\mu\nu} = \partial_{[\mu} A_{\nu]}$). These operators form an ultra-short representation of the $\mathcal{N} = 4$ algebra whose range of helicities is from (-1) to 1 (acting with more than two supercharges of the same helicity on any of these states gives either zero or derivatives, which are descendants of the conformal algebra). All other local gauge invariant operators in the theory involve derivatives or products of these operators. This representation is usually called the doubleton representation, and it does not appear in the $SU(N)$ SYM theory (though the representations which do appear can all be formed by tensor products of the doubleton representation).

In the context of AdS space one can think of this multiplet as living purely on the boundary of the space [46,86–95], as expected for the $U(1)$ part of the original $U(N)$ gauge group of the D3-branes (see the discussion in Sect. 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 (3.2) 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 operators of the form $\mathcal{O}^{I_1 I_2 \cdots I_n} \equiv \mathrm{Tr}(\phi^{I_1}\phi^{I_2}\cdots\phi^{I_n})$. First we can ask if this operator can be written as $\{Q,\psi\}$ for any field ψ. In the SUSY algebra (3.2) only commutators of ϕ^I's appear on the right hand side, so we see that if some of the indices are antisymmetric the field will be a descendant. Thus, only symmetric combinations of the indices will be lowest components of primary multiplets. Next, we should ask if the multiplet built on such an operator is a (short) chiral primary multiplet or not. There are several different ways to answer this question. One possibility is to use the relation between the dimension of chiral primary operators and their R-symmetry representation [96–100], and to check if this relation is obeyed in the free field theory, where $[\mathcal{O}^{I_1 I_2 \cdots I_n}] = n$. In this way we find that the representation is chiral primary if and only if the indices form a symmetric traceless product of n $\mathbf{6}$'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 $SU(4)_{\mathrm{R}}$; in this section we will refer to $SU(4)_{\mathrm{R}}$ 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 (3.2) it is easy to see that for symmetric traceless representations we get "null vectors" while for other representations we do not. For instance, let us analyze in detail the case $n=2$. The symmetric product of two $\mathbf{6}$'s is given by $\mathbf{6}\times\mathbf{6}\rightarrow\mathbf{1}+\mathbf{20}'$. The field in the $\mathbf{1}$ representation is $\mathrm{Tr}(\phi^I\phi^I)$, for which $[Q^A_\alpha, \mathrm{Tr}(\phi^I\phi^I)] \sim C^{AJB}\mathrm{Tr}(\lambda_{\alpha B}\phi^J)$ where C^{AIB} is a Clebsch-Gordan coefficient for $\bar{\mathbf{4}}\times\mathbf{6}\rightarrow\mathbf{4}$. The right-hand side is in the $\mathbf{4}$ representation, which is the most general representation that can appear in the product $\mathbf{4} \times \mathbf{1}$, so we find no null vectors at this level. On the other hand, if we look at the symmetric traceless product $\mathrm{Tr}(\phi^{\{I}\phi^{J\}}) \equiv \mathrm{Tr}(\phi^I\phi^J) - \frac{1}{6}\delta^{IJ}\mathrm{Tr}(\phi^K\phi^K)$ in the $\mathbf{20'}$ representation, we find that $\{Q^A_\alpha, \mathrm{Tr}(\phi^{\{I}\phi^{J\}})\} \sim \mathrm{Tr}(\lambda_{\alpha B}\phi^K)$ with the right-hand side being in the $\mathbf{20}$ representation (appearing in $\bar{\mathbf{4}} \times \mathbf{6} \to \mathbf{4} + \mathbf{20}$), while the left-hand side could in principle be in the $\mathbf{4} \times \mathbf{20'} \to \mathbf{20} + \mathbf{60}$. Since the $\mathbf{60}$ does not appear on the right-hand side (it is a "null vector") we identify that the representation built on the $\mathbf{20'}$ is a short representation of the SUSY algebra. By similar manipulations (see [24,81,84,101] for more details) one can verify that chiral primary representations correspond exactly to symmetric traceless products of $\mathbf{6}$'s.

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

The last issue we should discuss is what is the range of values of n. The product of more than N commuting[2] $N \times N$ matrices can always be written as a sum of products of traces of less than N of the matrices, so it does not form an independent operator. This means that for $n > N$ we can express the operator $\mathcal{O}^{I_1I_2\cdots I_n}$ 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 $\mathcal{O}_n = \mathcal{O}^{\{I_1I_2\cdots I_n\}}$ with $n = 2, 3, \cdots, N$, for which the indices are in the symmetric traceless product of n $\mathbf{6}$'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

[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.

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" [81, 101]; see also [102] which checks some components of these superfields against the couplings to supergravity modes predicted on the basis of the DBI action for D3-branes in anti-de Sitter space [4].

It is easy to find the form of all the fields in such a multiplet by using the algebra (3.2). For example, let us analyze here in detail the bosonic primary fields of dimension $n+1$ in the multiplet. To get a field of dimension $n+1$ we need to act on $\mathcal{O}_n$ with two supercharges (recall that $[Q]=\frac{1}{2}$). If we act with two supercharges Q^A_α of the same chirality, their Lorentz indices can be either antisymmetrized or symmetrized. In the first case we get a Lorentz scalar field in the $(2,n-2,0)$ representation of $SU(4)_\mathrm{R}$, which is of the schematic form

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

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

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

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 $SU(4)_\mathrm{R}$, of the form

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

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

- a complex vector field in the $(1, n-4, 1)$ representation, given by $Q^3\bar{Q}\mathcal{O}_n$, of the form $\mathrm{Tr}(F_{\mu\nu}\mathcal{D}^\nu\phi^J\phi^{I_1}\cdots\phi^{I_{n-2}})+\cdots$;
- an complex anti-symmetric 2-form field in the $(2, n-3, 0)$ representation, given by $Q^2\bar{Q}^2\mathcal{O}_n$, of the form $\mathrm{Tr}(F_{\mu\nu}[\phi^{J_1}, \phi^{J_2}]\phi^{I_1}\cdots\phi^{I_{n-2}})+\cdots$;
- a symmetric tensor field in the $(0, n-2, 0)$ representation, given by $Q^2\bar{Q}^2\mathcal{O}_n$, of the form $\mathrm{Tr}(\mathcal{D}_{\{\mu}\phi^J\mathcal{D}_{\nu\}}\phi^K\phi^{I_1}\cdots\phi^{I_{n-2}})+\cdots$

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

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

3.1.2 The string theory spectrum and the matching

As discussed in Section 2.1, fields on AdS_5 are in a one-to-one correspondence with operators in the dual conformal field theory. Thus, the spectrum of operators described in Section 3.1.1 should agree with the spectrum of fields of type IIB string theory on $AdS_5 \times S^5$. Fields on AdS naturally lie in the same multiplets of the conformal group as primary operators; the second Casimir of these representations is $C_2 = \Delta(\Delta-4)$ for a primary scalar field of dimension Δ in the field theory, and $C_2 = m^2R^2$ 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 $1/l_{\rm s}$ or of the Planck scale $1/l_{\rm p}$. Such states would correspond (using the mass/dimension relation described above) to operators in the field theory with dimensions of order $\Delta \sim (g_{\rm s}N)^{1/4}$ or $\Delta \sim N^{1/4}$ for large $N, g_{\rm s}N$. 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 field τ may be expanded as $\tau(x,y) = \sum_{k=0}^{\infty} \tau^k(x) Y^k(y)$ where x is a coordinate on AdS_5, y is a coordinate on S^5, and the Y^k are the scalar spherical harmonics on S^5. These spherical harmonics are in representations corresponding to symmetric traceless products of **6**'s of $SU(4)_{\rm R}$; they may be written as $Y^k(y) \sim y^{I_1} y^{I_2} \cdots y^{I_k}$ where the y^I, for $I = 1, 2, \cdots, 6$ and with $\sum_{I=1}^{6} (y^I)^2 = 1$, are coordinates on S^5. Thus, we find a field $\tau^k(x)$ on AdS_5 in each such $(0,k,0)$ representation of $SU(4)_{\rm R}$, and the equations of motion determine the mass of this field to be $m_k^2 = k(k+4)/R^2$. A similar expansion may be performed for all other fields.

If we organize the results of [103] into representations of the superconformal algebra [80], we find representations of the form described in the previous section, which are built on a lowest dimension field which is a scalar in the $(0,n,0)$ representation of $SU(4)_{\rm R}$ 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

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

$n+1$ 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 $SU(4)_{\mathrm{R}}$ gauge fields which correspond to (or couple to) the global $SU(4)_{\mathrm{R}}$ currents in the field theory.

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

Comparing the results described above with the results of Section 3.1.1, we see that we find the same spectrum of chiral primary operators for $n = 2, 3, \cdots, N$. The supergravity results cannot be trusted for masses above the order of the string scale (which corresponds to $n \sim (g_{\mathrm{s}}N)^{1/4}$) 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 $m^2 \sim N^2/R^2 \sim N^{3/2}M_{\mathrm{p}}^2$ 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 $g_{\mathrm{s}} = 0$ we must take $N = \infty$ to obtain a finite value of $g_{\mathrm{s}}N$. 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 [106–108]) 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 $g_{\rm s}N$, the dimension of all non-chiral operators in the field theory, such as $\mathrm{Tr}(\phi^I\phi^I)$, grows at least as $(g_{\rm s}N)^{1/4} \sim (g_{\rm YM}^2 N)^{1/4}$. 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 $SU(4)_{\rm R}$ 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 $SU(4)_{\rm R}$ 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 $SU(4)_{\rm R}$ global currents. These currents are chiral since the fermions $\lambda_{\alpha A}$ are in the $\bar{\mathbf{4}}$ representation while the fermions of the opposite chirality $\bar\lambda^A_{\dot\alpha}$ are in the $\mathbf{4}$ representation. Thus, if we gauge the $SU(4)_{\rm R}$ global symmetry, we will find an Adler-Bell-Jackiw anomaly from the triangle diagram of three $SU(4)_{\rm R}$ 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 $SU(4)_{\rm R}$ global currents,

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

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

$$(\mathcal{D}^\mu J_\mu)^a = \frac{N^2-1}{384\pi^2}id^{abc}\epsilon^{\mu\nu\rho\sigma}F^b_{\mu\nu}F^c_{\rho\sigma}. \tag{3.3}$$

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 (3.3), 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 (3.3) from the point of view of the AdS theory [24]. In the AdS theory we have gauge fields A^a_μ which couple, as explained above, to the $SU(4)_{\mathrm{R}}$ global currents J^a_μ of the gauge theory, but the anomaly means that when we turn on non-zero field strengths for these fields the theory should no longer be gauge invariant. This effect is precisely reproduced by a Chern-Simons term which exists in the low-energy supergravity theory arising from the compactification of type IIB supergravity on $AdS_5 \times S^5$, which is of the form

$$\frac{iN^2}{96\pi^2}\int_{AdS_5} \mathrm{d}^5x(\mathrm{d}^{abc}\epsilon^{\mu\nu\lambda\rho\sigma}A^a_\mu\partial_\nu A^b_\lambda\partial_\rho A^c_\sigma + \cdots). \tag{3.4}$$

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

$$-\frac{iN^2}{384\pi^2}\int_{\partial AdS_5} \mathrm{d}^4x\epsilon^{\mu\nu\rho\sigma}\mathrm{d}^{abc}\Lambda^a F^b_{\mu\nu}F^c_{\rho\sigma}. \tag{3.5}$$

From this we can read off the anomaly in $(\mathcal{D}^\mu J_\mu)$ since, when we have a coupling of the form $\int \mathrm{d}^4x A^\mu_a J^a_\mu$, the change in the action under a gauge transformation is given by $\int \mathrm{d}^4x(\mathcal{D}^\mu\Lambda)_a J^a_\mu = -\int \mathrm{d}^4x\Lambda_a(\mathcal{D}^\mu J^a_\mu)$, and we find exact agreement with (3.3) for large N.

The other anomaly in the $\mathcal{N}=4$ SYM theory is the conformal (or Weyl) anomaly (see [111, 112] and references therein), indicating the breakdown of conformal invariance when the theory is coupled to a curved external metric (there is a similar breakdown of conformal invariance when the theory is coupled to external $SU(4)_{\mathrm{R}}$ 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 [113–116]. In four dimensions, the general form of the conformal anomaly is

$$\langle g^{\mu\nu}T_{\mu\nu}\rangle = -aE_4 - cI_4, \tag{3.6}$$

where

$$\begin{aligned} E_4 &= \frac{1}{16\pi^2}(\mathcal{R}^2_{\mu\nu\rho\sigma} - 4\mathcal{R}^2_{\mu\nu} + \mathcal{R}^2), \\ I_4 &= -\frac{1}{16\pi^2}\left(\mathcal{R}^2_{\mu\nu\rho\sigma} - 2\mathcal{R}^2_{\mu\nu} + \frac{1}{3}\mathcal{R}^2\right), \end{aligned} \tag{3.7}$$

where $\mathcal{R}_{\mu\nu\rho\sigma}$ is the curvature tensor, $\mathcal{R}_{\mu\nu} \equiv \mathcal{R}^{\rho}_{\mu\rho\nu}$ is the Riemann tensor, and $\mathcal{R} \equiv \mathcal{R}^{\mu}_{\mu}$ is the scalar curvature. A free field computation in the $SU(N)$ $\mathcal{N}=4$ SYM theory leads to $a=c=(N^2-1)/4$. In supersymmetric theories the supersymmetry algebra relates $g^{\mu\nu}T_{\mu\nu}$ to derivatives of the R-symmetry current, so it is protected from any quantum corrections. Thus, the same result should be obtained in type IIB string theory on $AdS_5 \times S^5$, and to leading order in the large N limit it should be obtained from type IIB supergravity on $AdS_5 \times S^5$. This was indeed found to be true in [117–120][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 [117–120] 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 $1/N \sim g_{\mathrm{s}}(\alpha'/R^2)^2$; such corrections were discussed in [122] and successfully compared in [123–125][5]. It would be interesting to compare other 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 2.1, 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 $\mathrm{e}^{-I_{\mathrm{SUGRA}}}$, where I_{SUGRA} is the supergravity action evaluated on $AdS_5 \times S^5$ (or on small deformations of this space). This approximation amounts to ignoring all the stringy α' corrections that cure the divergences of supergravity, and also all the loop corrections, which are controlled essentially by the gravitational coupling $\kappa \sim g_{\mathrm{st}}\alpha'^2$. On the gauge theory side,

[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.

as explained in Section 2.1, this approximation amounts to taking both N and $g_{\rm YM}^2 N$ large, and the basic relation becomes

$$\mathrm{e}^{-I_{\rm SUGRA}} \simeq Z_{\rm string} = Z_{\rm gauge} = \mathrm{e}^{-W}, \tag{4.1}$$

where W is the generating functional for connected Green's functions in the gauge theory. At finite temperature, $W = \beta F$ where β is the inverse temperature and F is the free energy of the gauge theory. When we apply this relation to a Schwarzschild black hole in AdS_5, which is thought to be reflected in the gauge theory by a thermal state at the Hawking temperature of the black hole, we arrive at the relation $I_{\rm SUGRA} \simeq \beta F$. 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 (2.15)), with value $\phi_0(\vec{x})$ for $z = \epsilon$. The true boundary of anti-de Sitter space is $z = 0$, and $\epsilon \neq 0$ serves as a cutoff which will eventually be removed. In the supergravity approximation, we think of choosing the values ϕ_0 arbitrarily and then extremizing the action $I_{\rm SUGRA}[\phi]$ in the region $z > \epsilon$ subject to these boundary conditions. In short, we solve the equations of motion in the bulk subject to Dirichlet boundary conditions on the boundary, and evaluate the action on the solution. If there is more than one solution, then we have more than one saddle point contributing to the string theory partition function, and we must determine which is most important. In this section, multiple saddle points will not be a problem. So, we can write

$$W_{\rm gauge}[\phi_0] = -\log \left\langle \mathrm{e}^{\int \mathrm{d}^4 x\, \phi_0(x) \mathcal{O}(x)} \right\rangle_{\rm CFT} \simeq \underset{\phi\big|_{z=\epsilon} = \phi_0}{\rm extremum}\, I_{\rm SUGRA}[\phi]\,. \tag{4.2}$$

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

Note that in (4.2) we have not attempted to be prescient about inserting factors of ϵ. Instead our strategy will be to use (4.2) without modification to compute two-point functions of $\mathcal{O}$, and then perform a wave-function renormalization on either $\mathcal{O}$ or ϕ so that the final answer is independent of the cutoff. This approach should be workable even in a space (with boundary) which is not asymptotically anti-de Sitter, corresponding to a field theory which does not have a conformal fixed point in the ultraviolet.

A remark is in order regarding the relation of (4.2) 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 (4.2) 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 [109, 110, 127–141]–have undertaken the program of extracting explicit n-point correlation functions of gauge singlet operators by developing both sides of (4.2) in a power series in ϕ_0. Because the right hand side is the extremization of a classical action, the power series has a graphical representation in terms of tree-level Feynman graphs for fields in the supergravity. There is one difference: in ordinary Feynman graphs one assigns the wavefunctions of asymptotic states to the external legs of the graph, but in the present case the external leg factors reflect the boundary values ϕ_0. They are special limits of the usual gravity propagators in the bulk, and are called bulk-to-boundary propagators. We will encounter their explicit form in the next two sections.

4.1 Two-point functions

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

$$S = \eta \int \mathrm{d}^5x \sqrt{g} \left[\frac{1}{2}(\partial\phi)^2 + \frac{1}{2} m^2 \phi^2 \right], \tag{4.3}$$

where η is some normalization which in principle follows from the ten-dimensional origin of the action. The bulk-to-boundary propagator is a particular solution of the equation of motion, $(\square - m^2)\phi = 0$, 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

$$\mathrm{d}s^2 = \frac{R^2}{z^2}\left(\mathrm{d}\vec{x}^2 + \mathrm{d}z^2\right) \tag{4.4}$$

provides manifest Euclidean symmetry on the directions parametrized by $\vec{x}$. To avoid divergences associated with the small z region of integration in (4.3), we will employ an explicit cutoff, $z \geq \epsilon$.

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

$$\left[u^5 \partial_u \frac{1}{u^3} \partial_u - u^2 - m^2 R^2\right] Z(u) = 0\ . \tag{4.5}$$

There are two independent solutions to (4.5), namely $Z(u) = u^2 I_{\Delta-2}(u)$ and $Z(u) = u^2 K_{\Delta-2}(u)$, where I_ν and K_ν are Bessel functions and

$$\Delta = 2 + \sqrt{4 + m^2 R^2}. \tag{4.6}$$

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 $\phi(\vec{x}, z) = \phi_0(\vec{x}) = \mathrm{e}^{i\vec{p}\cdot\vec{x}}$ at $z = \epsilon$, we find the bulk-to-boundary propagator

$$\phi(\vec{x}, z) = K_{\vec{p}}(\vec{x}, z) = \frac{(pz)^2 K_{\Delta-2}(pz)}{(p\epsilon)^2 K_{\Delta-2}(p\epsilon)} \mathrm{e}^{i\vec{p}\cdot\vec{x}}. \tag{4.7}$$

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

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

$$\begin{aligned}
\langle \mathcal{O}(\vec{p})\mathcal{O}(\vec{q})\rangle &= \left.\frac{\partial^2 W\left[\phi_0 = \lambda_1 \mathrm{e}^{i\vec{p}\cdot x} + \lambda_2 \mathrm{e}^{i\vec{q}\cdot x}\right]}{\partial\lambda_1\partial\lambda_2}\right|_{\lambda_1=\lambda_2=0} \\
&= (\text{leading analytic terms in } (\epsilon p)^2) \\
&\quad - \eta\epsilon^{2\Delta-8}(2\Delta-4)\frac{\Gamma(3-\Delta)}{\Gamma(\Delta-1)}\delta^4(\vec{p}+\vec{q})\left(\frac{\vec{p}}{2}\right)^{2\Delta-4} \qquad (4.8)\\
&\quad + (\text{higher order terms in } (\epsilon p)^2), \\
\langle \mathcal{O}(\vec{x})\mathcal{O}(\vec{y})\rangle &= \eta\epsilon^{2\Delta-8}\frac{2\Delta-4}{\Delta}\frac{\Gamma(\Delta+1)}{\pi^2\Gamma(\Delta-2)}\frac{1}{|\vec{x}-\vec{y}|^{2\Delta}}.
\end{aligned}$$

Several explanatory remarks are in order:

- To establish the second equality in (4.8) we have used (4.7), substituted in (4.3), performed the integral and expanded in ϵ. The leading analytic terms give rise to contact terms in position space, and the higher order terms are unimportant in the limit where we remove the cutoff. Only the leading nonanalytic term is essential. We have given the expression for generic real values of Δ. Expanding around integer $\Delta \geq 2$ one obtains finite expressions involving $\log \epsilon p$.

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

- We have assumed a coupling $\int \mathrm{d}^4x\, \phi(\vec{x}, z=\epsilon)\mathcal{O}(\vec{x})$ to compute the Green's functions. The explicit powers of the cutoff in the final position space answer can be eliminated by absorbing a factor of $\epsilon^{\Delta-4}$ into the definition of $\mathcal{O}$. From here on we will take that convention, which amounts to inserting a factor of $\epsilon^{4-\Delta}$ on the right hand side of (4.7). In fact, precise matchings between the normalizations in field theory and in string theory for all the chiral primary operators have not been worked out. In part this is due to the difficulty of determining the coupling of bulk fields to field theory operators (or in stringy terms, the coupling of closed string states to composite open string operators on the brane). See [15] for an early approach to this problem. For the dilaton, the graviton, and their superpartners (including gauge fields in AdS_5), the couplings can be worked out explicitly. In some of these cases all normalizations have been worked out unambiguously and checked against field theory predictions (see for example [23, 109, 134]).

- The mass-dimension relation (4.6) 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 $\Delta \sim (g_{\rm YM}^2 N)^{1/4}$. The remarkable fact is that all the string theory modes with $m \sim 1/R$ (which is to say, all closed string states which arise from massless ten dimensional fields) fall in short multiplets of the supergroup $SU(2,2|4)$. All other states have a much larger mass. The operators in short multiplets have algebraically protected dimensions. The obvious conclusion is that all operators whose dimensions are not algebraically protected have large dimension in the strong 't Hooft coupling, large N limit to which supergravity applies. This is no longer true for theories of reduced supersymmetry: the supergroup gets smaller, but the Kaluza-Klein states are roughly as numerous as before, and some of them escape the short multiplets and live in long multiplets of the smaller supergroups. They still have a mass on the order of $1/R$, and typically correspond to dimensions which are finite (in the large $g_{\rm YM}^2 N$ limit) but irrational.

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

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

- spin-3/2: $\Delta = \frac{1}{2}(d + 2|m|)$;
- massless spin-2: $\Delta = d$.

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

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

4.2 Three-point functions

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

$$K_\Delta(\vec{x}, z; \vec{y}) \to z^{4-\Delta}\delta^4(\vec{x} - \vec{y}) \quad \text{as} \quad z \to 0. \tag{4.9}$$

Here $\vec{y}$ is the point on the boundary where we insert the operator, and $(\vec{x}, z)$ is a point in the bulk. The unique regular K_Δ solving the equation of motion and satisfying (4.9) is

$$K_\Delta(\vec{x}, z; \vec{y}) = \frac{\Gamma(\Delta)}{\pi^2\Gamma(\Delta - 2)} \left(\frac{z}{z^2 + (\vec{x} - \vec{y})^2}\right)^\Delta \cdot \tag{4.10}$$

At a fixed cutoff, $z = \epsilon$, the bulk-to-boundary propagator $K_\Delta(\vec{x}, \epsilon; \vec{y})$ is a continuous function which approximates $\epsilon^{4-\Delta}\delta^4(\vec{x} - \vec{y})$ better and better as $\epsilon \to 0$. Thus at any finite ϵ, the Fourier transform of (4.10) only approximately coincides with (4.7) (modified by the factor of $\epsilon^{4-\Delta}$ as explained after (4.8)). 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 (4.8). Calculating two-point functions directly using the position-space propagators (4.9), but cutting the bulk integrals off again at ϵ, and finally taking the same $\epsilon \to 0$ answer, one arrives at a different answer. This is not

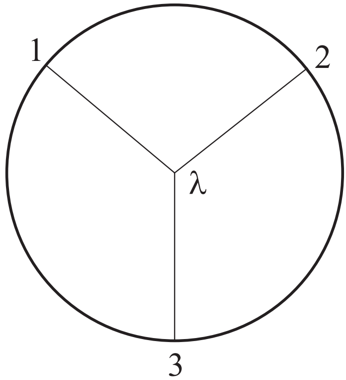


Fig. 2. The Feynman graph for the three-point function as computed in supergravity. The legs correspond to factors of K_{Δ_i}, and the cubic vertex to a factor of λ. The position of the vertex is integrated over AdS_5.

surprising since the $z = \epsilon$ boundary conditions were not used consistently. The authors of [109] checked that using the cutoff consistently ([illegible] with the momentum space propagators) gave two-point functions $\langle \mathcal{O}(\vec{x}_1)\mathcal{O}(\vec{x}_2)\rangle$ a normalization such that Ward identities involving the three-point function $\langle \mathcal{O}(\vec{x}_1)\mathcal{O}(\vec{x}_2)J_\mu(\vec{x}_3)\rangle$, where J_μ is a conserved current, were obeyed. Two-point functions are uniquely difficult because of the poor convergence properties of the integrals over z. The integrals involved in three-point functions are sufficiently benign that one can ignore the issue of how to impose the cutoff.

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

$$S = \int \mathrm{d}^5x\,\sqrt{g}\left[\sum_i \frac{1}{2}(\partial\phi_i)^2 + \tfrac{1}{2}m_i^2\phi_i^2 + \lambda\phi_1\phi_2\phi_3\right], \qquad (4.11)$$

and if the ϕ_i couple to operators in the field theory by interaction terms $\int \mathrm{d}^4x\,\phi_i\mathcal{O}_i$, then the calculation of $\langle \mathcal{O}_1\mathcal{O}_2\mathcal{O}_3\rangle$ reduces, [illegible] (4.2), to the evaluation of the graph shown in Figure 2. That is,

$$\begin{aligned}\langle \mathcal{O}_1(\vec{x}_1)\mathcal{O}_2(\vec{x}_2)\mathcal{O}_3(\vec{x}_3)\rangle &= -\lambda\int \mathrm{d}^5x\,\sqrt{g}K_{\Delta_1}(x;\vec{x}_1)K_{\Delta_2}(x;\vec{x}_2)K_{\Delta_3}(x;\vec{x}_3)\\ &= \frac{\lambda a_1}{|\vec{x}_1-\vec{x}_2|^{\Delta_1+\Delta_2-\Delta_3}|\vec{x}_1-\vec{x}_3|^{\Delta_1+\Delta_3-\Delta_2}|\vec{x}_2-\vec{x}_3|^{\Delta_2+\Delta_3-\Delta_1}},\end{aligned} \qquad (4.12)$$

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 x. The result [109] is

$$a_1 = -\frac{\Gamma\left[\frac{1}{2}(\Delta_1+\Delta_2-\Delta_3)\right]\Gamma\left[\frac{1}{2}(\Delta_1+\Delta_3-\Delta_2)\right]\Gamma\left[\frac{1}{2}(\Delta_2+\Delta_3-\Delta_1)\right]}{2\pi^4\Gamma(\Delta_1-2)\Gamma(\Delta_2-2)\Gamma(\Delta_3-2)} \cdot \Gamma\left[\tfrac{1}{2}(\Delta_1+\Delta_2+\Delta_3)-2\right]. \tag{4.13}$$

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

$$\mathcal{O}^I = \mathcal{C}^I_{J_1\dots J_\ell}\mathrm{Tr}\phi^{(J_1}\dots\phi^{J_\ell)} \tag{4.14}$$

follow from an action of the form

$$S = \frac{4N^2}{(2\pi)^5}\int \mathrm{d}^5x\,\sqrt{g}\Bigg\{\sum_I \frac{A_I(w^I)^2}{2}\left[-(\nabla\tilde{s}_I)^2 - l(l-4)\tilde{s}_I^2\right] + \sum_{I_1,I_2,I_3}\frac{\mathcal{G}_{I_1I_2I_3}w^{I_1}w^{I_2}w^{I_3}}{3}\tilde{s}_{I_1}\tilde{s}_{I_2}\tilde{s}_{I_3}\Bigg\}. \tag{4.15}$$

Derivative couplings of the form $\tilde{s}\partial\tilde{s}\partial\tilde{s}$ are expected [illegible] to enter into (4.15), but an appropriate field redefinition eliminates them. The notation in (4.14) and (4.15) 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 $SU(4)_{\mathrm{R}}$. These are the representations whose Young diagrams are $\yng(1,1)$, $\yng(2,2)$, $\yng(3,3)$, $\cdots$. $\mathcal{C}^I_{J_1\ldots J_\ell}$ is a basis transformation matrix, chosen so that $\mathcal{C}^I_{J_1\ldots J_\ell}\mathcal{C}^J_{J_1\ldots J_\ell} = \delta^{IJ}$. As commented in the previous section, there is generally a normalization ambiguity on how supergravity fields couple to operators in the gauge theory. We have taken the coupling to be $\int \mathrm{d}^4x\, \tilde{s}_I \mathcal{O}^I$, and the normalization ambiguity is represented by the "leg factors" w^I. It is the combination $s^I = w^I \tilde{s}^I$ rather than $\tilde{s}^I$ itself which has a definite relation to supergravity fields. We refer the reader to [132] for explicit expressions for A_I and the symmetric tensor $\mathcal{G}_{I_1I_2I_3}$. To get rid of factors of w^I, we introduce operators $\mathcal{O}^I = \tilde{w}^I \mathcal{O}^I$. One can choose $\tilde{w}^I$ so that a two-point function computation along the lines of Section 4.1 leads to

$$\langle \mathcal{O}^{I_1}(\vec{x})\mathcal{O}^{I_2}(0)\rangle = \frac{\delta^{I_1I_2}}{x^{2\Delta_1}} \cdot \tag{4.16}$$

With this choice, the three-point function, as calculated using (4.12), is

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

where we have defined

$$\langle \mathcal{C}^{I_1}\mathcal{C}^{I_2}\mathcal{C}^{I_3}\rangle = \mathcal{C}^{I_1}_{J_1\cdots J_iK_1\cdots K_j}\mathcal{C}^{I_2}_{J_1\cdots J_iL_1\cdots L_k}\mathcal{C}^{I_3}_{K_1\cdots K_jL_1\cdots L_k}. \tag{4.17}$$

Remarkably, (4.17) is the same result one obtains from free field theory by Wick contracting all the ϕ^J fields in the three operators. This suggests that there is a non-renormalization theorem for this correlation function, but such a theorem has not yet been proven (see however comments at the end of Sect. 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.

5 Wilson loops

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

$$W(\mathcal{C}) = \mathrm{Tr}\left[P\exp\left(i\oint_{\mathcal{C}} A\right)\right] \tag{5.1}$$

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

$$\langle W\rangle \sim \mathrm{e}^{-TV(L)}\ , \tag{5.2}$$

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

5.1 Wilson loops and minimum surfaces

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

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

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

$$W(\mathcal{C}) = \mathrm{Tr}\left[P \exp\left(\oint (iA_\mu \dot{x}^\mu + \theta^I \phi^I \sqrt{\dot{x}^2}) \mathrm{d}\tau\right)\right], \tag{5.3}$$

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

Generalizing the prescription of section 4 for computing correlation functions, the discussion above implies that in order to compute the expectation value of the operator (5.3) 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 3 [166, 167]. In the supergravity regime, when $g_\mathrm{s}N$ is large, the leading contribution to this partition function will come from the area of the string worldsheet. This area is measured with the AdS metric, and it is generally not the same as the area enclosed by the loop $\mathcal{C}$ in four dimensions.

The area as defined above is divergent. The divergence arises from the fact that the string worldsheet is going all the way to the boundary of AdS. If we evaluate the area up to some radial distance $U = r$, we see that for large r it diverges as $r|\mathcal{C}|$, where $|\mathcal{C}|$ is the length of the loop in the field theory [166, 167]. On the other hand, the perturbative computation in the field theory shows that $\langle W \rangle$, for W given by (5.3), is finite, as it should be since a divergence in the Wilson loop would have implied a mass renormalization of the BPS particle. The apparent discrepancy between the divergence of the area of the minimum surface in AdS and the finiteness of the field theory computation can be reconciled by noting that the appropriate action for the string worldsheet is not the area itself but its Legendre transform with respect to the string coordinates corresponding to θ^I and the

[7] The difference in the factor of i between the Euclidean and the Minkowski cases can be traced to the analytic continuation of $\sqrt{\dot{x}^2}$. A detailed derivation of (5.3) can be found in [168].

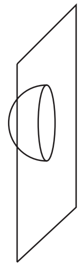


Fig. 3. The Wilson loop operator creates a string worldsheet ending on the corresponding loop on the boundary of AdS.

radial coordinate u [168]. This is because these string coordinates obey the Neumann boundary conditions rather than the Dirichlet conditions. When the loop is smooth, the Legendre transformation simply subtracts the divergent term $r|\mathcal{C}|$, leaving the resulting action finite.

As an example let us consider a circular Wilson loop. Take $\mathcal{C}$ to be a circle of radius a on the boundary, and let us work in the Poincaré coordinates. We could find the surface that minimizes the area by solving the Euler-Lagrange equations. However, in this case it is easier to use conformal invariance. Note that there is a conformal transformation in the field theory that maps a line to a circle. In the case of the line, the minimum area surface is clearly a plane that intersects the boundary and goes all the way to the horizon (which is just a point on the boundary in the Euclidean case). Using the conformal transformation to map the line to a circle we obtain the minimal surface we want. It is, using the coordinates (2.15) for AdS_5,

$$\vec{x} = \sqrt{a^2 - z^2}(\vec{e}_1 \cos\theta + \vec{e}_2 \sin\theta), \tag{5.4}$$

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

$$S \sim \frac{1}{2\pi\alpha'}\mathcal{A} = \frac{R^2}{2\pi\alpha'}\int \mathrm{d}\theta \int_\epsilon^a \frac{\mathrm{d}z a}{z^2} = \frac{R^2}{\alpha'}\left(\frac{a}{\epsilon} - 1\right), \tag{5.5}$$

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

$$\langle W \rangle \sim \mathrm{e}^{-S} \sim \mathrm{e}^{R^2/\alpha'} = \mathrm{e}^{\sqrt{4\pi g_s N}}. \tag{5.6}$$

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 (5.6) but with $g_{\mathrm{s}} \to 1/g_{\mathrm{s}}$.

Using (5.2) it is possible to compute the quark-antiquark potential in the supergravity approximation [166, 167]. 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

$$S=\frac{1}{2\pi\alpha'}\int \mathrm{d}\tau\mathrm{d}\sigma\sqrt{\det(G_{MN}\partial_\alpha X^M\partial_\beta X^N)}, \tag{5.7}$$

where G_{MN} is the Euclidean $AdS_5\times S^5$ metric. Note that the factors of α' cancel out in (5.7), as they should. Since we are interested in a static configuration we take $\tau=t,\ \sigma=x$, and then the action becomes

$$S=\frac{TR^2}{2\pi}\int_{-L/2}^{L/2}\mathrm{d}x\frac{\sqrt{(\partial_x z)^2+1}}{z^2}\,. \tag{5.8}$$

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

$$\frac{1}{z^2\sqrt{(\partial_x z)^2+1}}=\text{constant}. \tag{5.9}$$

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]

$$x=z_0\int_{z/z_0}^{1}\frac{\mathrm{d}y y^2}{\sqrt{1-y^4}}, \tag{5.10}$$

where z_0 is determined by the condition

$$\frac{L}{2}=z_0\int_0^1\frac{\mathrm{d}y y^2}{\sqrt{1-y^4}}=z_0\frac{\sqrt{2}\pi^{3/2}}{\Gamma(1/4)^2}\,. \tag{5.11}$$

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

The qualitative form of the solution is shown in Figure 4b. Notice that the string quickly approaches $x = L/2$ for small z (close to the boundary),

$$\frac{L}{2} - x \sim z^3 \;, \qquad z \to 0 \;. \tag{5.12}$$

Now we compute the total energy of the configuration. We just plug in the solution (5.10) in (5.8), subtract the infinity as explained above (which can be interpreted as the energy of two separated massive quarks, as in Fig. 4a), and we find

$$E = V(L) = -\frac{4\pi^2 (2g_{\rm YM}^2 N)^{1/2}}{\Gamma(\frac{1}{4})^4 L} \,\cdot \tag{5.13}$$

We see that the energy goes as $1/L$, a fact which is determined by conformal invariance. Note that the energy is proportional to $(g_{\rm YM}^2 N)^{1/2}$, as opposed to $g_{\rm YM}^2 N$ 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 $g_{\rm s}N$ is large, independently of the value of $g_{\rm s}$. Some subleading corrections coming from quantum fluctuations of the worldsheet were calculated in [170–172].

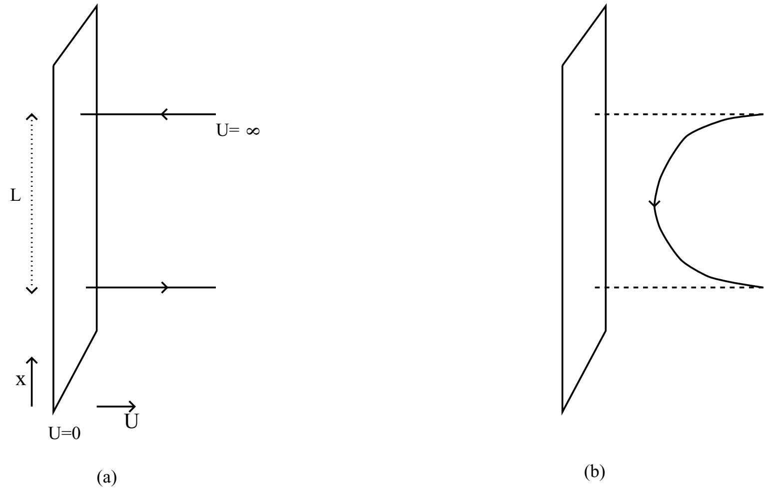


Fig. 4. a) Initial configuration corresponding to two massive quarks before we turn on their coupling to the $U(N)$ gauge theory. **b)** Configuration after we consider the coupling to the $U(N)$ gauge theory. This configuration minimizes the action. The quark-antiquark energy is given by the difference of the total length of the strings in **a)** and **b)**.

In 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 (5.13) but with $g_{\rm YM} \to 4\pi/g_{\rm YM}$. 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 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 $g_{\rm YM}$ [173], but in the limit that $g_{\rm YM}$ is small (but still with $g_{\rm YM}^2 N$ 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 [175–179].

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

$$S = m \int \mathrm{d}s = -2mR \int_\epsilon^a \frac{a\,\mathrm{d}z}{z\sqrt{a^2 - z^2}}, \tag{5.14}$$

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

$$\mathrm{e}^{-S} \sim \mathrm{e}^{-2mR\log a} \sim \frac{1}{a^{2mR}}, \tag{5.15}$$

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 (2.12), 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 [180, 181].

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

6 Theories at finite temperature

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

6.1 Construction

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

$$
\begin{aligned}
\mathrm{d}s^2 &= R^2 \left[u^2(-h\mathrm{d}t^2 + \mathrm{d}x_1^2 + \mathrm{d}x_2^2 + \mathrm{d}x_3^2) + \frac{\mathrm{d}u^2}{hu^2} + \mathrm{d}\Omega_5^2 \right] \\
h &= 1 - \frac{u_0^4}{u^4}\,, \qquad u_0 = \pi T.
\end{aligned}
\tag{6.1}
$$

It will often be useful to Wick rotate by setting $t_{\mathrm{E}} = it$, 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 $g_{\mathrm{YM}}^2 N$. There is no problem on the gauge theory side in working at large N, but large $g_{\mathrm{YM}}^2 N$ at finite temperature is difficult indeed. The analysis of [182] was limited to a free field computation in the field theory, but nevertheless the two results for the entropy agreed up to a factor of a power of 4/3. In the canonical ensemble, where temperature and volume are the independent variables, one identifies the field theory volume with the world-volume of the D3-branes, and one sets the field theory temperature equal to the Hawking temperature in supergravity. The result is

$$
\begin{aligned}
F_{\mathrm{SUGRA}} &= -\frac{\pi^2}{8} N^2 V T^4, \\
F_{\mathrm{SYM}} &= \frac{4}{3} F_{\mathrm{SUGRA}}.
\end{aligned}
\tag{6.2}
$$

The supergravity result is at leading order in l_{s}/R, and it would acquire corrections suppressed by powers of TR if we had considered the full D3-brane metric rather than the near-horizon limit, equation (6.1). 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

$$F = -f(g_{\rm YM}^2 N)\frac{\pi^2}{6}N^2 V T^4, \tag{6.3}$$

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

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

$$\begin{aligned} f(g_{\rm YM}^2 N) &= 1 - \frac{3}{2\pi^2}g_{\rm YM}^2 N + \ldots && \text{for small } g_{\rm YM}^2 N, \\ f(g_{\rm YM}^2 N) &= \frac{3}{4} + \frac{45}{32}\frac{\zeta(3)}{(g_{\rm YM}^2 N)^{3/2}} + \ldots && \text{for large } g_{\rm YM}^2 N. \end{aligned} \tag{6.4}$$

The weak coupling result is a straightforward although somewhat tedious application of the diagrammatic methods of perturbative finite-temperature field theory. The constant term is from one loop, and the leading correction is from two loops. The strong coupling result follows from considering the leading α' corrections to the supergravity action. The relevant one involves a particular contraction of four powers of the Weyl tensor. It is important now to work with the Euclidean solution, and one restricts attention

[9]The result of [182], $S_{\rm SYM} = (4/3)^{1/4} S_{\rm SUGRA}$, differs superficially from (6.2), 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.

further to the near-horizon limit. The Weyl curvature comes from the non-compact part of the metric, which is no longer AdS_5 but rather the AdS-Schwarzschild solution which we will discuss in more detail in Section 6.2. The action including the α' corrections no longer has the Einstein-Hilbert form, and correspondingly the Bekenstein-Hawking prescription no longer agrees with the free energy computed as βI where I is the Euclidean action. In keeping with the basic prescription for computing Green's functions, where a free energy in field theory is equated (in the appropriate limit) with a supergravity action, the relation $I = \beta F$ is regarded as the correct one (see [188]). It has been conjectured that the interpolating function $f(g_{\rm YM}^2 N)$ 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 $g_{\rm YM}^2 N$ 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 $e^{-I_{\rm SUGRA}}$ 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 $I_{\rm SUGRA}$ is the one that dominates the path integral.

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

$$I = -\frac{1}{16\pi G_5}\int \mathrm{d}^5x\,\sqrt{g}\left(\mathcal{R}+\frac{12}{R^2}\right)\cdot \tag{6.5}$$

Explicitly, the metric is

$$\begin{aligned} \mathrm{d}s^2 &= f\mathrm{d}t^2+\frac{1}{f}\mathrm{d}r^2+r^2\mathrm{d}\Omega_3^2,\\ f &= 1+\frac{r^2}{R^2}-\frac{\mu}{r^2}\cdot \end{aligned} \tag{6.6}$$

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

$$\beta = \frac{2\pi R^2 r_+}{2r_+^2+R^2}\cdot \tag{6.7}$$

Topologically, this space is $S^3\times B^2$, and the boundary is $S^3\times S^1$ (which is the relevant space for the field theory on S^3 with finite temperature). We will call this space X_2. Another space with the same boundary which is also a local extremum of (6.5) is given by the metric in (6.6) with $\mu=0$ and again with periodic time. This space, which we will call X_1, is not only metrically distinct from the first (being locally conformally flat), but also topologically $B^4\times S^1$ rather than $S^3\times B^2$. Because the S^1 factor is not simply connected, there are two possible spin structures on X_1, corresponding to thermal (anti-periodic) or supersymmetric (periodic) boundary conditions on fermions. In contrast, X_2 is simply connected and hence admits a unique spin structure, corresponding to thermal boundary conditions. For the purpose of computing the twisted partition function, $\mathrm{Tr}(-1)^{\mathrm{F}}\mathrm{e}^{-\beta\mathrm{H}}$, in a saddle-point approximation, only X_1 contributes. But, X_1 and X_2 make separate saddle-point contributions to the usual thermal partition function, $\mathrm{Tr}\mathrm{e}^{-\beta\mathrm{H}}$, and the more important one is the one with the smaller Euclidean action.

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

$$I(X_2) - I(X_1) = \frac{\pi^2 r_+^3 (R^2 - r_+^2)}{4G_5(2r_+^2 + R^2)} , \tag{6.8}$$

where again r_+ is the largest root of $f = 0$. The fact that (6.8) (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 [192–198]. 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 [201–204] for other interesting contributions to the finite temperature literature.

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

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

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

In [68] the Euclidean black hole solution (6.6) 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].

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 in part by by the Packard foundation and the DOE grant DE-FGO2-91ER40654.

References

[1] O. Aharony, S.S. Gubser, J. Maldacena, H. Ooguri and Y. Oz, *Phys. Rept.* **323** (2000) 183.
[2] H.J. Boonstra, B. Peeters and K. Skenderis, *Phys. Lett. B* **411** (1997) 59.
[3] K. Sfetsos and K. Skenderis, *Nucl. Phys. B* **517** (1998) 179.
[4] S.R. Das and S.P. Trivedi, *Phys. Lett. B* **445** (1998) 142.
[5] F. Gonzalez-Rey, I. Park and K. Schalm, *Phys. Lett. B* **448** (1999) 37.
[6] A. Brandhuber, N. Itzhaki, J. Sonnenschein and S. Yankielowicz, *Phys. Lett. B* **434** (1998) 36.
[7] G. 't Hooft, *Nucl. Phys. B* **72** (1974) 461.
[8] A.M. Polyakov, *Phys. Lett. B* **103** (1981) 207.
[9] J. Maldacena, *Adv. Theor. Math. Phys.* **2** (1998) 231 [hep-th/9711200].
[10] J. Polchinski, *Phys. Rev. Lett.* **75** (1995) 4724 [hep-th/9510017].
[11] G. 't Hooft, *Nucl. Phys. B* **79** (1974) 276.

[12] A.M. Polyakov, *JETP Lett.* **20** (1974) 194.
[13] E. Witten, *Nucl. Phys. B* **460** (1996) 335-350 [hep-th/9510135].
[14] I.R. Klebanov, *Nucl. Phys. B* **496** (1997) 231 [hep-th/9702076].
[15] S.S. Gubser, I.R. Klebanov and A.A. Tseytlin, *Nucl. Phys. B* **499** (1997) 217 [hep-th/9703040].
[16] S.S. Gubser and I.R. Klebanov, *Phys. Lett. B* **413** (1997) 41 [hep-th/9708005].
[17] A. Strominger and C. Vafa, *Phys. Lett. B* **379** (1996) 99 [hep-th/9601029].
[18] C.G. Callan and J.M. Maldacena, *Nucl. Phys. B* **472** (1996) 591 [hep-th/9602043].
[19] G.T. Horowitz and A. Strominger, *Phys. Rev. Lett.* **77** (1996) 2368 [hep-th/9602051].
[20] S.R. Das and S.D. Mathur, *Nucl. Phys. B* **478** (1996) 561 [hep-th/9606185].
[21] A. Dhar, G. Mandal and S.R. Wadia, *Phys. Lett. B* **388** (1996) 51 [hep-th/9605234].
[22] J. Maldacena and A. Strominger, *Phys. Rev. D* **55** (1997) 861 [hep-th/9609026].
[23] S.S. Gubser, I.R. Klebanov and A.M. Polyakov, *Phys. Lett. B* **428** (1998) 105 [hep-th/9802109].
[24] E. Witten, *Adv. Theor. Math. Phys.* **2** (1998) 253 [hep-th/9802150].
[25] A. Salam and E. Sezgin, *Supergravities in Diverse Dimensions*, Vols. 1 and 2 (North-Holland, Amsterdam, Netherlands, 1989).
[26] M.J. Duff, B.E.W. Nilsson and C.N. Pope, *Phys. Rept.* **130** (1986) 1.
[27] E. Witten, *Commun. Missing Math. Phys.* **121** (1989) 351.
[28] T. Banks, W. Fischler, S.H. Shenker and L. Susskind, *Phys. Rev. D* **55** (1997) 5112 [hep-th/9610043].
[29] G. 't Hooft, *Dimensional Reduction in Quantum Gravity* [gr-qc/9310026].
[30] L. Susskind, *J. Math. Phys.* **36** (1995) 6377 [hep-th/9409089].
[31] J.D. Bekenstein, *Phys. Rev. D* **49** (1994) 1912 [gr-qc/9307035].
[32] P.D. Vecchia, *An Introduction to AdS/CFT correspondence* [hep-th/9903007].
[33] M.R. Douglas and S. Randjbar-Daemi, *Two lectures on the AdS/CFT correspondence* [hep-th/9902022].
[34] J.L. Petersen, *Introduction to the Maldacena conjecture on AdS/CFT* [hep-th/9902131].
[35] I.R. Klebanov, *From three-branes to large N gauge theories* [hep-th/9901018].
[36] J. Polchinski, *String Theory* (Cambridge University Press, 1998).
[37] R.G. Leigh, *Mod. Phys. Lett. A* **4** (1989) 2767.
[38] G.T. Horowitz and A. Strominger, *Nucl. Phys. B* **360** (1991) 197.
[39] G.W. Gibbons and P.K. Townsend, *Phys. Rev. Lett.* **71** (1993) 3754 [hep-th/9307049].
[40] E. Witten, *JHEP* **12** (1998) 012 [hep-th/9812012].
[41] N. Seiberg, private communication (1999).
[42] S. Elitzur, G. Moore, A. Schwimmer and N. Seiberg, *Nucl. Phys. B* **326** (1989) 108.
[43] C. Fronsdal, *Phys. Rev. D* **26** (1982) 1988.
[44] D.Z. Freedman and H. Nicolai, *Nucl. Phys. B* **237** (1984) 342.
[45] K. Pilch, P. van Nieuwenhuizen and P.K. Townsend, *Nucl. Phys. B* **242** (1984) 377.
[46] M. Gunaydin, P. van Nieuwenhuizen and N.P. Warner, *Nucl. Phys. B* **255** (1985) 63.
[47] M. Gunaydin and N.P. Warner, *Nucl. Phys. B* **272** (1986) 99.
[48] M. Gunaydin, B.E.W. Nilsson, G. Sierra and P.K. Townsend, *Phys. Lett. B* **176** (1986) 45.

[49] E. Bergshoeff, A. Salam, E. Sezgin and Y. Tanii, *Phys. Lett. B* **205** (1988) 237.

[50] E. Bergshoeff, A. Salam, E. Sezgin and Y. Tanii, *Nucl. Phys. B* **305** (1988) 497.

[51] E. Bergshoeff, M.J. Duff, C.N. Pope and E. Sezgin, the Eleven-Dimensional Supermembrane, *Phys. Lett. B* **224** (1989) 71.

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

[53] C.M. Hull and P.K. Townsend, *Nucl. Phys. B* **438** (1995) 109 [hep-th/9410167].

[54] A.A. Tseytlin, *Nucl. Phys. B* **469** (1996) 51 [hep-th/9602064].

[55] M.B. Green and M. Gutperle, *Phys. Lett. B* **377** (1996) 28 [hep-th/9602077].

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

[57] N. Seiberg and E. Witten, *Nucl. Phys. B* **431** (1994) 484 [hep-th/9408099].

[58] M. Dine and N. Seiberg, *Phys. Lett. B* **409** (1997) 239 [hep-th/9705057].

[59] M.R. Douglas, D. Kabat, P. Pouliot and S.H. Shenker, *Nucl. Phys. B* **485** (1997) 85 [hep-th/9608024].

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

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

[62] S.R. Das, *JHEP* **02** (1999) 012 [hep-th/9901004].

[63] A. Kehagias, *Phys. Lett. B* **435** (1998) 337 [hep-th/9805131].

[64] S.S. Gubser, *Phys. Rev. D* **59** (1999) 025006 [hep-th/9807164].

[65] L.J. Romans, *Phys. Lett. B* **153** (1985) 392.

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

[67] O. Aharony and E. Witten, *JHEP* **11** (1998) 018 [hep-th/9807205].

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

[69] F. Gonzalez-Rey, B. Kulik, I.Y. Park and M. Rocek, *Nucl. Phys. B* **544** (1999) 218 [hep-th/9810152].

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

[71] O. Bergman, *Nucl. Phys. B* **525** (1998) 104 [hep-th/9712211].

[72] K. Hashimoto, H. Hata and N. Sasakura, *Phys. Lett. B* **431** (1998) 303 [hep-th/9803127].

[73] T. Kawano and K. Okuyama, *Phys. Lett. B* **432** (1998) 338 [hep-th/9804139].

[74] O. Bergman and B. Kol, *Nucl. Phys. B* **536** (1998) 149 [hep-th/9804160].

[75] K. Hashimoto, H. Hata and N. Sasakura, *Nucl. Phys. B* **535** (1998) 83 [hep-th/9804164].

[76] K. Lee and P. Yi, *Phys. Rev. D* **58** (1998) 066005 [hep-th/9804174].

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

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

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

[80] M. Gunaydin and N. Marcus, *Class. Quant. Grav.* **2** (1985) L11.

[81] L. Andrianopoli and S. Ferrara, *Phys. Lett. B* **430** (1998) 248 [hep-th/9803171].

[82] L. Andrianopoli and S. Ferrara, *Lett. Math. Phys.* **46** (1998) 265 [hep-th/9807150].

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

[84] M. Gunaydin, D. Minic and M. Zagermann, *Nucl. Phys. B* **534** (1998) 96 [hep-th/9806042].

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

[86] M. Flato and C. Fronsdal, *Lett. Math. Phys.* **2** (1978) 421.

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

[88] M. Flato and C. Fronsdal, *J. Math. Phys.* **22** (1981) 1100.

[89] E. Angelopoulos, M. Flato, C. Fronsdal and D. Sternheimer, *Phys. Rev. D* **23** (1981) 1278.

[90] H. Nicolai and E. Sezgin, *Phys. Lett. B* **143** (1984) 389.

[91] M. Gunaydin and N. Marcus, *Class. Quant. Grav.* **2** (1985) L19.

[92] M. Flato and C. Fronsdal, *Phys. Lett. B* **172** (1986) 412.

[93] S. Ferrara and C. Fronsdal, *Class. Quant. Grav.* **15** (1998) 2153 [hep-th/9712239].

[94] S. Ferrara and C. Fronsdal, *Lett. Math. Phys.* **46** (1998) 157 [hep-th/9806072].

[95] S. Ferrara and C. Fronsdal, *Phys. Lett. B* **433** (1998) 19 [hep-th/9802126].

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

[97] V.K. Dobrev and V.B. Petkova, *Fortschr. Phys.* **35** (1987) 537.

[98] V.K. Dobrev and V.B. Petkova, *Phys. Lett. B* **162** (1985) 127.

[99] N. Seiberg, *Nucl. Phys. Proc. Suppl.* **67** (1998) 158 [hep-th/9705117].

[100] S. Minwalla, *Adv. Theor. Math. Phys.* **2** (1998) 781 [hep-th/9712074].

[101] S. Ferrara, C. Fronsdal and A. Zaffaroni, *Nucl. Phys. B* **532** (1998) 153 [hep-th/9802203].

[102] S. Ferrara, M.A. Lledo and A. Zaffaroni, *Phys. Rev. D* **58** (1998) 105029 [hep-th/9805082].

[103] H.J. Kim, L.J. Romans and P. van Nieuwenhuizen, *Phys. Rev. D* **32** (1985) 389.

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

[105] O. Aharony, M. Berkooz, D. Kutasov and N. Seiberg, *JHEP* **10** (1998) 004 [hep-th/9808149].

[106] E. Witten, *JHEP* **07** (1998) 006 [hep-th/9805112].

[107] I.R. Klebanov and E. Witten, *Nucl. Phys. B* **536** (1998) 199 [hep-th/9807080].

[108] S.S. Gubser and I.R. Klebanov, *Phys. Rev. D* **58** (1998) 125025 [hep-th/9808075].

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

[110] G. Chalmers, H. Nastase, K. Schalm and R. Siebelink, *Nucl. Phys. B* **540** (1999) 247 [hep-th/9805105].

[111] S. Deser and A. Schwimmer, *Phys. Lett. B* **309** (1993) 279 [hep-th/9302047].

[112] M.J. Duff, *Class. Quant. Grav.* **11** (1994) 1387 [hep-th/9308075].

[113] H. Osborn and A. Petkos, *Ann. Phys.* **231** (1994) 311 [hep-th/9307010].

[114] D. Anselmi, M. Grisaru and A. Johansen, *Nucl. Phys. B* **491** (1997) 221 [hep-th/9601023].

[115] J. Erdmenger and H. Osborn, *Nucl. Phys. B* **483** (1997) 431 [hep-th/9605009].

[116] D. Anselmi, D.Z. Freedman, M.T. Grisaru and A.A. Johansen, *Nucl. Phys. B* **526** (1998) 543 [hep-th/9708042].

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

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

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

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

[121] S. Nojiri and S.D. Odintsov, *Phys. Lett. B* **444** (1998) 92 [hep-th/9810008].

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

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

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

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

[126] G.W. Gibbons and S.W. Hawking, *Phys. Rev. D* **15** (1977) 2752.

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

[128] W. Muck and K.S. Viswanathan, *Phys. Rev. D* **58** (1998) 041901 [hep-th/9804035].

[129] H. Liu and A.A. Tseytlin, *Nucl. Phys. B* **533** (1998) 88 [hep-th/9804083].

[130] W. Muck and K.S. Viswanathan, *Phys. Rev. D* **58** (1998) 106006 [hep-th/9805145].

[131] S.N. Solodukhin, *Nucl. Phys. B* **539** (1999) 403 [hep-th/9806004].

[132] S. Lee, S. Minwalla, M. Rangamani and N. Seiberg, *Adv. Theor. Math. Phys.* **2** (1999) 697 [hep-th/9806074].

[133] H. Liu and A.A. Tseytlin, *Phys. Rev. D* **59** (1999) 086002 [hep-th/9807097].

[134] E. D'Hoker, D.Z. Freedman and W. Skiba, *Phys. Rev. D* **59** (1999) 045008 [hep-th/9807098].

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

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

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

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

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

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

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

[142] D.Z. Freedman, K. Johnson and J.I. Latorre, *Nucl. Phys. B* **371** (1992) 353.

[143] M. Henningson and K. Sfetsos, *Phys. Lett. B* **431** (1998) 63 [hep-th/9803251].

[144] A.M. Ghezelbash, K. Kaviani, I. Shahrokh AF Tehran Parvizi and A.H. Fatollahi, *Phys. Lett. B* **435** (1998) 291 [hep-th/9805162].

[145] G.E. Arutyunov and S.A. Frolov, *Nucl. Phys. B* **544** (1999) 576 [hep-th/9806216].

[146] G.E. Arutyunov and S.A. Frolov, *Phys. Lett. B* **441** (1998) 173 [hep-th/9807046].

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

[148] A. Volovich, *JHEP* **09** (1998) 022 [hep-th/9809009].

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

[150] W.S. l'Yi, *Phys. Lett. B* **448** (1999) 218 [hep-th/9811097].

[151] A.S. Koshelev and O.A. Rytchkov, *Phys. Lett. B* **450** (1999) 368 [hep-th/9812238].

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

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

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

[155] G. Dall'Agata, K. Lechner and D. Sorokin, *Class. Quant. Grav.* **14** (1997) L195 [hep-th/9707044].

[156] G. Dall'Agata, K. Lechner and M. Tonin, *JHEP* **07** (1998) 017 [hep-th/9806140].

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

[158] J.H. Schwarz, *Nucl. Phys. B* **226** (1983) 269.

[159] P.S. Howe and P.C. West, *Nucl. Phys. B* **238** (1984) 181.

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

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

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

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

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

[165] D.J. Gross and H. Ooguri, *Phys. Rev. D* **58** (1998) 106002 [hep-th/9805129].

[166] J. Maldacena, *Phys. Rev. Lett.* **80** (1998) 4859 [hep-th/9803002].

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

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

[169] G. 't Hooft, *Nucl. Phys. B* **153** (1979) 141.

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

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

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

[173] J.A. Minahan, *Adv. Theor. Math. Phys.* **2** (1998) 559 [hep-th/9803111].

[174] J.A. Minahan and N.P. Warner, *JHEP* **06** (1998) 005 [hep-th/9805104].

[175] I.I. Kogan and O.A. Solovev, *Phys. Lett. B* **442** (1998) 136 [hep-th/9807223].

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

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

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

[179] E. Alvarez, C. Gomez and T. Ortin, *Nucl. Phys. B* **545** (1999) 217 [hep-th/9806075].

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

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

[182] S.S. Gubser, I.R. Klebanov and A.W. Peet, *Phys. Rev. D* **54** (1996) 3915 [hep-th/9602135].

[183] A. Strominger, unpublished notes (1997).

[184] S.S. Gubser, I.R. Klebanov and A.A. Tseytlin, *Nucl. Phys. B* **534** (1998) 202 [hep-th/9805156].

[185] G.T. Horowitz and J. Polchinski, *Phys. Rev. D* **55** (1997) 6189 [hep-th/9612146].

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

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

[188] R.M. Wald, *Phys. Rev. D* **48** (1993) 3427 [gr-qc/9307038].

[189] M. Li, *JHEP* **03** (1999) 004 [hep-th/9807196].

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

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

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

[193] K. Landsteiner, *Mod. Phys. Lett. A* **14** (1999) 379 [hep-th/9901143].

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

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

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

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

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

[199] P. Kraus, F. Larsen and S.P. Trivedi, *JHEP* **03** (1999) 003 [hep-th/9811120].

[200] A.A. Tseytlin and S. Yankielowicz, *Nucl. Phys. B* **541** (1999) 145 [hep-th/9809032].

[201] D. Birmingham, *Class. Quant. Grav.* **16** (1999) 1197 [hep-th/9808032].

[202] J. Louko and D. Marolf, *Phys. Rev. D* **59** (1999) 066002 [hep-th/9808081].

[203] S.W. Hawking, C.J. Hunter and M.M. Taylor-Robinson, *Phys. Rev. D* **59** (1999) 064005 [hep-th/9811056].

[204] A.W. Peet and S.F. Ross, *JHEP* **12** (1998) 020 [hep-th/9810200].

[205] S.-J. Rey, S. Theisen and J. Yee, *Nucl. Phys. B* **527** (1998) 171 [hep-th/9803135].

LECTURE 5

D-BRANES ON THE CONIFOLD AND $\mathcal{N} = 1$ GAUGE/GRAVITY DUALITIES

I.R. KLEBANOV

Department of Physics, Princeton University Princeton, NJ 08544, U.S.A.

Contents

D-BRANES ON THE CONIFOLD AND $\mathcal{N} = 1$ GAUGE/GRAVITY DUALITIES

I.R. Klebanov, C.P. Herzog and P. Ouyang

Abstract

We review extensions of the AdS/CFT correspondence to gauge/gravity dualities with $\mathcal{N} = 1$ supersymmetry. In particular, we describe the gauge/gravity dualities that emerge from placing D3-branes at the apex of the conifold. We consider first the conformal case, with discussions of chiral primary operators and wrapped D-branes. Next, we break the conformal symmetry by adding a stack of partially wrapped D5-branes to the system, changing the gauge group and introducing a logarithmic renormalization group flow. In the gravity dual, the effect of these wrapped D5-branes is to turn on the flux of 3-form field strengths. The associated RR 2-form potential breaks the $U(1)$ R-symmetry to $\mathbb{Z}_{2\mathrm{M}}$ and we study this phenomenon in detail. This extra flux also leads to deformation of the cone near the apex, which describes the chiral symmetry breaking and confinement in the dual gauge theory.

1 Introduction

Comparison of a stack of D3-branes with the geometry it produces leads to a formulation of duality between $\mathcal{N} = 4$ supersymmetric Yang-Mills theory and type II strings on $AdS_5 \times \mathbf{S}^5$ [1–3]. It is of obvious interest to consider more general dualities between gauge theories and string theories where some of the supersymmetry and/or conformal invariance are broken. These notes are primarily devoted to extensions of the AdS/CFT correspondence to theories with $\mathcal{N} = 1$ supersymmetry.

We first show how to break some of the supersymmetry without destroying conformal invariance. This may be accomplished through placing a stack of D3-branes at the apex of a Ricci flat 6-dimensional cone [4–7]. Then we show how to break the conformal invariance in this set-up and to introduce logarithmic RG flow into the field theory. A convenient way to make the coupling constants run logarithmically is to introduce fractional D3-branes at the apex of the cone [8–10]; these fractional branes may be thought of as D5-branes wrapped over 2-cycles in the base of the cone. In the gravity dual the effect of these wrapped D5-branes is to turn on the flux of 3-form field strengths. This extra flux may lead to deformation of the cone near the apex, which describes the chiral symmetry breaking and confinement in the dual gauge theory [11]. We will start the notes with a very brief review of some of the basic facts about the AdS/CFT correspondence. For more background the reader may consult, for example, the review papers [12,13].

To make the discussion more concrete, we consider primarily one particular example of a cone, the conifold. There are two reasons for this focus. The conifold has enough structure that many new aspects of AdS/CFT correspondence emerge that are not immediately visible for the simplest case, where the conifold is replaced with $\mathbb{R}^6$. At the same time, the conifold is simple enough that we can follow the program outlined in the paragraph above in great detail. This program eventually leads to the warped deformed conifold [11], a solution of type IIB supergravity that is dual to a certain $\mathcal{N}=1$ supersymmetric $SU(N+M)\times SU(N)$ gauge theory in the limit of strong 't Hooft coupling. This solution encodes various interesting gauge theory phenomena in a dual geometrical language, such as the chiral anomaly, the logarithmic running of couplings, the duality cascade in the UV, and chiral symmetry breaking and confinement in the IR.

First, however, we review the original AdS/CFT correspondence. The duality between $\mathcal{N}=4$ supersymmetric $SU(N)$ gauge theory and the $AdS_5\times\mathbf{S}^5$ background of type IIB string theory [1–3] is usually motivated by considering a stack of a large number N of D3-branes. The SYM theory is the low-energy limit of the gauge theory on the stack of D3-branes. On the other hand, the curved background produced by the stack is

$$\mathrm{d}s^2 = h^{-1/2}\left(-\mathrm{d}t^2+\mathrm{d}x_1^2+\mathrm{d}x_2^2+\mathrm{d}x_3^2\right)+h^{1/2}\left(\mathrm{d}r^2+r^2\mathrm{d}\Omega_5^2\right), \qquad (1.1)$$

where $\mathrm{d}\Omega_5^2$ is the metric of a unit 5-sphere and

$$h(r) = 1+\frac{L^4}{r^4}\cdot \qquad (1.2)$$

This 10-dimensional metric may be thought of as a "warped product" of the $\mathbb{R}^{3,1}$ along the branes and the transverse space $\mathbb{R}^6$. Note that the dilaton, $\Phi=0$, is constant, and the selfdual 5-form field strength is given by

$$F_5=\mathcal{F}_5+\star\mathcal{F}_5, \qquad \mathcal{F}_5 = 16\pi(\alpha')^2N\mathrm{vol}(\mathbf{S}^5). \qquad (1.3)$$

The normalization above is dictated by the quantization of Dp-brane tension which implies

$$\int_{\mathbf{S}^{8-p}}\star F_{p+2} = \frac{2\kappa^2\tau_{\mathrm{p}}N}{g_{\mathrm{s}}}, \qquad (1.4)$$

where

$$\tau_{\mathrm{p}} = \frac{\sqrt{\pi}}{\kappa}(4\pi^2\alpha')^{(3-p)/2} \tag{1.5}$$

and $\kappa = 8\pi^{7/2} g_{\mathrm{s}} \alpha'^2$ is the 10-dimensional gravitational constant. In particular, for $p=3$ we have

$$\int_{\mathbf{S}^5} F_5 = (4\pi^2\alpha')^2 N, \tag{1.6}$$

which is consistent with (1.3) since the volume of a unit 5-sphere is

$$\mathrm{Vol}(\mathbf{S}^5) = \pi^3.$$

Note that the 5-form field strength may also be written as

$$g_{\mathrm{s}} F_5 = \mathrm{d}^4 x \wedge \mathrm{d}h^{-1} - r^5 \frac{\mathrm{d}h}{\mathrm{d}r} \mathrm{vol}(\mathbf{S}^5). \tag{1.7}$$

Then it is not hard to see that the Einstein equation

$$R_{\mathrm{MN}} = \frac{g_{\mathrm{s}}^2}{96} F_{\mathrm{MPQRS}} F_{\mathrm{N}}{}^{\mathrm{PQRS}}$$

is satisfied. Since $-r^5 \frac{\mathrm{d}h}{\mathrm{d}r} = 4L^4$, we find by comparing with (1.3) that

$$L^4 = 4\pi g_{\mathrm{s}} N \alpha'^2. \tag{1.8}$$

A related way to determine the scale factor L is to equate the ADM tension of the supergravity solution with N times the tension of a single D3-brane [14]:

$$\frac{2}{\kappa^2} L^4 \mathrm{Vol}(\mathbf{S}^5) = \frac{\sqrt{\pi}}{\kappa} N. \tag{1.9}$$

This way we find

$$L^4 = \frac{\kappa N}{2\pi^{5/2}} = 4\pi g_{\mathrm{s}} N \alpha'^2 \tag{1.10}$$

in agreement with the preceding paragraph.

The radial coordinate r is related to the scale in the dual gauge theory. The low-energy limit corresponds to $r \to 0$. In this limit the metric becomes

$$\mathrm{d}s^2 = \frac{L^2}{z^2}\left(-\mathrm{d}t^2 + \mathrm{d}\vec{x}^2 + \mathrm{d}z^2\right) + L^2 \mathrm{d}\Omega_5^2, \tag{1.11}$$

where $z = \frac{L^2}{r}$. This describes the direct product of 5-dimensional Anti-de Sitter space, AdS_5, and the 5-dimensional sphere, $\mathbf{S}^5$, with equal radii of curvature L.

An interesting generalization of the basic AdS/CFT correspondence [1–3] is found by studying branes at conical singularities [4–7]. Consider a stack of D3-branes placed at the apex of a Ricci-flat 6-d cone Y_6 whose base is a 5-d Einstein manifold X_5. Comparing the metric with the D-brane description leads one to conjecture that type IIB string theory on $AdS_5 \times X_5$ is dual to the low-energy limit of the world volume theory on the D3-branes at the singularity. The equality of tensions now requires [15]

$$L^4 = \frac{\sqrt{\pi}\kappa N}{2\mathrm{Vol}(X_5)} = 4\pi g_{\mathrm{s}} N {\alpha'}^2 \frac{\pi^3}{\mathrm{Vol}(X_5)}, \tag{1.12}$$

an important normalization formula which we will use in the following section.

The simplest examples of X_5 are the orbifolds $\mathbf{S}^5/\Gamma$ where Γ is a discrete subgroup of $SO(6)$ [4]. In these cases X_5 has the local geometry of a 5-sphere. The dual gauge theory is the IR limit of the world volume theory on a stack of N D3-branes placed at the orbifold singularity of $\mathbb{R}^6/\Gamma$. Such theories typically involve product gauge groups $SU(N)^k$ coupled to matter in bifundamental representations [16].

Constructions of the dual gauge theories for Einstein manifolds X_5 which are not locally equivalent to $\mathbf{S}^5$ are also possible. The simplest example is the Romans compactification on $X_5 = T^{1,1} = (SU(2)\times SU(2))/U(1)$ [6,17]. The dual gauge theory is the conformal limit of the world volume theory on a stack of N D3-branes placed at the singularity of a Calabi-Yau manifold known as the conifold [6], which is a cone over $T^{1,1}$. Let us explain this connection in more detail.

2 D3-branes on the conifold

The conifold may be described by the following equation in four complex variables,

$$\sum_{a=1}^{4} z_a^2 = 0. \tag{2.1}$$

Since this equation is invariant under an overall real rescaling of the coordinates, this space is a cone. Remarkably, the base of this cone is precisely the space $T^{1,1}$ [6,18]. In fact, the metric on the conifold may be cast in the form [18]

$$\mathrm{d}s_6^2 = \mathrm{d}r^2 + r^2 \mathrm{d}s_{T^{1,1}}^2, \tag{2.2}$$

where

$$\mathrm{d}s^2_{T^{1,1}} = \frac{1}{9}\left(\mathrm{d}\psi + \sum_{i=1}^{2}\cos\theta_i \mathrm{d}\phi_i\right)^2 + \frac{1}{6}\sum_{i=1}^{2}\left(\mathrm{d}\theta_i^2 + \sin^2\theta_i \mathrm{d}\phi_i^2\right) \tag{2.3}$$

is the metric on $T^{1,1}$. Here ψ is an angular coordinate which ranges from 0 to 4π, while (θ_1, ϕ_1) and (θ_2, ϕ_2) parametrize two $\mathbf{S}^2$s in a standard way. Therefore, this form of the metric shows that $T^{1,1}$ is an $\mathbf{S}^1$ bundle over $\mathbf{S}^2 \times \mathbf{S}^2$.

Now placing N D3-branes at the apex of the cone we find the metric

$$\begin{aligned}\mathrm{d}s^2 &= \left(1+\frac{L^4}{r^4}\right)^{-1/2}\left(-\mathrm{d}t^2 + \mathrm{d}x_1^2 + \mathrm{d}x_2^2 + \mathrm{d}x_3^2\right) \\ &\quad + \left(1+\frac{L^4}{r^4}\right)^{1/2}\left(\mathrm{d}r^2 + r^2 \mathrm{d}s^2_{T^{1,1}}\right)\end{aligned} \tag{2.4}$$

whose near-horizon limit is $AdS_5 \times T^{1,1}$. Using the metric (2.3) it is not hard to find that the volume of $T^{1,1}$ is $\frac{16\pi^3}{27}$ [8]. From (1.12) it then follows that

$$L^4 = 4\pi g_\mathrm{s} N\ (\alpha')^2 \frac{27}{16} = \frac{27\kappa N}{32\pi^{5/2}}\,\cdot \tag{2.5}$$

The same logic that leads us to the maximally supersymmetric version of the AdS/CFT correspondence now shows that the type IIB string theory on this space should be dual to the infrared limit of the field theory on N D3-branes placed at the singularity of the conifold. Since Calabi-Yau spaces preserve 1/4 of the original supersymmetries we find that this should be an $\mathcal{N}=1$ superconformal field theory. This field theory was constructed in [6]: it is $SU(N) \times SU(N)$ gauge theory coupled to two chiral superfields, A_i, in the $(\mathbf{N}, \overline{\mathbf{N}})$ representation and two chiral superfields, B_j, in the $(\overline{\mathbf{N}}, \mathbf{N})$ representation. The A's transform as a doublet under one of the global $SU(2)$s while the B's transform as a doublet under the other $SU(2)$.

A simple way to motivate the appearance of the fields A_i, B_j is to rewrite the defining equation of the conifold, (2.1), as

$$\det_{i,j} z_{ij} = 0, \qquad z_{ij} = \frac{1}{\sqrt{2}}\sum_n \sigma^n_{ij} z_n \tag{2.6}$$

where σ^n are the Pauli matrices for $n = 1, 2, 3$ and σ^4 is i times the unit matrix. This quadratic constraint may be "solved" by the substitution

$$z_{ij} = A_i B_j, \tag{2.7}$$

where $A_i,\ B_j$ are unconstrained variables. If we place a single D3-brane at the singularity of the conifold, then we find a $U(1)\times U(1)$ gauge theory coupled to fields A_1, A_2 with charges $(1,-1)$ and B_1, B_2 with charges $(-1,1)$.

In constructing the generalization to the non-abelian theory on N D3-branes, cancellation of the anomaly in the $U(1)$ R-symmetry requires that the A's and the B's each have R-charge 1/2. For consistency of the duality it is necessary that we add an exactly marginal superpotential which preserves the $SU(2)\times SU(2)\times U(1)_{\rm R}$ symmetry of the theory (this superpotential produces a critical line related to the radius of $AdS_5\times T^{1,1}$). Since a marginal superpotential has R-charge equal to 2 it must be quartic, and the symmetries fix it uniquely up to overall normalization:

$$W = \epsilon^{ij}\epsilon^{kl}{\rm tr} A_i B_k A_j B_l. \tag{2.8}$$

Therefore, it was proposed in [6] that the $SU(N)\times SU(N)$ SCFT with this superpotential is dual to type IIB strings on $AdS_5\times T^{1,1}$.

This proposal can be checked in an interesting way by comparing to a certain $AdS_5\times \mathbf{S}^5/\mathbb{Z}_2$ background. If $\mathbf{S}^5$ is described by an equation

$$\sum_{i=1}^{6} x_i^2 = 1, \tag{2.9}$$

with real variables $x_1,\ldots,x_6$, then the $\mathbb{Z}_2$ acts as -1 on four of the x_i and as $+1$ on the other two. The importance of this choice is that this particular $\mathbb{Z}_2$ orbifold of $AdS_5\times\mathbf{S}^5$ has $\mathcal{N}=2$ superconformal symmetry. Using orbifold results for D-branes [16], this model has been identified [4] as an AdS dual of a $U(N)\times U(N)$ theory with hypermultiplets transforming in $(\mathbf{N},\overline{\mathbf{N}})\oplus(\overline{\mathbf{N}},\mathbf{N})$. From an $\mathcal{N}=1$ point of view, the hypermultiplets correspond to chiral multiplets $A_k, B_l,\ k,l=1,2$ in the $(\mathbf{N},\overline{\mathbf{N}})$ and $(\overline{\mathbf{N}},\mathbf{N})$ representations respectively. The model also contains, from an $\mathcal{N}=1$ point of view, chiral multiplets Φ and $\tilde\Phi$ in the adjoint representations of the two $U(N)$'s. The superpotential is

$$g{\rm Tr}\Phi(A_1B_1 - A_2B_2) + g{\rm Tr}\tilde\Phi(B_1A_1 - B_2A_2).$$

Now, let us add to the superpotential of this $\mathbb{Z}_2$ orbifold a relevant term,

$$\frac{m}{2}\left({\rm Tr}\Phi^2 - {\rm Tr}\tilde\Phi^2\right). \tag{2.10}$$

It is straightforward to see what this does to the field theory. We simply integrate out Φ and $\tilde\Phi$, to find the superpotential

$$-\frac{g^2}{m}\left[{\rm Tr}(A_1B_1A_2B_2) - {\rm Tr}(B_1A_1B_2A_2)\right].$$

This expression is the same as (2.8), so the $\mathbb{Z}_2$ orbifold with relevant perturbation (2.10) apparently flows to the $T^{1,1}$ model associated with the conifold.

Let us try to understand why this works from the point of view of the geometry of $\mathbf{S}^5/\mathbb{Z}_2$. The perturbation in (2.10) is odd under exchange of the two $U(N)$'s. The exchange of the two $U(N)$'s is the quantum symmetry of the $AdS_5 \times \mathbf{S}^5/\mathbb{Z}_2$ orbifold – the symmetry that acts as -1 on string states in the twisted sector and $+1$ in the untwisted sector. Therefore we associate this perturbation with a twisted sector mode of string theory on $AdS_5 \times \mathbf{S}^5/\mathbb{Z}_2$. The twisted sector mode which is a relevant perturbation of the field theory is the blowup of the orbifold singularity of $\mathbf{S}^5/\mathbb{Z}_2$ into the smooth space $T^{1,1}$. A somewhat different derivation of the field theory on D3-branes at the conifold singularity, which is based on blowing up a $\mathbb{Z}_2 \times \mathbb{Z}_2$ orbifold, was given in [7].

It is interesting to examine how various quantities change under the RG flow from the $\mathbf{S}^5/\mathbb{Z}_2$ theory to the $T^{1,1}$ theory. The behavior of the conformal anomaly (which is equal to the $U(1)^3_{\mathrm{R}}$ anomaly) was studied in [15]. Using the fact that the chiral superfields carry R-charge equal to 1/2, on the field theory side it was found that

$$\frac{c_{\mathrm{IR}}}{c_{\mathrm{UV}}} = \frac{27}{32}\,. \tag{2.11}$$

On the other hand, all 3-point functions calculated from supergravity on $AdS_5 \times X_5$ carry normalization factor inversely proportional to $\mathrm{Vol}(X_5)$. Thus, on the supergravity side

$$\frac{c_{\mathrm{IR}}}{c_{\mathrm{UV}}} = \frac{\mathrm{Vol}\ (\mathbf{S}^5/\mathbb{Z}_2)}{\mathrm{Vol}\ (T^{1,1})} = \frac{27}{32}\,. \tag{2.12}$$

Thus, the supergravity calculation is in exact agreement with the field theory result (2.11) [15]. This is a striking and highly sensitive test of the $\mathcal{N}=1$ dual pair constructed in [6, 7].

2.1 Dimensions of chiral operators

There are a number of further convincing checks of the duality between this field theory and type IIB strings on $AdS_5 \times T^{1,1}$. Here we discuss the supergravity modes which correspond to chiral primary operators. (For a more extensive analysis of the spectrum of the model, see [19].) For the $AdS_5 \times \mathbf{S}^5$ case, these modes are mixtures of the conformal factors of the AdS_5 and $\mathbf{S}^5$ and the 4-form field. The same has been shown to be true for the $T^{1,1}$ case [15, 19, 20]. In fact, we may keep the discussion of such modes quite general and consider $AdS_5 \times X_5$ where X_5 is any Einstein manifold.

The diagonalization of such modes carried out in [22] for the $\mathbf{S}^5$ case is easily generalized to any X_5. The mixing of the conformal factor and 4-form modes results in the following mass-squared matrix,

$$m^2 = \begin{pmatrix} E+32 & 8E \\ 4/5 & E \end{pmatrix} \tag{2.13}$$

where $E \geq 0$ is the eigenvalue of the Laplacian on X_5. The eigenvalues of this matrix are

$$m^2 = 16 + E \pm 8\sqrt{4+E}. \tag{2.14}$$

We will be primarily interested in the modes which correspond to picking the minus branch: they turn out to be the chiral primary fields. For such modes there is a possibility of m^2 falling in the range

$$-4 < m^2 < -3 \tag{2.15}$$

where there is a two-fold ambiguity in defining the corresponding operator dimension [21].

First, let us recall the $\mathbf{S}^5$ case where the spherical harmonics correspond to traceless symmetric tensors of $SO(6)$, $d^{(k)}_{i_1\ldots i_k}$. Here $E = k(k+4)$, and it seems that the bound (2.15) is satisfied for $k = 1$. However, this is precisely the special case where the corresponding mode is missing: for $k = 1$ one of the two mixtures is the singleton [22]. Thus, all chiral primary operators in the $\mathcal{N} = 4$ $SU(N)$ theory correspond to the conventional branch of dimension, Δ_+. It is now well-known that this family of operators with dimensions $\Delta = k$, $k = 2, 3, \ldots$ is $d^{(k)}_{i_1\ldots i_k} \mathrm{Tr}(X^{i_1} \ldots X^{i_k})$. The absence of $k = 1$ is related to the gauge group being $SU(N)$ rather than $U(N)$. Thus, in this case we do not encounter operator dimensions lower than 2.

The situation is different for $T^{1,1}$. Here there is a family of wave functions labeled by non-negative integer k, transforming under $SU(2) \times SU(2)$ as $(k/2, k/2)$, and with $U(1)_{\mathrm{R}}$ charge k [15,19,20]. The corresponding eigenvalues of the Laplacian are

$$E(k) = 3\left(k(k+2) - \frac{k^2}{4}\right) \cdot \tag{2.16}$$

In [6] it was argued that the dual chiral operators are

$$\mathrm{tr}(A_{i_1} B_{j_1} \ldots A_{i_k} B_{j_k}). \tag{2.17}$$

Since the F-term constraints in the gauge theory require that the i and the j indices are separately symmetrized, we find that their $SU(2) \times SU(2) \times U(1)$ quantum numbers agree with those given by the supergravity analysis.

In the field theory the A's and the B's have dimension $3/4$, hence the dimensions of the chiral operators are $3k/2$.

In studying the dimensions from the supergravity point of view, one encounters an interesting subtlety discussed in [21]. While for $k>1$ only the dimension Δ_+ is admissible, for $k=1$ one could pick either branch. Indeed, from (2.16) we have $E(1)=33/4$ which falls within the range (2.15). Here we find that $\Delta_-=3/2$, while $\Delta_+=5/2$. Since the supersymmetry requires the corresponding dimension to be $3/2$, in this case we have to pick the unconventional Δ_- branch [21]. Choosing this branch for $k=1$ and Δ_+ for $k>1$ we indeed find following [15,19,20] that the supergravity analysis based on (2.14), (2.16) reproduces the dimensions $3k/2$ of the chiral operators (2.17). Thus, the conifold theory provides a simple example of AdS/CFT duality where the Δ_- branch has to be chosen for certain operators.

Let us also note that substituting $E(1)=33/4$ into (2.14) we find $m^2=-15/4$ which corresponds to a conformally coupled scalar in AdS_5 [22]. In fact, the short chiral supermultiplet containing this scalar includes another conformally coupled scalar and a massless fermion [19]. One of these scalar fields corresponds to the lower component of the superfield $\mathrm{Tr}(A_iB_j)$, which has dimension $3/2$, while the other corresponds to the upper component which has dimension $5/2$. Thus, the supersymmetry requires that we pick dimension Δ_+ for one of the conformally coupled scalars, and Δ_- for the other.

2.2 Wrapped D3-branes as "dibaryons"

It is of further interest to consider various branes wrapped over the cycles of $T^{1,1}$ and attempt to identify these states in the field theory [8]. For example, wrapped D3-branes turn out to correspond to baryon-like operators A^N and B^N where the indices of both $SU(N)$ groups are fully antisymmetrized. For large N the dimensions of such operators calculated from the supergravity are found to be $3N/4$ [8]. This is in complete agreement with the fact that the dimension of the chiral superfields at the fixed point is $3/4$ and may be regarded as a direct supergravity calculation of an anomalous dimension in the dual gauge theory.

To show how this works in detail, we need to calculate the mass of a D3-brane wrapped over a minimal volume 3-cycle. An example of such a 3-cycle is the subspace at a constant value of (θ_2,ϕ_2), and its volume is found to be $V_3=8\pi^2L^3/9$ [8]. The mass of the D3-brane wrapped over the 3-cycle is, therefore,

$$m=V_3\frac{\sqrt{\pi}}{\kappa}=\frac{8\pi^{5/2}L^3}{9\kappa}\,\cdot \tag{2.18}$$

For large mL, the corresponding operator dimension Δ approaches

$$mL = \frac{8\pi^{5/2}L^4}{9\kappa} = \frac{3}{4}N, \tag{2.19}$$

where in the last step we used (2.5).

Let us construct the corresponding operators in the dual gauge theory. Since the fields $A^{\alpha}_{k\beta}$, $k = 1, 2$, carry an index α in the $\mathbf{N}$ of $SU(N)_1$ and an index β in the $\overline{\mathbf{N}}$ of $SU(N)_2$, we can construct color-singlet "dibaryon" operators by antisymmetrizing completely with respect to both groups:

$$\mathcal{B}_{1l} = \epsilon_{\alpha_1 \ldots \alpha_N} \epsilon^{\beta_1 \ldots \beta_N} D_l^{k_1 \ldots k_N} \prod_{i=1}^{N} A^{\alpha_i}_{k_i \beta_i}, \tag{2.20}$$

where $D_l^{k_1 \ldots k_N}$ is the completely symmetric $SU(2)$ Clebsch-Gordon coefficient corresponding to forming the $\mathbf{N+1}$ of $SU(2)$ out of N $\mathbf{2}$'s. Thus the $SU(2) \times SU(2)$ quantum numbers of $\mathcal{B}_{1l}$ are $(\mathbf{N+1}, \mathbf{1})$. Similarly, we can construct "dibaryon" operators which transform as $(\mathbf{1}, \mathbf{N+1})$,

$$\mathcal{B}_{2l} = \epsilon^{\alpha_1 \ldots \alpha_N} \epsilon_{\beta_1 \ldots \beta_N} D_l^{k_1 \ldots k_N} \prod_{i=1}^{N} B^{\beta_i}_{k_i \alpha_i}. \tag{2.21}$$

Under the duality these operators map to D3-branes classically localized at a constant (θ_1, ϕ_1). Thus, the existence of two types of "dibaryon" operators is related on the supergravity side to the fact that the base of the $U(1)$ bundle is $\mathbf{S}^2 \times \mathbf{S}^2$. At the quantum level, the collective coordinate for the wrapped D3-brane has to be quantized, and this explains its $SU(2) \times SU(2)$ quantum numbers [8]. The most basic check on the operator identification is that, since the exact dimension of the A's and the B's is 3/4, the dimension of the "dibaryon" operators agrees exactly with the supergravity calculation.

2.3 Other ways of wrapping D-branes over cycles of $T^{1,1}$

There are many other admissible ways of wrapping branes over cycles of $T^{1,1}$ (for a complete list, see [23]). For example, a D3-brane may be wrapped over a 2-cycle, which produces a string in AdS_5. The tension of such a "fat" string scales as $L^2/\kappa \sim N(g_s N)^{-1/2}/\alpha'$. The non-trivial dependence of the tension on the 't Hooft coupling $g_s N$ indicates that such a string is not a BPS saturated object. This should be contrasted with the tension of a BPS string obtained in [24] by wrapping a D5-brane over $\mathbf{RP}^4$: $T \sim N/\alpha'$.

In discussing wrapped 5-branes, we will limit explicit statements to D5-branes: since a (p, q) 5-brane is an $SL(2, \mathbb{Z})$ transform of a D5-brane, our discussion may be generalized to wrapped (p, q) 5-branes using the $SL(2, \mathbb{Z})$

symmetry of the type IIB string theory. If a D5-brane is wrapped over the entire $T^{1,1}$ then, according to the arguments in [24,25], it serves as a vertex connecting N fundamental strings. Since each string ends on a charge in the fundamental representation of one of the $SU(N)$'s, the resulting field theory state is a baryon built out of external quarks.

If a D5-brane is wrapped over an $\mathbf{S}^3$, with its remaining two dimensions parallel to $\mathbf{R}^{3,1}$, then we find a domain wall in the dual field theory. Consider positioning a "fat" string made of a wrapped D3-brane orthogonally to the domain wall. As the string is brought through the membrane, a fundamental string stretched between them is created. The origin of this effect is creation of fundamental strings by crossing D5 and D3 branes, as shown in [26,27].

We should note, however, that the domain wall positioned at some arbitrary AdS_5 radial coordinate r is not stable: its energy scales as r^3. Therefore, the only stable position is at $r = 0$ which is the horizon. The domain wall is tensionless there, and it is unlikely that this object really exists in the dual CFT. We will see, however, that the domain wall made of a wrapped D5-brane definitely exists in the $SU(N) \times SU(N+M)$ generalization of the gauge theory. This theory is confining and, correspondingly, the dual background does not have a horizon. In this case the wrapped D5-brane again falls to the minimum value of the radial coordinate, but its tension there is non-vanishing. This is the BPS domain wall which separates adjacent inequivalent vacua distinguished by the phase of the gluino condensate.

Finally, we show how to construct the $SU(N) \times SU(N+M)$ theories mentioned above. Consider a D5-brane wrapped over the 2-cycle, with its remaining directions filling $\mathbf{R}^{3,1}$. If this object is located at some fixed r, then it is a domain walls in AdS_5. The simplest domain wall is a D3-brane which is not wrapped over the compact manifold. Through an analysis of the five-form flux carried over directly from [24] one can conclude that when one crosses the domain wall, the effect in field theory is to change the gauge group from $SU(N) \times SU(N)$ to $SU(N+1) \times SU(N+1)$.

The field theory interpretation of a D5-brane wrapped around $\mathbf{S}^2$ is more interesting: if on one side of the domain wall we have the original $SU(N) \times SU(N)$ theory, then on the other side the theory is $SU(N) \times SU(N+1)$ [8]. The matter fields A_k and B_k are still bifundamentals, filling out 2 $(\mathbf{N}, \overline{\mathbf{N}+\mathbf{1}}) \oplus 2(\overline{\mathbf{N}}, \mathbf{N}+1)$. One piece of evidence for this claim is the way the D3-branes wrapped over the $\mathbf{S}^3$ behave when crossing the D5-brane domain wall. In homology there is only one $\mathbf{S}^3$, but for definiteness let us wrap the D3-brane around a particular three-sphere $\mathbf{S}^3_{(1)}$ which is invariant under the group $SU(2)_B$ under which the fields B_k transform. The corresponding state in the $SU(N) \times SU(N)$ field theory is $\mathcal{B}_1$ of (2.21).

In the $SU(N) \times SU(N+1)$ theory, one has instead

$$\epsilon_{\alpha_1 \ldots \alpha_N} \epsilon^{\beta_1 \ldots \beta_{N+1}} A^{\alpha_1}_{\beta_1} \ldots A^{\alpha_N}_{\beta_N} \qquad \text{or}$$

$$\epsilon_{\alpha_1 \ldots \alpha_N} \epsilon^{\beta_1 \ldots \beta_{N+1}} A^{\alpha_1}_{\beta_1} \ldots A^{\alpha_N}_{\beta_N} A^{\alpha_{N+1}}_{\beta_{N+1}} \tag{2.22}$$

where we have omitted $SU(2)$ indices. Either the upper index β_{N+1}, indicating a fundamental of $SU(N+1)$, or the upper index α_{N+1}, indicating a fundamental of $SU(N)$, is free.

How can this be in supergravity? The answer is simple: the wrapped D3-brane must have a string attached to it. Indeed, after a wrapped D3-brane has passed through the wrapped D5-brane domain wall, it emerges with a string attached to it due to the string creation by crossing D-branes which together span 8 dimensions [26, 27]. Calculating the tension of a wrapped D5-brane as a function of r shows that it scales as r^4/L^2. Hence, the domain wall is not stable, but in fact wants to move towards $r = 0$. We will assume that the wrapped D5-branes "fall" behind the horizon and are replaced by their flux in the SUGRA background. This gives a well-defined way of constructing the SUGRA duals of the $SU(N) \times SU(N+M)$ gauge theories.

The D5-branes wrapped over 2-cycles are examples of a more general phenomenon. For many singular spaces Y_6 there are fractional D3-branes which can exist only within the singularity [8, 9, 28, 29]. These fractional D3-branes are D5-branes wrapped over (collapsed) 2-cycles at the singularity. In the case of the conifold, the singularity is a point. The addition of M fractional branes at the singular point changes the gauge group to $SU(N+M) \times SU(N)$; the four chiral superfields remain, now in the representation $(\mathbf{N+M}, \overline{\mathbf{N}})$ and its conjugate, as does the superpotential [8,9]. The theory is no longer conformal. Instead, the relative gauge coupling $g_1^{-2} - g_2^{-2}$ runs logarithmically, as pointed out in [9], where the supergravity equations corresponding to this situation were solved to leading order in M/N. In [10] this solution was completed to all orders; the conifold suffers logarithmic warping, and the relative gauge coupling runs logarithmically at all scales. The D3-brane charge, *i.e.* the 5-form flux, decreases logarithmically as well. However, the logarithm in the solution is not cut off at small radius; the D3-brane charge eventually becomes negative and the metric becomes singular.

In [10] it was conjectured that this solution corresponds to a flow in which the gauge group factors repeatedly drop in size by M units, until finally the gauge groups are perhaps $SU(2M) \times SU(M)$ or simply $SU(M)$. It was further suggested that the strong dynamics of this gauge theory would resolve the naked singularity in the metric. The flow is in fact an infinite series of Seiberg duality transformations – a "duality cascade" – in which the number of colors repeatedly drops by M units [11]. Once the number of colors in the smaller gauge group is less than M, non-perturbative effects become essential. We will show that these gauge theories have an exact anomaly-free $\mathbb{Z}_{2\mathrm{M}}$ R-symmetry, which is broken dynamically, as in pure $\mathcal{N} = 1$ Yang-Mills theory, to $\mathbb{Z}_2$. In the supergravity, this occurs through the deformation of the conifold. In short, the resolution of the naked singularity found in [10] occurs through the chiral symmetry breaking of the gauge theory. The resulting space, *a warped deformed conifold*, is completely nonsingular and without a horizon, leading to confinement [11].

3 The RG cascade

The addition of M fractional 3-branes (wrapped D5-branes) at the singular point changes the gauge group to $SU(N+M)\times SU(N)$. Let us consider the effect on the dual supergravity background of adding M wrapped D5-branes. The D5-branes serve as sources of the magnetic RR 3-form flux through the $\mathbf{S}^3$ of $T^{1,1}$. Therefore, the supergravity dual of this field theory involves M units of the 3-form flux, in addition to N units of the 5-form flux:

$$\frac{1}{4\pi^2\alpha'}\int_{\mathbf{S}^3} F_3 = M, \qquad\qquad \frac{1}{(4\pi^2\alpha')^2}\int_{T^{1,1}} F_5 = N. \tag{3.1}$$

The coefficients above follow from the quantization rule (1.4). The warped conifold (KT) solution with such fluxes was constructed in [10].

It will be useful to employ the following basis of 1-forms on the compact space [30]:

$$\begin{aligned} g^1 &= \frac{e^1 - e^3}{\sqrt{2}}, \qquad g^2 = \frac{e^2 - e^4}{\sqrt{2}},\\ g^3 &= \frac{e^1 + e^3}{\sqrt{2}}, \qquad g^4 = \frac{e^2 + e^4}{\sqrt{2}},\\ &\qquad\qquad g^5 = e^5, \end{aligned} \tag{3.2}$$

where

$$\begin{aligned} e^1 &\equiv -\sin\theta_1 \mathrm{d}\phi_1, \qquad e^2 \equiv \mathrm{d}\theta_1,\\ e^3 &\equiv \cos\psi\sin\theta_2 \mathrm{d}\phi_2 - \sin\psi \mathrm{d}\theta_2,\\ e^4 &\equiv \sin\psi\sin\theta_2 \mathrm{d}\phi_2 + \cos\psi \mathrm{d}\theta_2,\\ e^5 &\equiv \mathrm{d}\psi + \cos\theta_1 \mathrm{d}\phi_1 + \cos\theta_2 \mathrm{d}\phi_2. \end{aligned} \tag{3.3}$$

In terms of this basis, the Einstein metric on $T^{1,1}$ assumes the form

$$\mathrm{d}s^2_{T^{1,1}} = \frac{1}{9}(g^5)^2 + \frac{1}{6}\sum_{i=1}^{4}(g^i)^2. \tag{3.4}$$

Keeping track of the normalization factors, in order to be consistent with the quantization conditions (3.1),

$$F_3 = \frac{M\alpha'}{2}\omega_3, \qquad B_2 = \frac{3g_{\mathrm{s}}M\alpha'}{2}\omega_2 \ln(r/r_0), \tag{3.5}$$

$$H_3 = \mathrm{d}B_2 = \frac{3g_{\mathrm{s}}M\alpha'}{2r}\mathrm{d}r \wedge \omega_2, \tag{3.6}$$

where

$$\omega_2 = \frac{1}{2}\left(g^1 \wedge g^2 + g^3 \wedge g^4\right) = \frac{1}{2}\left(\sin\theta_1 \mathrm{d}\theta_1 \wedge \mathrm{d}\phi_1 - \sin\theta_2 \mathrm{d}\theta_2 \wedge \mathrm{d}\phi_2\right), \tag{3.7}$$

$$\omega_3 = \frac{1}{2}g^5 \wedge \left(g^1 \wedge g^2 + g^3 \wedge g^4\right). \tag{3.8}$$

One can show that [31]

$$\int_{\mathbf{S}^2} \omega_2 = 4\pi, \qquad \int_{\mathbf{S}^3} \omega_3 = 8\pi^2 \tag{3.9}$$

where the S^2 is parametrized by $\psi = 0$, $\theta_1 = \theta_2$ and $\phi_1 = -\phi_2$, and the S^3 by $\theta_2 = \phi_2 = 0$. As a result, the quantization condition for RR 3-form flux is obeyed.

Both ω_2 and ω_3 are closed. Note also that

$$g_{\mathrm{s}} \star_6 F_3 = H_3, \qquad g_{\mathrm{s}}F_3 = - \star_6 H_3, \tag{3.10}$$

where $\star_6$ is the Hodge dual with respect to the metric $\mathrm{d}s_6^2$. Thus, the complex 3-form G_3 satisfies the self-duality condition

$$\star_6 G_3 = iG_3, \qquad G_3 = F_3 - \frac{i}{g_{\mathrm{s}}}H_3. \tag{3.11}$$

Note that the self-duality fixes the relative factor of 3 in (3.5) (see (2.2), (2.3)). We will see that this geometrical factor is crucial for reproducing the well-known factor of 3 in the $\mathcal{N} = 1$ beta functions.

It follows from (3.10) that

$$g_{\mathrm{s}}^2 F_3^2 = H_3^2, \tag{3.12}$$

which implies that the dilaton is constant, $\Phi = 0$. Since $F_{3\mu\nu\lambda}H_3^{\mu\nu\lambda} = 0$, the RR scalar vanishes as well.

The 10-d metric found in [10] has the structure of a "warped product" of $\mathbb{R}^{3,1}$ and the conifold:

$$\mathrm{d}s_{10}^2 = h^{-1/2}(r)\mathrm{d}x_n \mathrm{d}x_n + h^{1/2}(r)(\mathrm{d}r^2 + r^2 \mathrm{d}s_{T^{1,1}}^2). \tag{3.13}$$

The solution for the warp factor h may be determined from the trace of the Einstein equation:

$$R = \frac{1}{24}(H_3^2 + g_\mathrm{s}^2 F_3^2) = \frac{1}{12} H_3^2. \tag{3.14}$$

This implies

$$-h^{-3/2}\frac{1}{r^5}\frac{\mathrm{d}}{\mathrm{d}r}(r^5 h') = \frac{1}{6}H_3^2. \tag{3.15}$$

Integrating this differential equation, we find that

$$h(r) = \frac{27\pi(\alpha')^2[g_\mathrm{s}N + a(g_\mathrm{s}M)^2\ln(r/r_0) + a(g_\mathrm{s}M)^2/4]}{4r^4} \tag{3.16}$$

with $a = 3/(2\pi)$.

An important feature of this background is that $\tilde{F}_5$ acquires a radial dependence [10]. This is because

$$\tilde{F}_5 = F_5 + B_2 \wedge F_3, \qquad F_5 = \mathrm{d}C_4, \tag{3.17}$$

and $\omega_2 \wedge \omega_3 = 54\mathrm{vol}(T^{1,1})$. Thus, we may write

$$\tilde{F}_5 = \mathcal{F}_5 + \star\mathcal{F}_5, \qquad \mathcal{F}_5 = 27\pi{\alpha'}^2 N_\mathrm{eff}(r)\mathrm{vol}(T^{1,1}), \tag{3.18}$$

and

$$N_\mathrm{eff}(r) = N + \frac{3}{2\pi} g_\mathrm{s} M^2 \ln(r/r_0). \tag{3.19}$$

The novel phenomenon in this solution is that the 5-form flux present at the UV scale $r = r_0$ may completely disappear by the time we reach a scale where $N_\mathrm{eff} = 0$. The non-conservation of the flux is due to the type IIB SUGRA equation

$$\mathrm{d}\tilde{F}_5 = H_3 \wedge F_3. \tag{3.20}$$

A related fact is that $\int_{\mathbf{S}^2} B_2$ is no longer a periodic variable in the SUGRA solution once the M fractional branes are introduced: as the B_2 flux goes through a period, $N_\mathrm{eff}(r) \to N_\mathrm{eff}(r) - M$ which has the effect of decreasing the 5-form flux by M units. Note from (3.19) that for a single cascade step $N_\mathrm{eff}(r) \to N_\mathrm{eff}(r) - M$ the radius changes by a factor $r_2/r_1 = \exp(-2\pi/3g_\mathrm{s}M)$, agreeing with a result of [32].

Due to the non-vanishing RHS of (3.20), $\frac{1}{(4\pi^2\alpha')^2}\int_{T^{1,1}}\tilde{F}_5$ is not quantized. We may identify this quantity with $N_{\rm eff}$ defining the gauge group $SU(N_{\rm eff}+M)\times SU(N_{\rm eff})$ only at special radii $r_k = r_0\exp(-2\pi k/3g_{\rm s}M)$ where k is an integer. Thus, $N_{\rm eff} = N - kM$. Furthermore, we believe that the continuous logarithmic variation of $N_{\rm eff}(r)$ is related to continuous reduction in the number of degrees of freedom as the theory flows to the IR. Some support for this claim comes from studying the high-temperature phase of this theory using black holes embedded into an asymptotic KT geometry [33]. The effective number of degrees of freedom computed from the Bekenstein–Hawking entropy grows logarithmically with the temperature, in agreement with (3.19).

The metric (3.13) has a naked singularity at $r = r_{\rm s}$ where $h(r_{\rm s}) = 0$. Writing

$$h(r) = \frac{L^4}{r^4}\ln(r/r_{\rm s}), \qquad L^2 = \frac{9g_{\rm s}M\alpha'}{2\sqrt{2}}, \tag{3.21}$$

we find a purely logarithmic RG cascade:

$$\mathrm{d}s^2 = \frac{r^2}{L^2\sqrt{\ln(r/r_{\rm s})}}\mathrm{d}x_n\mathrm{d}x_n + \frac{L^2\sqrt{\ln(r/r_{\rm s})}}{r^2}\mathrm{d}r^2 + L^2\sqrt{\ln(r/r_{\rm s})}\mathrm{d}s^2_{T^{1,1}}. \tag{3.22}$$

Since $T^{1,1}$ expands slowly toward large r, the curvatures decrease there so that corrections to the SUGRA become negligible. Therefore, even if $g_{\rm s}M$ is very small, this SUGRA solution is reliable for sufficiently large radii where $g_{\rm s}N_{\rm eff}(r) \gg 1$. In this regime the separation between the cascade steps is very large, so that the SUGRA calculation of the β-functions may be compared with $SU(N_{\rm eff}+M)\times SU(N_{\rm eff})$ gauge theory. We will work near $r = r_0$ where $N_{\rm eff}$ may be replaced by N.

3.1 Matching of the β-functions

In order to match the two gauge couplings to the moduli of the type IIB theory on $AdS_5\times T^{1,1}$, one notes that the integrals over the $\mathbf{S}^2$ of $T^{1,1}$ of the NS-NS and R-R 2-form potentials, B_2 and C_2, are moduli. In particular, the two gauge couplings are determined as follows [6, 7][1]:

$$\frac{4\pi^2}{g_1^2} + \frac{4\pi^2}{g_2^2} = \frac{\pi}{g_{\rm s}e^{\Phi}}, \tag{3.23}$$

[1] Exactly the same relations apply to the $\mathcal{N}=2$ supersymmetric $\mathbb{Z}_2$ orbifold theory [4, 34].

$$\left[\frac{4\pi^2}{g_1^2}-\frac{4\pi^2}{g_2^2}\right]g_{\mathrm{s}}e^{\Phi}=\frac{1}{2\pi\alpha'}\left(\int_{\mathbf{S}^2}B_2\right)-\pi \quad (\mathrm{mod}\ 2\pi). \tag{3.24}$$

From the quantization condition on H_3, $\frac{1}{2\pi\alpha'}(\int_{\mathbf{S}^2}B_2)$ must be a periodic variable with period 2π. This periodicity is crucial for the cascade phenomenon. These equations are crucial for relating the SUGRA background to the field theory β-functions when the theory is generalized to $SU(N+M)\times SU(N)$ [9,10].

In gauge/gravity duality the 5-dimensional radial coordinate defines the RG scale of the dual gauge theory [1–3,35,36]. There are different ways of establishiing the precise relation. The simplest one is to identify the field theory energy scale Λ with the energy of a stretched string ending on a probe brane positioned at radius r. For all metrics of the form (3.13) this gives

$$\Lambda\sim r. \tag{3.25}$$

In this section we adopt this UV/IR relation, which typically corresponds to the Wilsonian renormalization group.

Now we are ready to interpret the solution of [10] in terms of RG flow in the dual $SU(N+M)\times SU(N)$ gauge theory. The constancy of the dilaton translates into the vanishing of the β-function for $\frac{8\pi^2}{g_1^2}+\frac{8\pi^2}{g_2^2}$. Substituting the solution for B_2 into (3.24) we find

$$\frac{8\pi^2}{g_1^2}-\frac{8\pi^2}{g_2^2}=6M\ln(r/r_{\mathrm{s}})+\mathrm{const}. \tag{3.26}$$

Since $\ln(r/r_{\mathrm{s}})=\ln(\Lambda/\mu)$, (3.26) implies a logarithmic running of $\frac{1}{g_1^2}-\frac{1}{g_2^2}$ in the $SU(N+M)\times SU(N)$ gauge theory. As we mentioned earlier, this SUGRA result is reliable for any value of $g_{\mathrm{s}}M$ provided that $g_{\mathrm{s}}N\gg 1$. We may consider, for instance, $g_{\mathrm{s}}M\ll 1$ so that the cascade jumps are well-separated.

Let us compare with the Shifman–Vainshtein β-functions [37][2]:

$$\frac{\mathrm{d}}{\mathrm{d}\log(\Lambda/\mu)}\frac{8\pi^2}{g_1^2}=3(N+M)-2N(1-\gamma), \tag{3.27}$$

$$\frac{\mathrm{d}}{\mathrm{d}\log(\Lambda/\mu)}\frac{8\pi^2}{g_2^2}=3N-2(N+M)(1-\gamma), \tag{3.28}$$

[2]These expressions for the β-functions differ from the standard NSVZ form [38] by a factor of $1/(1-g^2N_{\mathrm{c}}/8\pi^2)$. The difference comes from the choice of normalization of the vector superfields. We choose the normalization so that the relevant kinetic term in the field theory action is $\frac{1}{2g^2}\int \mathrm{d}^4x\mathrm{d}^2\theta\mathrm{Tr}(W^\alpha W_\alpha)+$ h.c.; this choice is dictated by the form of the supergravity action and differs from the canonical normalization by a factor of $1/g^2$. With this convention the additional factor in the β-function does not appear. A nice review of the derivation of the exact β-functions is in [39].

where γ is the anomalous dimension of operators $\mathrm{Tr} A_i B_j$. The conformal invariance of the field theory for $M = 0$, and symmetry under $M \to -M$, require that $\gamma = -\frac{1}{2} + O[(M/N)^{2n}]$ where n is a positive integer [11]. Taking the difference of the two equations in (3.27) we then find

$$\begin{aligned} \frac{8\pi^2}{g_1^2} - \frac{8\pi^2}{g_2^2} &= M \ln(\Lambda/\mu)[3 + 2(1-\gamma)] \\ &= 6M \ln(\Lambda/\mu)(1 + O[(M/N)^{2n}]). \end{aligned} \tag{3.29}$$

Remarkably, the coefficient $6M$ is in *exact* agreement with the result (3.26) found on the SUGRA side. This consitutes a geometrical explanation of a field theory β-function, including its normalization.

We may also trace the jumps in the rank of the gauge group to a well-known phenomenon in the dual $\mathcal{N} = 1$ field theory, namely, Seiberg duality [40]. The essential observation is that $1/g_1^2$ and $1/g_2^2$ flow in opposite directions and, according to (3.27), there is a scale where the $SU(N + M)$ coupling, g_1, diverges. To continue past this infinite coupling, we perform a $\mathcal{N} = 1$ duality transformation on this gauge group factor. The $SU(N+M)$ gauge factor has $2N$ flavors in the fundamental representation. Under a Seiberg duality transformation, this becomes an $SU(2N - [N + M]) = SU(N - M)$ gauge group. Thus we obtain an $SU(N) \times SU(N - M)$ theory which resembles closely the theory we started with [11].

As the theory flows to the IR, the cascade must stop, however, because negative N is physically nonsensical. Thus, we should not be able to continue the solution (3.22) to the region where N_{eff} is negative. To summarize, the fact that the solution of [10] is singular tells us that it has to be modified in the IR. The necessary modification proceeds *via* the deformation of the conifold, and is discussed in Section 5.

4 The chiral anomaly

In theories with $\mathcal{N} = 1$ supersymmetry, β-functions are related to chiral anomalies [37]. The essential mechanism is the β-functions contribute to the trace anomaly, $\langle T_i^i \rangle$, which is related by supersymmetry to the divergence of the $U(1)_{\mathrm{R}}$ current, $\partial_i J^i$. In the previous section we showed how the logarithmic running of the gauge couplings manifests itself in the dual supergravity solution of [10]. Here we show that the chiral anomaly can be read off the solution as well. Although the metric has a continuous $U(1)_{\mathrm{R}}$ symmetry, the full supergravity solution is only invariant under a $\mathbb{Z}_{2\mathrm{M}}$ subgroup of this $U(1)$. In the dual quantum field theory there are chiral fermions charged under the $U(1)_{\mathrm{R}}$, and so we can understand the R-symmetry breaking as an effect of the chiral anomaly. Anomalies are especially interesting creatures for the gauge/gravity duality, because the

Adler-Bardeen theorem [41] guarantees that anomaly coefficients computed at one loop are exact, with no radiative corrections; the significance of this fact is that we can compute anomaly coefficients in the field theory at weak coupling, then extrapolate the results to strong coupling, where we can use dual gravity methods to check the calculation. In this section we will study some aspects of the anomaly in detail for the cascading gauge theory.

There are three lessons that we can take away from this analysis [42]. First, the anomaly coefficients computed on each side of the duality agree exactly, even for our non-conformal cascading theory with only $\mathcal{N} = 1$ supersymmetry; although this result is hardly surprising, it is a nice check of the duality. Second, the symmetry breaking is a classical effect on the gravity side. There is no need to appeal to instantons, which is a good thing as they do not appear anywhere explicitly in the gravity dual. Finally, the R-symmetry is broken *spontaneously* in the supergravity solution – the bulk vector field dual to the R-current of the gauge theory acquires a mass. The symmetry breaking then appears "anomalous" if one insists on a four-dimensional description.

4.1 The anomaly as a classical effect in supergravity

The asymptotic UV metric (3.13, 2.3) has a $U(1)$ symmetry associated with the rotations of the angular coordinate $\beta = \psi/2$, normalized so that β has period 2π. This is the R-symmetry of the dual gauge theory. It is crucial, however, that the background value of the R-R 2-form C_2 does not have this continuous symmetry. Indeed, although F_3 is $U(1)$ symmetric, there is no smooth global expression for C_2. Locally, we may write

$$C_2 \to M\alpha'\beta\omega_2. \tag{4.1}$$

This expression is not single-valued as a function of the angular variable β, but it is single-valued up to a gauge transformation, so that $F_3 = \mathrm{d}C_2$ is single-valued. In fact, F_3 is completely independent of β. Because of the explicit β dependence, C_2 is not $U(1)$-invariant. Under the transformation $\beta \to \beta + \epsilon$,

$$C_2 \to C_2 + M\alpha'\epsilon\omega_2. \tag{4.2}$$

Since $\int_{\mathbf{S}^2} C_2$ is defined modulo $4\pi^2\alpha'$, a gauge transformation can shift $C_2/(4\pi^2\alpha')$ by an arbitrary integer multiple of $\omega_2/(4\pi)$, so $\beta \to \beta + \epsilon$ is a symmetry precisely if ϵ is an integer multiple of π/M. Because ϵ is anyway only defined mod 2π, a $\mathbb{Z}_{2\mathrm{M}}$ subgroup of the $U(1)$ leaves fixed the asymptotic values of the fields, and thus corresponds to a symmetry of the system. This $\mathbb{Z}_{2\mathrm{M}}$ is a symmetry since it respects the asymptotic values of the fields.

Let us compare the above analysis with the gauge theory. (A similar comparison for the case of an $\mathcal{N}=2$ orbifold theory appeared in [43].) As pointed out in [6], the integral of the RR 2-form potential C_2 over the $\mathbf{S}^2$ of $T^{1,1}$ is a modulus. Because the integral of B_2 was dual to the difference of gauge couplings for the two gauge groups, it is natural that the integral of C_2 is dual to the difference of Θ-angles (it is possible to check this statement explicitly in orbifold backgrounds). The Θ-angles are given by

$$\Theta_1 - \Theta_2 = \frac{1}{\pi\alpha'}\int_{S^2} C_2, \qquad \Theta_1 + \Theta_2 \sim C, \tag{4.3}$$

where C is the RR scalar, which vanishes for the case under consideration. Using the fact that $\int_{\mathbf{S}^2}\omega_2 = 4\pi$, we find that the small $U(1)$ rotation $\beta \to \beta + \epsilon$ induces

$$\Theta_1 = -\Theta_2 = 2M\epsilon. \tag{4.4}$$

With a conventional normalization, the Θ terms appear in the gauge theory action as

$$\int \mathrm{d}^4x \left(\frac{\Theta_1}{32\pi^2}F^a_{ij}\tilde{F}^{aij} + \frac{\Theta_2}{32\pi^2}G^b_{ij}\tilde{G}^{bij}\right). \tag{4.5}$$

If we assume that ϵ is a function of the 4 world volume coordinates x^i, then under the $U(1)$ rotation (4.4) the terms linear in ϵ in the dual gauge theory (4.5) are

$$\int \mathrm{d}^4x \left[-\epsilon\partial_i J^i + \frac{M\epsilon}{16\pi^2}(F^a_{ij}\tilde{F}^{aij} - G^b_{ij}\tilde{G}^{bij})\right], \tag{4.6}$$

where J^i is the chiral R-current. The appearance of the second term is due to the non-invariance of C_2 under the $U(1)$ rotation. Varying with respect to ϵ, we therefore obtain

$$\partial_i J^i = \frac{M}{16\pi^2}\left(F^a_{ij}\tilde{F}^{aij} - G^b_{ij}\tilde{G}^{bij}\right). \tag{4.7}$$

This anomaly equation, derived from supergravity, agrees exactly with our expectations from the gauge theory. A standard result of quantum field theory is that in a theory with chiral fermions charged under a global $U(1)$ symmetry of the classical Lagrangian, the Noether current associated with that symmetry is not generally conserved but instead obeys the equation

$$\partial_i J^i = \frac{1}{32\pi^2}\sum_m n_m R_m F^a_{ij}\tilde{F}^{aij} \tag{4.8}$$

where n_m is the number of chiral fermions with R-charge R_m circulating in the loop of the relevant triangle diagram. In the case of interest, there are two gauge groups, so let us define F^a_{ij} and G^b_{ij} to be the field strengths of $SU(N+M)$ and $SU(N)$ respectively. Now, the chiral superfields A_i, B_j contribute $2N$ flavors to the gauge group $SU(N+M)$, and each one carries R-charge 1/2. The chiral fermions which are their superpartners have R-charge $-1/2$ while the gluinos have R-charge 1. Therefore, the anomaly coefficient is $\frac{M}{16\pi^2}$. An equivalent calculation for the $SU(N)$ gauge group with $2(N+M)$ flavors produces the opposite anomaly, so the anomaly equation as computed from field theory is just (4.7).

The upshot of the calculation presented above is that the chiral anomaly of the $SU(N+M) \times SU(N)$ gauge theory is encoded in the ultraviolet (large r) behavior of the dual classical supergravity solution. No additional fractional D-instanton effects are needed to explain the anomaly. Thus, as often occurs in the gauge/gravity duality, a quantum effect on the gauge theory side turns into a classical effect in supergravity. Similar methods have been used to describe chiral anomalies in other supersymmetric gauge theories [42–44].

4.2 The anomaly as spontaneous symmetry breaking in AdS_5

Let us look for a deeper understanding of the anomaly from the dual gravity point of view. On the gauge theory side, the R-symmetry is global, but in the gravity dual it as usual becomes a gauge symmetry, which must not be anomalous, or the theory would not make sense at all. Rather, we will find that the gauge symmetry is spontaneously broken: the 5-d vector field dual to the R-current of the gauge theory "eats" the scalar dual to the difference of the theta angles and acquires a mass[3]. A closely related mechanism was observed in studies of RG flows from the dual gravity point of view [46, 48]. There R-current conservation was violated not through anomalies but by turning on relevant perturbations or expectation values for fields. In these cases it was shown [46, 48] that the 5-d vector field dual to the R-current acquires a mass through the Higgs mechanism. We will show that symmetry breaking through anomalies can also have the bulk Higgs mechanism as its dual.

In the absence of fractional branes there are no background three-form fluxes, so the $U(1)$ R-symmetry is a true symmetry of the field theory. Because the R-symmetry is realized geometrically by invariance under a rigid shift of the angle β, it becomes a local symmetry in the full gravity

[3]The connection between anomalies in a D-brane field theory and spontaneous symmetry breaking in string theory was previously noted in [45] (and probably elsewhere in the literature).

theory, and the associated gauge fields $A = A_\mu \mathrm{d}x^\mu$ appear as fluctuations of the ten-dimensional metric and RR four-form potential [19,22]. The natural metric ansatz is of the familiar Kaluza-Klein form:

$$\begin{aligned} \mathrm{d}s^2 &= h(r)^{-1/2}\left(\mathrm{d}x_n \mathrm{d}x^n\right) \\ &+h(r)^{1/2} r^2\left[\frac{\mathrm{d}r^2}{r^2}+\frac{1}{9}\left(g^5-2A\right)^2+\frac{1}{6}\sum_{r=1}^4\left(g^r\right)^2\right], \end{aligned} \tag{4.9}$$

where $h(r) = L^4/r^4$, and $L^4 = \frac{27}{16}(4\pi\alpha'^2 g_{\mathrm{s}} N)$. It is convenient to define the one-form $\chi = g^5 - 2A$, which is invariant under the combined gauge transformations

$$\beta \to \beta + \lambda, \qquad A \to A + \mathrm{d}\lambda. \tag{4.10}$$

The equations of motion for the field A_μ appear as the $\chi\mu$ components of Einstein's equations,

$$R_{\mathrm{MN}} = \frac{g_{\mathrm{s}}^2}{4\cdot 4!}\tilde{F}_{\mathrm{MPQRS}}\tilde{F}_{\mathrm{N}}^{\ \mathrm{PQRS}}. \tag{4.11}$$

The five-form flux will also fluctuate when we activate the Kaluza-Klein gauge field; indeed, the unperturbed $\tilde{F}_5$ of (3.18) is not self-dual with respect to the gauged metric (4.9). An appropriate ansatz to linear order in A is

$$\begin{aligned} \tilde{F}_5 = \mathrm{d}C_4 &= \frac{1}{g_{\mathrm{s}}}\mathrm{d}^4x\wedge \mathrm{d}h^{-1} + \frac{\pi\alpha'^2 N}{4}\Big[\chi\wedge g^1\wedge g^2\wedge g^3\wedge g^4 \\ &\quad -\mathrm{d}A\wedge g^5\wedge \mathrm{d}g^5 + \frac{3}{L}\star_5 \mathrm{d}A\wedge \mathrm{d}g^5\Big]. \end{aligned} \tag{4.12}$$

The five-dimensional Hodge dual $\star_5$ is defined with respect to the AdS_5 metric ${\mathrm{d}s_5}^2 = h^{-1/2}\mathrm{d}x_n\mathrm{d}x^n + h^{1/2}\mathrm{d}r^2$. It is straightforward to show that the supergravity field equation $\mathrm{d}\tilde{F}_5 = 0$ implies that the field A satisfies the equation of motion for a massless vector field in AdS_5 space:

$$\mathrm{d}\star_5 \mathrm{d}A = 0. \tag{4.13}$$

Using the identity $\mathrm{d}g^5\wedge \mathrm{d}g^5 = -2g^1\wedge g^2\wedge g^3\wedge g^4$, we can check that the expression for C_4 is

$$\begin{aligned} C_4 &= \frac{1}{g_{\mathrm{s}}}h^{-1}\mathrm{d}^4x + \frac{\pi\alpha'^2 N}{2}\Big[\beta g^1\wedge g^2\wedge g^3\wedge g^4 - \frac{1}{2}A\wedge \mathrm{d}g^5\wedge g^5 \\ &\quad -\frac{3}{2r}h^{-1/4}\star_5 \mathrm{d}A\wedge g^5\Big]. \end{aligned} \tag{4.14}$$

Another way to see that A is a massless vector in AdS_5 is to consider the Ricci scalar for the metric (4.9)

$$R = R(A=0) - \frac{h^{1/2}r^2}{9} F_{\mu\nu}F^{\mu\nu} \tag{4.15}$$

so that on reduction from ten dimensions the five-dimensional supergravity action will contain the action for a massless vector field.

The story changes when we add wrapped D5-branes. As described in Section 2, the 5-branes introduce M units of RR flux through the three-cycle of $T^{1,1}$. Now, the new wrinkle is that the RR three-form flux of (3.5) is not gauge-invariant with respect to shifts of β (4.10). To restore the gauge invariance, we introduce a new field $\theta \sim \int_{S^2} C_2$:

$$F_3 = \mathrm{d}C_2 = \frac{M\alpha'}{2}\left(g^5 + 2\partial_\mu\theta \mathrm{d}x^\mu\right) \wedge \omega_2 \tag{4.16}$$

so that F_3 is invariant under the gauge transformation $\beta \to \beta+\lambda$, $\theta \to \theta-\lambda$. Let us also define $W_\mu = A_\mu + \partial_\mu\theta$. In terms of the gauge invariant forms χ and $W = W_\mu \mathrm{d}x^\mu$,

$$F_3 = \frac{M\alpha'}{2}(\chi + 2W) \wedge \omega_2. \tag{4.17}$$

From (4.17) we can immediately see how the anomaly will appear in the gravity dual. Assuming that the NS-NS three form is still given by (3.6), we find that up to terms of order $g_{\mathrm{s}}M^2/N$ the three-form equation implies

$$\mathrm{d}\star_5 W = 0 \Rightarrow \frac{L^2}{r^2}\partial_i W^i + \frac{1}{r^5}\partial_{\mathrm{r}} r^5 W_{\mathrm{r}} = 0 \tag{4.18}$$

which is just what one would expect for a massive vector field in five dimensions. To a four dimensional observer, however, a massive vector field would satisfy $\partial_i W^i = 0$. Thus in the field theory one cannot interpret the $U(1)$ symmetry breaking as being spontaneous, and the additional W_{r} term in (4.18) appears in four dimensions to be an anomaly.

Another way to see that the vector field becomes massive is to compute its equation of motion. To do this calculation precisely, we should derive the $\chi\mu$ components of Einstein's equations, and also find the appropriate expressions for the five-form and metric up to quadratic order in $g_{\mathrm{s}}M$ and linear order in fluctuations. This approach is somewhat nontrivial. A more heuristic approach is to consider the type IIB supergravity action to quadratic order in W, ignoring the 5-form field strength contributions:

$$\begin{aligned} S &= -\frac{1}{2\kappa_{10}^2}\int \mathrm{d}^{10}x\sqrt{-G_{10}}\left[R_{10} - \frac{{g_\mathrm{s}}^2}{12}|F_3|^2\right] + \dots \qquad (4.19)\\ &\sim -\frac{1}{2\kappa_{10}^2}\int \mathrm{d}^{10}x\sqrt{-G_{10}}\left[-\frac{h^{1/2}r^2}{9}F_{\mu\nu}F^{\mu\nu}\right.\\ &\quad \left. - \left(\frac{g_\mathrm{s}M\alpha'}{2}\right)^2\frac{36}{hr^4}W_\mu W^\mu\right] + \dots \qquad (4.20)\end{aligned}$$

This is clearly the action for a massive four-dimensional vector field, which has as its equation of motion

$$\partial_\mu(hr^7F^{\mu\nu}) = \tilde{m}^2hr^7W^\nu \qquad (4.21)$$

which in differential form notation is $\mathrm{d}(h^{7/4}r^7\star_5\mathrm{d}W) = -\tilde{m}^2h^{7/4}r^7\star_5 W$. From the action (4.20), we see that the mass-squared is given by

$$\tilde{m}^2 = (g_\mathrm{s}M\alpha')^2\frac{81}{2h^{3/2}r^6}. \qquad (4.22)$$

This result, however, ignores the subtlety of the type IIB action in presence of the self-dual 5-form field. A more precise calculation [49], which takes the mixing into account, gives instead the following equation for the transverse vector modes:

$$\left(\frac{1}{hr^7}\partial_\mathrm{r}hr^7\partial_\mathrm{r} + h\partial_i\partial_i - \frac{(9M\alpha')^4}{64h^2r^{10}}\right)W_i = 0. \qquad (4.23)$$

This shows that the 10-d mass actually appears at a higher order in perturbation theory compared to the result (4.22) that ignores the mixing with the 5-form.

Let us compare this result to earlier work. In [46–48] it was shown that the 5-d vector field associated with a $U(1)_\mathrm{R}$ symmetry acquires a mass in the presence of a symmetry-breaking relevant perturbation, and that this mass is related in a simple way to the warp factor of the geometry[4]. It is conventional to write the 5-d gauged supergravity metric in the form

$$\tilde{G}_{\mu\nu}\mathrm{d}x^\mu\mathrm{d}x^\nu = \mathrm{e}^{2T(q)}\eta_{ij}\mathrm{d}x^i\mathrm{d}x^j + \mathrm{d}q^2. \qquad (4.24)$$

The result of [46] is that $m^2 = -2T''$. To relate the 5-d metric (4.24) to the 10-d metric (5.2) we must normalize the 5-d metric so that the graviton has a canonical kinetic term. Doing this carefully we find

$$\tilde{G}_{\mu\nu}\mathrm{d}x^\mu\mathrm{d}x^\nu = \left(hr^4/L^4\right)^{5/6}(h^{-1/2}\eta_{ij}\mathrm{d}x^i\mathrm{d}x^j + h^{1/2}\mathrm{d}r^2). \qquad (4.25)$$

[4]We are grateful to O. DeWolfe and K. Skenderis for pointing out the relevance of this work to the present calculation.

The factor $\left(hr^4/L^4\right)^{5/6}$ arises due to the radial dependence of the size of $T^{1,1}$ through the usual Kaluza–Klein reduction. The radial variables q and r are related, at leading order in $g_\mathrm{s}M^2/N$, by

$$\log(r) \sim \frac{q}{L} - \frac{g_\mathrm{s}M^2}{2\pi N}\left(\frac{q}{L}\right)^2 \cdot \tag{4.26}$$

We can also show that $-2T = -2\log(r) +$(terms which do not affect the mass to leading order in $g_\mathrm{s}M^2/N$), so now computing the mass-squared by the prescription of [46] we obtain

$$m^2 = \frac{4}{\alpha'(3\pi)^{3/2}}\frac{(g_\mathrm{s}M)^2}{(g_\mathrm{s}N)^{3/2}}\cdot \tag{4.27}$$

where this mass applies to a vector field V with a canonical kinetic term for the metric (4.25). For these calculations it is convenient to work with the transverse 4-d vector modes V_i and to decouple the longitudinal modes such as V_r. The equation of motion of V is

$$\left(\mathrm{e}^{-2T}\frac{\partial}{\partial q}\mathrm{e}^{2T}\frac{\partial}{\partial q} + \mathrm{e}^{-2T}\partial_i\partial_i - m^2\right)V_i = 0. \tag{4.28}$$

In fact, this equation follows from (4.23) after a rescaling [49]

$$V_i = (hr^4/L^4)^{2/3}W_i. \tag{4.29}$$

The nonvanishing vector mass is consistent with gauge invariance because the massless vector field A has eaten the scalar field θ, spontaneously breaking the gauge symmetry, as advertised. It is interesting that the anomaly appears as a bulk effect in AdS space, in contrast to some earlier examples [3, 50] where anomalies arose from boundary terms.

The appearance of a mass implies that the R-current operator should acquire an anomalous dimension. From (4.27) it follows that

$$(mL)^2 = \frac{2(g_\mathrm{s}M)^2}{\pi(g_\mathrm{s}N)}\cdot \tag{4.30}$$

Using the AdS/CFT correspondence (perhaps naively, as the KT metric is not asymptotically AdS) we find that the dimension of the current J^μ dual to the vector field W^μ is

$$\Delta = 2 + \sqrt{1+(mL)^2}. \tag{4.31}$$

Therefore, the anomalous dimension of the current is

$$\Delta - 3 \approx (mL)^2/2 = \frac{(g_\mathrm{s}M)^2}{\pi(g_\mathrm{s}N)}\cdot \tag{4.32}$$

We can obtain a rough understanding of this result by considering the relevant weak coupling calculation in the gauge theory. The leading correction to the current-current two-point function comes from the three-loop Feynman diagram composed of two triangle diagrams glued together, and the resulting anomalous dimension $\gamma_{\rm J}$ is quadratic in M and N. $\gamma_{\rm J}$ must vanish when $M=0$, and it must be invariant under the map $M \to -M$, $N \to N+M$, which simply interchanges the two gauge groups. Thus, the lowest order piece of the anomalous dimension will be of order $(g_{\rm s}M)^2$. Our supergravity calculation predicts that this anomalous dimension is corrected at large $g_{\rm s}N$ by an extra factor of $1/(g_{\rm s}N)$. Of course, it would be interesting to understand this result better from the gauge theory point of view.

5 Deformation of (KS) the conifold

It was shown in [11] that, to remove the naked singularity found in [10] the conifold (2.1) should be replaced by the deformed conifold

$$\sum_{i=1}^{4} z_i^2 = \varepsilon^2, \tag{5.1}$$

in which the singularity of the conifold is removed through the blowing-up of the $\mathbf{S}^3$ of $T^{1,1}$. The 10-d metric of [11] takes the following form:

$$\mathrm{d}s_{10}^2 = h^{-1/2}(\tau)\mathrm{d}x_n \mathrm{d}x_n + h^{1/2}(\tau)\mathrm{d}s_6^2, \tag{5.2}$$

where $\mathrm{d}s_6^2$ is the metric of the deformed conifold (5.3). This is the same type of “D-brane” ansatz as (3.13), but with the conifold replaced by the deformed conifold as the transverse space.

The metric of the deformed conifold was discussed in some detail in [18, 30, 51]. It is diagonal in the basis (3.2):

$$\mathrm{d}s_6^2 = \frac{1}{2}\varepsilon^{4/3}K(\tau)\left[\frac{1}{3K^3(\tau)}(\mathrm{d}\tau^2 + (g^5)^2) + \cosh^2\left(\frac{\tau}{2}\right)[(g^3)^2 + (g^4)^2] \right. \\ \left. + \sinh^2\left(\frac{\tau}{2}\right)[(g^1)^2 + (g^2)^2]\right], \tag{5.3}$$

where

$$K(\tau) = \frac{(\sinh(2\tau) - 2\tau)^{1/3}}{2^{1/3}\sinh\tau}\,. \tag{5.4}$$

For large τ we may introduce another radial coordinate r *via*

$$r^2 = \frac{3}{2^{5/3}}\varepsilon^{4/3}\mathrm{e}^{2\tau/3}, \tag{5.5}$$

and in terms of this radial coordinate $\mathrm{d}s_6^2 \to \mathrm{d}r^2 + r^2\mathrm{d}s^2_{T^{1,1}}$.

At $\tau=0$ the angular metric degenerates into

$$\mathrm{d}\Omega_3^2 = \frac{1}{2}\varepsilon^{4/3}(2/3)^{1/3}\left[\frac{1}{2}(g^5)^2+(g^3)^2+(g^4)^2\right], \tag{5.6}$$

which is the metric of a round $\mathbf{S}^3$ [18, 30]. The additional two directions, corresponding to the $\mathbf{S}^2$ fibered over the $\mathbf{S}^3$, shrink as

$$\frac{1}{8}\varepsilon^{4/3}(2/3)^{1/3}\tau^2[(g^1)^2+(g^2)^2]. \tag{5.7}$$

The simplest ansatz for the 2-form fields is

$$\begin{aligned} F_3 &= \frac{M\alpha'}{2}\left\{g^5\wedge g^3\wedge g^4 + \mathrm{d}[F(\tau)(g^1\wedge g^3+g^2\wedge g^4)]\right\} \\ &= \frac{M\alpha'}{2}\left\{g^5\wedge g^3\wedge g^4(1-F)+g^5\wedge g^1\wedge g^2 F\right. \\ &\qquad \left. + F'\mathrm{d}\tau\wedge(g^1\wedge g^3+g^2\wedge g^4)\right\}, \end{aligned} \tag{5.8}$$

with $F(0)=0$ and $F(\infty)=1/2$, and

$$B_2 = \frac{g_{\mathrm{s}}M\alpha'}{2}[f(\tau)g^1\wedge g^2 + k(\tau)g^3\wedge g^4], \tag{5.9}$$

$$\begin{aligned} H_3 = \mathrm{d}B_2 &= \frac{g_{\mathrm{s}}M\alpha'}{2}\Big[\mathrm{d}\tau\wedge(f'g^1\wedge g^2+k'g^3\wedge g^4) \\ &\qquad +\frac{1}{2}(k-f)g^5\wedge(g^1\wedge g^3+g^2\wedge g^4)\Big]. \end{aligned} \tag{5.10}$$

As before, the self-dual 5-form field strength may be decomposed as $\tilde{F}_5 = \mathcal{F}_5 + \star\mathcal{F}_5$. We have

$$\mathcal{F}_5 = B_2\wedge F_3 = \frac{g_{\mathrm{s}}M^2(\alpha')^2}{4}\ell(\tau)g^1\wedge g^2\wedge g^3\wedge g^4\wedge g^5, \tag{5.11}$$

where

$$\ell = f(1-F)+kF, \tag{5.12}$$

and

$$\star\mathcal{F}_5 = 4g_{\mathrm{s}}M^2(\alpha')^2\varepsilon^{-8/3}\mathrm{d}x^0\wedge\mathrm{d}x^1\wedge\mathrm{d}x^2\wedge\mathrm{d}x^3\wedge\mathrm{d}\tau\frac{\ell(\tau)}{K^2h^2\sinh^2(\tau)}. \tag{5.13}$$

5.1 The first-order equations and their solution

In searching for BPS saturated supergravity backgrounds, the second order equations should be replaced by a system of first-order ones. Luckily, this is possible for our ansatz [11]:

$$\begin{aligned} f' &= (1-F)\tanh^2(\tau/2), \\ k' &= F\coth^2(\tau/2), \\ F' &= \frac{1}{2}(k-f), \end{aligned} \tag{5.14}$$

and

$$h' = -\alpha\frac{f(1-F)+kF}{K^2(\tau)\sinh^2\tau}, \tag{5.15}$$

where

$$\alpha = 4(g_{\mathrm{s}}M\alpha')^2\varepsilon^{-8/3}. \tag{5.16}$$

These equations follow from a superpotential for the effective radial problem [52].

Note that the first three of these equations (5.14), form a closed system and need to be solved first. In fact, these equations imply the self-duality of the complex 3-form with respect to the metric of the deformed conifold: $\star_6 G_3 = iG_3$. The solution is

$$\begin{aligned} F(\tau) &= \frac{\sinh\tau-\tau}{2\sinh\tau}, \\ f(\tau) &= \frac{\tau\coth\tau-1}{2\sinh\tau}(\cosh\tau-1), \\ k(\tau) &= \frac{\tau\coth\tau-1}{2\sinh\tau}(\cosh\tau+1). \end{aligned} \tag{5.17}$$

Now that we have solved for the 3-forms on the deformed conifold, the warp factor may be determined by integrating (5.15). First we note that

$$\ell(\tau) = f(1-F)+kF = \frac{\tau\coth\tau-1}{4\sinh^2\tau}(\sinh 2\tau - 2\tau). \tag{5.18}$$

This behaves as τ^3 for small τ. For large τ we impose, as usual, the boundary condition that h vanishes. The resulting integral expression for h is

$$h(\tau) = \alpha\frac{2^{2/3}}{4}I(\tau) = (g_{\mathrm{s}}M\alpha')^2 2^{2/3}\varepsilon^{-8/3}I(\tau), \tag{5.19}$$

where

$$I(\tau) \equiv \int_\tau^\infty \mathrm{d}x \frac{x \coth x - 1}{\sinh^2 x}(\sinh(2x) - 2x)^{1/3}. \tag{5.20}$$

We have not succeeded in evaluating this integral in terms of elementary or well-known special functions, but it is not hard to see that

$$I(\tau \to 0) \to a_0 + O(\tau^2); \tag{5.21}$$

$$I(\tau \to \infty) \to 3 \cdot 2^{-1/3}\left(\tau - \frac{1}{4}\right)\mathrm{e}^{-4\tau/3}, \tag{5.22}$$

where $a_0 \approx 0.71805$. This $I(\tau)$ is nonsingular at the tip of the deformed conifold and, from (5.5), matches the form of the large-τ solution (3.21). The small τ behavior follows from the convergence of the integral (5.19), while at large τ the integrand becomes $\sim x\mathrm{e}^{-4x/3}$.

Thus, for small τ the ten-dimensional geometry is approximately $\mathbb{R}^{3,1}$ times the deformed conifold:

$$\begin{aligned} \mathrm{d}s^2_{10} \quad \to \quad & \frac{\varepsilon^{4/3}}{2^{1/3}a_0^{1/2}g_\mathrm{s}M\alpha'}\mathrm{d}x_n\mathrm{d}x_n + a_0^{1/2}6^{-1/3}(g_\mathrm{s}M\alpha')\Big\{\frac{1}{2}\mathrm{d}\tau^2 + \frac{1}{2}(g^5)^2 \\ & +(g^3)^2 + (g^4)^2 + \frac{1}{4}\tau^2[(g^1)^2 + (g^2)^2]\Big\}. \end{aligned} \tag{5.23}$$

This metric will be useful in Section 6 where we investigate various infrared phenomenon of the gauge theory.

Very importantly, for large $g_\mathrm{s}M$ the curvatures found in our solution are small everywhere. This is true even far in the IR, since the radius-squared of the $\mathbf{S}^3$ at $\tau = 0$ is of order $g_\mathrm{s}M$ in string units. This is the 't Hooft coupling of the gauge theory found far in the IR. As long as this is large, the curvatures are small and the SUGRA approximation is reliable.

5.2 *SO*(4) *invariant expressions for the 3-forms*

In [53, 54] it was shown that the warped background of the previous section preserves $\mathcal{N}=1$ SUSY if and only if G_3 is a $(2,1)$ form on the CY space. Perhaps the easiest way to see the supersymmetry of the deformed conifold solution is through a T-duality. Performing a T-duality along one of the longitudinal directions, and lifting the result to M-theory maps our background to a Becker-Becker solution supported by a G_4 which is a $(2,2)$ form on $T^2 \times$ CY. G-flux of this type indeed produces a supersymmetric background [55].

While writing G_3 in terms of the angular 1-forms g^i is convenient for some purposes, the $(2,1)$ nature of the form is not manifest. That G_3 is indeed $(2,1)$ was demonstrated in [56] with the help of a holomorphic basis. Below we write the G_3 found in [11] in terms of the obvious 1-forms on the deformed conifold: $\mathrm{d}z^i$ and $\mathrm{d}\bar{z}^i$, $i = 1,2,3,4$:

$$G_3 = \frac{M\alpha'}{2\varepsilon^6 \sinh^4 \tau} \left\{ \frac{\sinh(2\tau) - 2\tau}{\sinh \tau} (\epsilon_{ijkl}\, z_i \bar{z}_j \,\mathrm{d}z_k \wedge \mathrm{d}\bar{z}_l) \wedge (\bar{z}_m \,\mathrm{d}z_m) \right.$$
$$\left. + 2(1 - \tau \coth \tau)(\epsilon_{ijkl}\, z_i \bar{z}_j \,\mathrm{d}z_k \wedge \mathrm{d}z_l) \wedge (z_m \,\mathrm{d}\bar{z}_m) \right\}. \quad (5.24)$$

We also note that the NS-NS 2-form potential is an $SO(4)$ invariant $(1,1)$ form:

$$B_2 = \frac{i g_{\mathrm{s}} M\alpha'}{2\varepsilon^4} \frac{\tau \coth \tau - 1}{\sinh^2 \tau} \epsilon_{ijkl}\, z_i \bar{z}_j \,\mathrm{d}z_k \wedge \mathrm{d}\bar{z}_l. \quad (5.25)$$

The derivation of these formulae is given in [31]. Our expressions for the gauge fields are manifestly $SO(4)$ invariant, and so is the metric.

6 Infrared physics

We have now seen that the deformation of the conifold allows the solution to be non-singular. In the following sections we point out some interesting features of the SUGRA background we have found and show how they realize the expected phenomena in the dual field theory. In particular, we will now demonstrate that there is confinement; that the theory has glueballs and baryons whose mass scale emerges through a dimensional transmutation; that there is a gluino condensate that breaks the $\mathbb{Z}_{2\mathrm{M}}$ chiral symmetry down to $\mathbb{Z}_2$ and that there are domain walls separating inequivalent vacua. Other stringy approaches to infrared phenomena in $\mathcal{N} = 1$ SYM theory have recently appeared in [57–59].

6.1 Dimensional transmutation and confinement

The resolution of the naked singularity *via* the deformation of the conifold is a supergravity realization of the dimensional transmutation. While the singular conifold has no dimensionful parameter, we saw that turning on the R-R 3-form flux produces the logarithmic warping of the KT solution. The scale necessary to define the logarithm transmutes into the the parameter ε that determines the deformation of the conifold. From (5.5) we see that $\varepsilon^{2/3}$ has dimensions of length and that

$$\tau = 3 \ln(r/\varepsilon^{2/3}) + \mathrm{const.} \quad (6.1)$$

Thus, the scale r_{s} entering the UV solution (3.21) should be identified with $\varepsilon^{2/3}$. On the other hand, the form of the IR metric (5.23) makes it clear that the dynamically generated 4-d mass scale, which sets the tension of the confining flux tubes, is

$$\frac{\varepsilon^{2/3}}{\alpha' \sqrt{g_{\mathrm{s}} M}} \cdot \qquad (6.2)$$

The reason the theory is confining is that in the metric for small τ (5.23) the function multiplying $\mathrm{d}x_n \mathrm{d}x_n$ approaches a constant. This should be contrasted with the AdS_5 metric where this function vanishes at the horizon, or with the singular metric of [10] where it blows up. Consider a Wilson contour positioned at fixed τ, and calculate the expectation value of the Wilson loop using the prescription [60, 61]. The minimal area surface bounded by the contour bends towards smaller τ. If the contour has a very large area A, then most of the minimal surface will drift down into the region near $\tau = 0$. From the fact that the coefficient of $\mathrm{d}x_n \mathrm{d}x_n$ is finite at $\tau = 0$, we find that a fundamental string with this surface will have a finite tension, and so the resulting Wilson loop satisfies the area law. A simple estimate shows that the string tension scales as

$$T_{\mathrm{s}} = \frac{1}{2^{4/3} a_0^{1/2} \pi} \frac{\varepsilon^{4/3}}{(\alpha')^2 g_{\mathrm{s}} M} \cdot \qquad (6.3)$$

We will return to these confining strings in the next section.

The masses of glueball and Kaluza-Klein (KK) states scale as

$$m_{\mathrm{glueball}} \sim m_{\mathrm{KK}} \sim \frac{\varepsilon^{2/3}}{g_{\mathrm{s}} M \alpha'} \cdot \qquad (6.4)$$

Comparing with the string tension, we see that

$$T_{\mathrm{s}} \sim g_{\mathrm{s}} M (m_{\mathrm{glueball}})^2. \qquad (6.5)$$

Due to the deformation, the full SUGRA background has a finite 3-cycle. We may interpret various branes wrapped over this 3-cycle in terms of the gauge theory. Note that the 3-cycle has the minimal volume near $\tau = 0$, hence all the wrapped branes will be localized there. A wrapped D3-brane plays the role of a baryon vertex which ties together M fundamental strings. Note that for $M = 0$ the D3-brane wrapped on the $\mathbf{S}^3$ gave a dibaryon [8]; the connection between these two objects becomes clearer when one notes that for $M > 0$ the dibaryon has M uncontracted indices, and therefore joins M external charges. Studying a probe D3-brane in the background of our solution show that the mass of the baryon scales as

$$M_{\mathrm{b}} \sim M \frac{\varepsilon^{2/3}}{\alpha'} \cdot \qquad (6.6)$$

6.2 Tensions of the q-strings

The existence of the blown up 3-cycle with M units of RR 3-form flux through it is responsible for another interesting infrared phenomenon, the appearance of composite confining strings. To explain what they are, let us recall that the basic string corresponds to the Wilson loop in the fundamental representation. The classic criterion for confinement is that this Wilson loop obey the area law

$$-\ln\langle W_1(C)\rangle = T_1 A(C) \tag{6.7}$$

in the limit of large area. An interesting generalization is to consider Wilson loops in antisymmetric tensor representations with q indices where q ranges from 1 to $M-1$. $q=1$ corresponds to the fundamental representation as denoted above, and there is a symmetry under $q \to M-q$ which corresponds to replacing quarks by anti-quarks. These Wilson loops can be thought of as confining strings which connect q probe quarks on one end to q corresponding probe anti-quarks on the other. For $q=M$ the probe quarks combine into a colorless state (a baryon); hence the corresponding Wilson loop does not have an area law.

It is interesting to ask how the tension of this class of confining strings depends on q. If it is a convex function,

$$T_{q+q'} < T_q + T_{q'}, \tag{6.8}$$

then the q-string will not decay into strings with smaller q. This is precisely the situation found by Douglas and Shenker (DS) [62] in softly broken $\mathcal{N}=2$ gauge theory, and later by Hanany *et al.* (HSZ) [63] in the MQCD approach to confining $\mathcal{N}=1$ supersymmetric gauge theory [64, 65]:

$$T_q = \Lambda^2 \sin\frac{\pi q}{M}, \qquad q = 1, 2, \ldots, M-1 \tag{6.9}$$

where Λ is the overall IR scale.

This type of behaviour is also found in the supergravity duals of $\mathcal{N}=1$ gauge theories [66]. Here the confining q-string is described by q coincident fundamental strings placed at $\tau=0$ and oriented along the $\mathbb{R}^{3,1}$.[5] In the deformed conifold solution analyzed above both F_5 and B_2 vanish at $\tau=0$, but it is important that there are M units of F_3 flux through the $\mathbf{S}^3$. In fact, this R-R flux blows up the q fundamental strings into a D3-brane wrapping an $\mathbf{S}^2$ inside the $\mathbf{S}^3$. Although the blow-up can be shown directly, for brevity we build on a closely related result of Bachas *et al.* [68]. In the S-dual of

[5] Qualitatively similar confining flux-tubes were examined in [67] where the authors use the near horizon geometry of non-extremal D3-branes to model confinement.

our type IIB gravity model, at $\tau=0$ we find the $\mathbb{R}^{3,1}\times\mathbf{S}^3$ geometry with M units of NS-NS H_3 flux through the $\mathbf{S}^3$ and q coincident D1-branes along the $\mathbb{R}^{3,1}$. T-dualizing along the D1-brane direction we find q D0-branes on an $\mathbf{S}^3$ with M units of NS-NS flux. This geometry is very closely related to the setup of [68] whose authors showed that the q D0-branes blow up into an $\mathbf{S}^2$. We will find the same phenomenon, but our probe brane calculation is somewhat different from [68] because the radius of our $\mathbf{S}^3$ is different.

After applying S-duality to the KS solution, at $\tau=0$ the metric is

$$\frac{\varepsilon^{4/3}}{2^{1/3}a_0^{1/2}g_s^2M\alpha'}\mathrm{d}x_n\mathrm{d}x_n+bM\alpha'\left(\mathrm{d}\psi^2+\sin^2\psi\mathrm{d}\Omega_2^2\right), \tag{6.10}$$

where $b=2a_0^{1/2}6^{-1/3}\approx 0.93266$. We are now using the standard round metric on $\mathbf{S}^3$ so that ψ is the azimuthal angle ranging from 0 to π. The NS-NS 2-form field at $\tau=0$ is

$$B_2=M\alpha'\left(\psi-\frac{\sin(2\psi)}{2}\right)\sin\theta\mathrm{d}\theta\wedge\mathrm{d}\phi, \tag{6.11}$$

while the world volume field is

$$F=-\frac{q}{2}\sin\theta\mathrm{d}\theta\wedge\mathrm{d}\phi. \tag{6.12}$$

Following [68] closely we find that the tension of a D3-brane which wraps an $\mathbf{S}^2$ located at the azimuthal angle ψ is

$$\frac{\epsilon^{4/3}}{12^{1/3}\pi^2g_s^2\alpha'^2b}\left[b^2\sin^4\psi+\left(\psi-\frac{\sin(2\psi)}{2}-\frac{\pi q}{M}\right)^2\right]^{1/2}. \tag{6.13}$$

Minimizing with respect to ψ we find

$$\psi-\frac{\pi q}{M}=\frac{1-b^2}{2}\sin(2\psi). \tag{6.14}$$

The tension of the wrapped brane is given in terms of the solution of this equation by

$$T_q=\frac{\epsilon^{4/3}}{12^{1/3}\pi^2g_s^2\alpha'^2}\sin\psi\sqrt{1+(b^2-1)\cos^2\psi}. \tag{6.15}$$

Note that under $q\to M-q$, we find $\psi\to\pi-\psi$, so that $T_{M-q}=T_q$. This is a crucial property needed for the connection with the q-strings of the gauge theory.

Although (6.14) is not exactly solvable, we note that $(1-b^2)/2 \approx 0.06507$ is small numerically. If we ignore the RHS of this equation, then $\psi \approx \pi q/M$ and

$$T_q \sim \sin \frac{\pi q}{M} \cdot \tag{6.16}$$

The deviations from this formulae are small: even when $\psi = \pi/4$ and correspondingly $q \approx M/4$, the tension in the KS case is approximately 96.7% of that in the $b = 1$ case.

It is interesting to compare (6.16) with the naive string tension (6.3) we obtained in the previous section. In the large M limit, we expect interactions among the strings to become negligible and the q-string tension to become just q times the ordinary string tension (6.3). Indeed, we find that $g_{\mathrm{s}} T_q = q T_{\mathrm{s}}$ in the large M limit. The extra g_{s} appears because we have been computing tensions in the dual background. When we S-dualize back to the original background with RR-flux and q F-strings, all the tensions are multiplied by g_{s}.

An analogous calculation for the MN background [57] proceeds almost identically. In this background only the F_3 flux is present; hence after the S-duality we find only $H_3 = \mathrm{d}B_2$. The value of B_2 at the minimal radius is again given by (6.11). There is a subtle difference however from the calculation for the KS background in that now the parameter b entering the radius of the $\mathbf{S}^3$ is equal to 1. This simplifies the probe calculation and makes it identical to that of [68]. In particular, now we find

$$\frac{T_q}{T_{q'}} = \frac{\sin \frac{\pi q}{M}}{\sin \frac{\pi q'}{M}}, \tag{6.17}$$

without making any approximations.

Our argument applied to the MN background leads very simply to the DS–HSZ formula for the ratios of q-string tensions (6.17). As we have shown earlier, this formula also holds approximately for the KS background. It is interesting to note that recent lattice simulations in non-supersymmetric pure glue gauge theory [69] appear to yield good agreement with (6.17).

6.3 Chiral symmetry breaking and gluino condensation

Our $SU(N+M) \times SU(N)$ field theory has an anomaly-free $\mathbb{Z}_{2\mathrm{M}}$ R-symmetry. In Section 3 we showed that the corresponding symmetry of the UV (large τ) limit of the metric is

$$\psi \to \psi + \frac{2\pi k}{M}, \qquad k = 1, 2, \ldots, M. \tag{6.18}$$

Recalling that ψ ranges from 0 to 4π, we see that the full solution, which depends on ψ through $\cos\psi$ and $\sin\psi$, has the $\mathbb{Z}_2$ symmetry generated by $\psi \to \psi + 2\pi$. As a result, there are M inequivalent vacua: there are exactly M different discrete orientations of the solution, corresponding to breaking of the $\mathbb{Z}_{2\mathrm{M}}$ UV symmetry through the IR effects. The domain walls constructed out of the wrapped D5-branes separate these inequivalent vacua.

Let us consider domain walls made of k D5-branes wrapped over the finite-sized $\mathbf{S}^3$ at $\tau = 0$, with remaining directions parallel to $\mathbb{R}^{3,1}$. Such a domain wall is obviously a stable object in the KS background and crossing it takes us from one ground state of the theory to another. Indeed, the wrapped D5-brane produces a discontinuity in $\int_B F_3$, where B is the cycle dual to the $\mathbf{S}^3$. If to the left of the domain wall $\int_B F_3 = 0$, as in the basic solution derived in the preceding sections, then to the right of the domain wall

$$\int_B F_3 = 4\pi^2\alpha' k, \tag{6.19}$$

as follows from the quantization of the D5-brane charge. The B-cycle is bounded by a 2-sphere at $\tau = \infty$, hence $\int_B F_3 = \int_{\mathbf{S}^2} \Delta C_2$. Therefore from (3.9) it is clear that to the right of the wall

$$\Delta C_2 \to \pi\alpha' k \omega_2 \tag{6.20}$$

for large τ. This change in C_2 is produced by the $\mathbb{Z}_{2\mathrm{M}}$ transformation (6.18) on the original field configuration (4.1).

It is expected that flux tubes can end on these domain walls [70]. Indeed, a fundamental string can end on the wrapped D5-brane. Also, baryons can dissolve in them. By studying a probe D5-brane in the metric, we find that the domain wall tension is

$$T_{\text{wall}} \sim \frac{1}{g_{\mathrm{s}}} \frac{\varepsilon^2}{(\alpha')^3} \cdot \tag{6.21}$$

In supersymmetric gluodynamics the breaking of chiral symmetry is associated with the gluino condensate $\langle\lambda\lambda\rangle$. A holographic calculation of the condensate was carried out by Loewy and Sonnenschein in [71] (see also [72] for previous work on gluino condensation in conifold theories.) They looked for the deviation of the complex 2-form field $C_2 - \frac{i}{g_{\mathrm{s}}} B_2$ from its asymptotic large τ form that enters the KT solution:

$$\delta\left(C_2 - \frac{i}{g_{\mathrm{s}}} B_2\right) \sim \frac{M\alpha'}{4} \tau \mathrm{e}^{-\tau} [g_1 \wedge g_3 + g_2 \wedge g_4 - i(g_1 \wedge g_2 - g_3 \wedge g_4)]$$

$$\sim \frac{M\alpha'\varepsilon^2}{r^3}\ln(r/\varepsilon^{2/3})\mathrm{e}^{i\psi}(\mathrm{d}\theta_1 - i\sin\theta_1\mathrm{d}\phi_1)\wedge(\mathrm{d}\theta_2 - i\sin\theta_2\mathrm{d}\phi_2). \qquad (6.22)$$

In a space-time that approaches AdS_5 a perturbation that scales as r^{-3} corresponds to the expectation value of a dimension 3 operator. The presence of an extra $\ln(r/\varepsilon^{2/3})$ factor is presumably due to the fact that the asymptotic KT metric differs from AdS_5 by such logarithmic factors. From the angular dependence of the perturbation we see that the dual operator is $SU(2)\times SU(2)$ invariant and carries R-charge 1. These are precisely the properties of $\lambda\lambda$. Thus, the holographic calculation tells us that

$$\langle\lambda\lambda\rangle \sim M\frac{\varepsilon^2}{(\alpha')^3}\,\cdot \qquad (6.23)$$

Thus, the parameter ε^2 which enters the deformed conifold equation has a dual interpretation as the gluino condensate[6].

I.R.K. is grateful to S. Gubser, N. Nekrasov, M. Strassler, A. Tseytlin and E. Witten for collaboration on parts of the material reviewed in these notes and for useful input. This work was supported in part by the NSF grant PHY-9802484.

References

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

[2] S.S. Gubser, I.R. Klebanov and A.M. Polyakov, *Phys. Lett. B* **428** (1998) 105 [hep-th/9802109].

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

[4] S. Kachru and E. Silverstein, *Phys. Rev. Lett.* **80** (1998) 4855 [hep-th/9802183]; A. Lawrence, N. Nekrasov and C. Vafa, *Nucl. Phys. B* **533** (1998) 199 [hep-th/9803015].

[5] A. Kehagias,*Phys. Lett. B* **435** (1998) 337 [hep-th/9805131].

[6] I.R. Klebanov and E. Witten, *Nucl. Phys. B* **536** (1998) 199 [hep-th/9807080].

[7] Non-Spherical Horizons, I, *Adv. Theor. Math. Phys.* **3** (1999) 1 [hep-th/9810201].

[8] S.S. Gubser and I.R. Klebanov, *Phys. Rev. D* **58** (1998) 125025 [hep-th/9808075].

[9] I.R. Klebanov and N. Nekrasov, *Nucl. Phys. B* **574** (2000) 263 [hep-th/9911096].

[10] I.R. Klebanov and A. Tseytlin, *Nucl. Phys. B* **578** (2000) 123 [hep-th/0002159].

[11] I.R. Klebanov and M. Strassler, *JHEP* **0008** (2000) 052 [hep-th/0007191].

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

[13] I.R. Klebanov, *TASI Lectures: Introduction to the AdS/CFT Correspondence* [hep-th/0009139].

[14] S.S. Gubser, I.R. Klebanov and A.W. Peet, *Phys. Rev. D* **54** (1996) 3915 [hep-th/9602135].

[6] It would be nice to understand the relative factor of $g_{\mathrm{s}}M$ between T_{wall} and $\langle\lambda\lambda\rangle$.

[15] S.S. Gubser, *Phys. Rev. D* **59** (1999) 025006 [hep-th/9807164].

[16] M.R. Douglas and G. Moore, *D-branes, quivers, and ALE instantons* [hep-th/9603167].

[17] L. Romans, *Phys. Lett. B* **153** (1985) 392.

[18] P. Candelas and X. de la Ossa, *Nucl. Phys. B* **342** (1990) 246.

[19] A. Ceresole, G. Dall'Agata, R. D'Auria and S. Ferrara, *Phys. Rev. D* **61** (2000) 066001 [hep-th/9905226].

[20] D. Jatkar and S. Randjbar-Daemi, *Phys. Lett. B* **460** (1999) 281 [hep-th/9904187].

[21] I.R. Klebanov and E. Witten, *Nucl. Phys. B* **556** (1999) 89 [hep-th/9905104].

[22] H.J. Kim, L.J. Romans and P. van Nieuwenhuizen, *Phys. Rev. D* **32** (1985) 389; M. Günaydin and N. Marcus, *Class. Quant. Grav.* **2** (1985) L11.

[23] S. Mukhi and N. Suryanarayana, *Stable Non-BPS States and Their Holographic Duals* [hep-th/0011185].

[24] E. Witten, *JHEP* **9807** (1998) 006 [hep-th/9805112].

[25] D.J. Gross and H. Ooguri, *Phys. Rev. D* **58** (1998) 106002 [hep-th/9805129].

[26] C. Bachas, M. Douglas and M. Green, *JHEP* **9707** (1997) 002 [hep-th/9705074].

[27] U. Danielsson, G. Ferretti and I.R. Klebanov, *Phys. Rev. Lett.* **79** (1997) 1984 [hep-th/9705084].

[28] E.G. Gimon and J. Polchinski, *Phys. Rev. D* **54** (1996) 1667 [hep-th/9601038].

[29] M.R. Douglas, *JHEP* **007** (1997) 004 [hep-th/9612126].

[30] R. Minasian and D. Tsimpis, *Nucl. Phys. B* **572** (2000) 499 [hep-th/9911042].

[31] C.P. Herzog, I.R. Klebanov and P. Ouyang, *Remarks on the warped deformed conifold* [hep-th/0108101].

[32] S.B. Giddings, S. Kachru and J. Polchinski, *Hierarchies from fluxes in string compactifications* [hep-th/0105097].

[33] A. Buchel, *Nucl. Phys. B* **600** (2001) 219 [hep-th/0011146]; A. Buchel, C.P. Herzog, I.R. Klebanov, L. Pando Zayas and A.A. Tseytlin, *JHEP* **0104** (2001) 033 [hep-th/0102105]; S.S. Gubser, C.P. Herzog, I.R. Klebanov and A.A. Tseytlin, *JHEP* **0105** (2001) 028 [hep-th/0102172].

[34] J. Polchinski, *Int. J. Mod. Phys. A* **16** (2001) 707 [hep-th/0011193].

[35] A.W. Peet and J. Polchinski, *Phys. Rev. D* **59** (1999) 065011 [hep-th/9809022].

[36] L. Susskind and E. Witten, *The holographic bound in anti-de Sitter space* [hep-th/9805114].

[37] M. Shifman and A. Vainshtein, *Nucl. Phys. B* **277** (1986) 456.

[38] V.A. Novikov, M.A. Shifman, A.I. Vainshtein and V.I. Zakharov, *Nucl. Phys. B* **229** (1983) 381.

[39] P.C. Argyres, Lecture Notes.

[40] N. Seiberg, *Nucl. Phys. B* **435** (1995) 129 [hep-th/9411149].

[41] S.L. Adler and W.A. Bardeen, *Phys. Rev.* **182** (1969) 1517.

[42] I.R. Klebanov, P. Ouyang and E. Witten, *Phys. Rev. D* **65** (2002) 105007 [hep-th/0202056].

[43] M. Bertolini, P. Di Vecchia, M. Frau, A. Lerda and R. Marotta, *Nucl. Phys. B* **621** (2002) 157 [hep-th/0107057].

[44] M. Bertolini, P. Di Vecchia, M. Frau, A. Lerda and R. Marotta, *More Anomalies from Fractional Branes* [hep-th/0202195].

[45] O. Aharony, S. Kachru and E. Silverstein, *Nucl. Phys. B* **488** (1997) 159 [hep-th/9610205].

[46] M. Bianchi, O. DeWolfe, D.Z. Freedman and K. Pilch, *JHEP* **0101** (2001) 021 [hep-th/0009156].

[47] A. Brandhuber and K. Sfetsos, *JHEP* **0012** (2000) 014 [hep-th/0010048].

[48] M. Bianchi, D.Z. Freedman and K. Skenderis, *Holographic Renormalization* [hep-th/0112119].

[49] M. Krasnitz, paper to appear.

[50] M. Henningson and K. Skenderis, *JHEP* **9807** 023 (1998) [hep-th/9806087].

[51] K. Ohta and T. Yokono, *JHEP* **0002** (2000) 023 [hep-th/9912266].

[52] L. Pando Zayas and A. Tseytlin, *JHEP* **0011** (2000) 028 [hep-th/0010088].

[53] M. Grana and J. Polchinski, *Phys. Rev. D* **63** (2001) 026001 [hep-th/0009211].

[54] S.S. Gubser, *Supersymmetry and F-theory realization of the deformed conifold with 3-form flux* [hep-th/0010010].

[55] K. Becker and M. Becker, *Nucl. Phys. B* **477** (1996) 155 [hep-th/9605053].

[56] M. Cvetič, H. Lü and C.N. Pope, *Nucl. Phys. B* **600** (2001) 103 [hep-th/0011023]; M. Cvetič, G.W. Gibbons, H. Lü and C.N. Pope, *Ricci-flat Metrics, Harmonic Forms and Brane Resolutions* [hep-th/0012011].

[57] J. Maldacena and C. Nunez, *Phys. Rev. Lett.* **86** (2001) 588 [hep-th/0008001].

[58] C. Vafa, *Superstrings and Topological Strings at Large N* [hep-th/0008142].

[59] K. Dasgupta, K.h. Oh, J. Park and R. Tatar, *JHEP* **0201** (2002) 031 [hep-th/0110050].

[60] J. Maldacena, *Phys. Rev. Lett.* **80** (1998) 4859 [hep-th/9803002].

[61] S.J. Rey and J. Yee, *Eur. Phys. J. C* **22** (2001) 379 [hep-th/9803001].

[62] M.R. Douglas and S.H. Shenker, *Nucl. Phys. B* **447** (1995) 271 [hep-th/9503163].

[63] A. Hanany, M.J. Strassler and A. Zaffaroni, *Nucl. Phys. B* **513** (1998) 87 [hep-th/9707244].

[64] E. Witten, *Nucl. Phys. B* **500** (1997) 3 [hep-th/9703166].

[65] K. Hori, H. Ooguri and Y. Oz, *Adv. Theor. Math. Phys.* **1** (1998) 1 [hep-th/9706082]; E. Witten, *Nucl. Phys. B* **507** (1997) 658 [hep-th/9706109]; A. Brandhuber, N. Itzhaki, V. Kaplunovsky, J. Sonnenschein and S. Yankielowicz, *Phys. Lett. B* **410** (1997) 27 [hep-th/9706127].

[66] C.P. Herzog and I.R. Klebanov, *Phys. Lett. B* **526** (2002) 388 [hep-th/0111078].

[67] C.G. Callan, A. Guijosa, K.G. Savvidy and O. Tafjord, *Nucl. Phys. B* **555** (1999) 183 [hep-th/9902197].

[68] C. Bachas, M. Douglas and C. Schweigert, *JHEP* **0005** (2000) 048 [hep-th/0003037].

[69] L. Del Debbio, H. Panagopoulos, P. Rossi and E. Vicari, *k-string tensions in $SU(N)$ gauge theories* [hep-th/0106185]; B. Lucini and M. Teper, *Phys. Rev. D* **64** 105019 (2001) 105019 [hep-lat/0107007].

[70] G. Dvali and M. Shifman, *Phys. Lett. B* **396** (1997) 64 [hep-th/9612128].

[71] A. Loewy and J. Sonnenschein, *JHEP* **0108** (2001) 007 [hep-th/0103163].

[72] F. Bigazzi, L. Girardello and A. Zaffaroni, *Nucl. Phys. B* **598** (2001) 530 [hep-th/0011041].

LECTURE 6

DE SITTER SPACE

A. STROMINGER

Department of Physics,
Harvard University, Cambridge,
MA 02138, U.S.A.

Contents

DE SITTER SPACE

M. Spradlin[1,2], A. Strominger[1] and A. Volovich[1,3]

Abstract

These lectures present an elementary discussion of some background material relevant to the problem of de Sitter quantum gravity. The first two lectures discuss the classical geometry of de Sitter space and properties of quantum field theory on de Sitter space, especially the temperature and entropy of de Sitter space. The final lecture contains a pedagogical discussion of the appearance of the conformal group as an asymptotic symmetry group, which is central to the dS/CFT correspondence. A (previously lacking) derivation of asymptotically de Sitter boundary conditions is also given.

1 Introduction

We begin these lectures with one of our favorite equations

$$S = \frac{A}{4G} \cdot \tag{1.1}$$

This is the Bekenstein-Hawking area-entropy law, which says that the entropy S associated with an event horizon is its area A divided by $4G$, where G is Newton's constant [1, 2]. This is a macroscopic formula. It should be viewed in the same light as the macroscopic thermodynamic formulae that were first studied in the 18th and 19th centuries. It describes how properties of event horizons in general relativity change as their parameters are

We would like to thank C. Bachas, A. Bilal, F. David, M. Douglas, and N. Nekrasov for organizing a very pleasant and productive summer school and for arranging financial support. This work was supported in part by DOE grant DE-FG02-91ER40655. M.S. is also supported by DOE grant DE-FG02-91ER40671, and A.V. is also supported by INTAS-OPEN-97-1312. We are grateful to R. Bousso and A. Maloney for useful discussions, and to C. Herzog and L. McAllister for comments on the manuscript.

[1]Harvard University, Cambridge, MA, U.S.A.
[2]Princeton University, Princeton NJ, U.S.A.
[3]L.D. Landau Institute for Theoretical Physics, Moscow, Russia

varied. This behavior can be succinctly summarized by ascribing to them an entropy given by (1.1).

One of the surprising and impressive features of this formula is its universality. It applies to all kinds of black holes with all kinds of charges, shapes and rotation, as well as to black strings and to all of the strange new objects we've found in string theory. It also applies to cosmological horizons, like the event horizon in de Sitter space [3].

After Boltzmann's work we tend to think of entropy in microscopic statistical terms as something which counts the number of microstates of a system. Such an interpretation for the entropy (1.1) was not given at the time that the law was discovered in the early 70s. A complete understanding of this law, and in particular of the statistical origin of this law, is undoubtedly one of the main keys to understanding what quantum gravity is and what the new notions are that replace space and time in quantum gravity.

There has been some definite but still limited progress in understanding the microscopic origin of (1.1) in very special cases of black holes which can be embedded into string theory [4]. That little piece of (1.1) that we have managed to understand has led to all kinds of interesting insights, ultimately culminating in the AdS/CFT correspondence [5]. Nevertheless the progress towards a complete understanding of (1.1) is still very limited, because we only understand special kinds of black holes–among which Schwarzschild black holes are not included–and we certainly don't understand much about cosmological event horizons, such as the horizon in de Sitter space.

In some ways cosmological horizons are much more puzzling than black hole horizons because in the black hole case one may expect that the black hole is a localized object with some quantum microstates. Then if you could provide the correct description of that localized object, you would be able to count those microstates and compare your result to the Bekenstein-Hawking formula and see that they agree. In some stringy cases this agreement has been achieved. On the other hand in de Sitter space the event horizon is observer dependent, and it is difficult even to see where the quantum microstates that we would like to count are supposed to be.

Why has there been significant progress in understanding black hole entropy, but almost no progress in understanding the entropy of de Sitter space? One reason is that one of the principal tools we've used for understanding black hole entropy is supersymmetry. Black holes can be supersymmetric, and indeed the first black holes whose entropy was counted microscopically were supersymmetric. Since then we've managed to creep away from the supersymmetric limit a little bit, but not very far, and certainly we never managed to get all the way to Schwarzschild black holes. So supersymmetry is a crutch that we will need to throw away before we can do anything about de Sitter space. Indeed there is a very simple observation [6] that de Sitter space is inconsistent with supersymmetry in the sense that there is no supergroup that includes the isometries of de Sitter space and has unitary representations[1]. A second, related, obstacle to progress in understanding de Sitter space is that so far we have not been able to embed it in a fully satisfactory manner into string theory.

While the importance of understanding de Sitter quantum gravity has been evident for decades, it has recently been receiving more attention [7–42]. One reason for this is the recent astronomical observations which indicate that the cosmological constant in our universe is positive [43–46]. A second reason is that recent progress in string theory and black holes provides new tools and suggests potentially fruitful new angles. So perhaps de Sitter quantum gravity is a nut ready to be cracked. These lectures are mostly an elementary discussion of the background material relevant to the problem of de Sitter quantum gravity. The classical geometry of de Sitter space is described in Section 2. Scalar quantum field theory in a fixed de Sitter background is in Section 3. Finally, in Section 4 we turn to some recent work on de Sitter quantum gravity. A pedagogical derivation is given of the appearance of the the two dimensional conformal group in three dimensional de Sitter space, which leads to the dS/CFT correspondence [27]. This section also contains a derivation, missing in previous treatments, of the asymptotically de Sitter boundary conditions on the metric. The Appendix contains a calculation of the asymptotic form of the Brown-York stress tensor.

2 Classical geometry of de Sitter space

The d-dimensional de Sitter space dS_d may be realized as the hypersurface described by the equation

$$-X_0^2 + X_1^2 + \cdots + X_d^2 = \ell^2 \tag{2.1}$$

in flat d+1-dimensional Minkowski space $\mathcal{M}^{d,1}$, where ℓ is a parameter with units of length called the de Sitter radius. This hypersurface in flat Minkowski space is a hyperboloid, as shown in Figure 1.

The de Sitter metric is the induced metric from the standard flat metric on $\mathcal{M}^{d,1}$. The embedding (2.1) is a nice way of describing de Sitter space because the $O(d,1)$ symmetry, which is the isometry group of dS_d, is manifest. Furthermore one can show that dS_d is an Einstein manifold with positive scalar curvature, and the Einstein tensor satisfies

$$G_{ab} + \Lambda g_{ab} = 0, \tag{2.2}$$

[1] See, however [7].

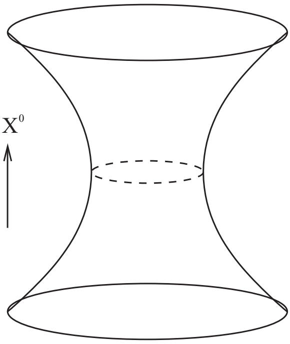


Fig. 1. Hyperboloid illustrating de Sitter space. The dotted line represents an extremal volume S^{d-1}.

where

$$\Lambda = \frac{(d-2)(d-1)}{2\ell^2} \tag{2.3}$$

is the cosmological constant. Henceforth we will set $\ell = 1$.

2.1 Coordinate systems and Penrose diagrams

We will now discuss a number of coordinate systems on dS_d which give different insights into the structure of dS_d. We will frequently make use of coordinates on the sphere S^{d-1}, which is conveniently parametrized by setting

$$\begin{aligned}
\omega^1 &= \cos\theta_1, \\
\omega^2 &= \sin\theta_1 \cos\theta_2, \\
&\vdots \\
\omega^{d-1} &= \sin\theta_1 \cdots \sin\theta_{d-2} \cos\theta_{d-1}, \\
\omega^d &= \sin\theta_1 \cdots \sin\theta_{d-2} \sin\theta_{d-1},
\end{aligned} \tag{2.4}$$

where $0 \le \theta_i < \pi$ for $1 \le i < d-1$, but $0 \le \theta_{d-1} < 2\pi$. Then it is clear that $\sum_{i=1}^{d} (\omega^i)^2 = 1$, and the metric on S^{d-1} is

$$\mathrm{d}\Omega_{\mathrm{d-1}}^2 = \sum_{i=1}^{d} (\mathrm{d}\omega^i)^2 = \mathrm{d}\theta_1^2 + \sin^2\theta_1 \mathrm{d}\theta_2^2 + \cdots + \sin^2\theta_1 \cdots \sin^2\theta_{d-2} \mathrm{d}\theta_{d-1}^2. \tag{2.5}$$

a. Global coordinates (τ, θ_i). This coordinate system is obtained by setting

$$X^0 = \sinh\tau,$$

$$X^i = \omega^i \cosh\tau, \qquad i = 1, \dots, d, \tag{2.6}$$

where $-\infty < \tau < \infty$ and the ω^i are as in (2.4). It is not hard to check that these satisfy (2.1) for any point (τ, ω_i).

From the flat metric on $\mathcal{M}^{d,1}$

$$\mathrm{d}s^2 = -\mathrm{d}X_0^2 + \mathrm{d}X_1^2 + \cdots + \mathrm{d}X_{\mathrm{d}}^2, \tag{2.7}$$

plugging in (2.6) we obtain the induced metric on dS_d,

$$\mathrm{d}s^2 = -\mathrm{d}\tau^2 + (\cosh^2\tau)\mathrm{d}\Omega_{d-1}^2. \tag{2.8}$$

In these coordinates dS_d looks like a $d-1$-sphere which starts out infinitely large at $\tau = -\infty$, then shrinks to a minimal finite size at $\tau = 0$, then grows again to infinite size as $\tau \to +\infty$.

b. Conformal coordinates (T, θ_i). These coordinates are related to the global coordinates by

$$\cosh\tau = \frac{1}{\cos T}, \tag{2.9}$$

so that we have $-\pi/2 < T < \pi/2$. The metric in these coordinates takes the form

$$\mathrm{d}s^2 = \frac{1}{\cos^2 T}(-\mathrm{d}T^2 + \mathrm{d}\Omega_{\mathrm{d}-1}^2). \tag{2.10}$$

This is a particularly useful coordinate system because it enables us to understand the causal structure of de Sitter space. If a geodesic is null with respect to the metric (2.10), then it is also null with respect to the conformally related metric

$$\mathrm{d}\tilde{s}^2 = (\cos^2 T)\mathrm{d}s^2 = -\mathrm{d}T^2 + \mathrm{d}\Omega_{d-1}^2. \tag{2.11}$$

So from the point of view of analyzing what null geodesics do in dS_d we are free to work with the metric (2.11), which looks simpler than (2.10).

The Penrose diagram 2 contains all the information about the causal structure of dS_d although distances are highly distorted. In this diagram each point is actually an S^{d-2} except for points on the left or right sides, which lie on the north or south pole respectively. Light rays travel at 45° angles in this diagram, while timelike surfaces are more vertical than horizontal and spacelike surfaces are more horizontal than vertical.

The surfaces marked $\mathcal{I}^-$, $\mathcal{I}^+$ are called past and future null infinity. They are the surfaces where all null geodesics originate and terminate. Note that a light ray which starts at the north pole at $\mathcal{I}^-$ will exactly reach the south pole by the time it reaches $\mathcal{I}^+$ infinitely far in the future.

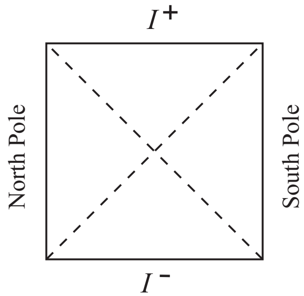


Fig. 2. Penrose diagram for dS_d. The north and south poles are timelike lines, while every point in the interior represents an S^{d-2}. A horizontal slice is an S^{d-1}. The dashed lines are the past and future horizons of an observer at the south pole. The conformal time coordinate T runs from $-\pi/2$ at $\mathcal{I}^-$ to $+\pi/2$ at $\mathcal{I}^+$.

One of the peculiar features of de Sitter space is that no single observer can access the entire spacetime. We often get into trouble in physics when we try to describe more than we are allowed to observe–position and momentum in quantum mechanics, for example. Therefore in attempting to develop de Sitter quantum gravity we should be aware of what can and cannot be observed. A classical observer sitting on the south pole will never be able to observe anything past the diagonal line stretching from the north pole at $\mathcal{I}^-$ to the south pole at $\mathcal{I}^+$. This region is marked as $\mathcal{O}^-$ in Figure 3. This is qualitatively different from Minkowski space, for example, where a timelike observer will eventually have the entire history of the universe in his/her past light cone.

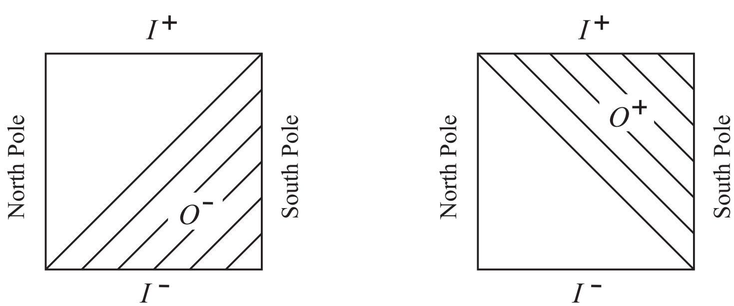


Fig. 3. These diagrams show the regions $\mathcal{O}^-$ and $\mathcal{O}^+$ corresponding respectively to the causal past and future of an observer at the south pole.

Also shown in Figure 3 is the region $\mathcal{O}^+$, which is the only part of de Sitter space that an observer on the south pole will ever be able to send a message to. The intersection $\mathcal{O}^+ \cap \mathcal{O}^-$ is called the (southern) causal diamond. It is this region that is fully accessible to the observer on the south pole. For example if she/he wishes to know the weather anywhere in the southern diamond, a query can be sent to the appropriately located weather station and the response received before $\mathcal{I}^+$ is reached. This is not possible in the lower diamond of $\mathcal{O}^-$, to which a query can never be sent, or the upper diamond of $\mathcal{O}^+$, from which a response cannot be received. The northern diamond on the left of 3 is completely inaccessible to an observer on the south pole.

c. Planar coordinates (t, x^i), $i = 1, \ldots, d-1$. To define this coordinate system we take

$$\begin{aligned} X^0 &= \sinh t - \frac{1}{2} x_i x^i \mathrm{e}^{-t}, \\ X^i &= x^i \mathrm{e}^{-t}, \qquad i = 1, \ldots, d-1, \\ X^d &= \cosh t - \frac{1}{2} x_i x^i \mathrm{e}^{-t}. \end{aligned} \tag{2.12}$$

The metric then takes the form

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + \mathrm{e}^{-2t} \mathrm{d}x_i \mathrm{d}x^i. \tag{2.13}$$

These coordinates do not cover all of de Sitter space, but only the region $\mathcal{O}^-$ and are therefore appropriate for an observer on the south pole. The slices of constant t are illustrated in Figure 4.

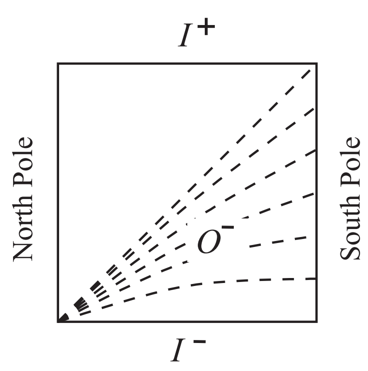


Fig. 4. The dashed lines are slices of constant t in planar coordinates. Note that each slice is an infinite flat $d-1$-dimensional plane which extends all the way down to $\mathcal{I}^-$.

The surfaces of constant t are spatial sections of de Sitter space which are infinite volume $d-1$-planes with the flat metric. From the diagram it is clear that every surface of constant t intersects $\mathcal{I}^-$ at the north pole. It may seem puzzling–and is certainly one of the salient features of de Sitter space–that a spatial plane can make it to the infinite past. This happens because $\mathcal{I}^-$ is very large, and you can get there along a spatial trajectory from anywhere in $\mathcal{O}^-$. In these coordinates the time t is not a Killing vector, and the only manifest symmetries are translations and rotations of the x^i coordinates.

d. Static coordinates (t, r, θ_a), $a = 1, \ldots, d-2$. The t in these coordinates is not the same as the t in planar coordinates, but we are running out of letters! Note also that for these coordinates and the following ones we will need a parametrization of S^{d-2}, not S^{d-1}. This coordinate system is constructed to have an explicit timelike Killing symmetry. If we write

$$\begin{aligned} X^0 &= \sqrt{1-r^2}\sinh t, \\ X^a &= r\omega^a, \qquad a = 1, \ldots, d-1, \\ X^d &= \sqrt{1-r^2}\cosh t, \end{aligned} \tag{2.14}$$

then the metric takes the form

$$\mathrm{d}s^2 = -(1-r^2)\mathrm{d}t^2 + \frac{\mathrm{d}r^2}{1-r^2} + r^2\mathrm{d}\Omega_{d-2}^2. \tag{2.15}$$

In this coordinate system $\partial/\partial t$ is a Killing vector and generates the symmetry $t \to t + \text{constant}$. The horizons are at $r^2 = 1$, and the southern causal diamond has $0 \leq r \leq 1$, with the south pole at $r = 0$.

One of the reasons to want a timelike Killing vector is so that we can use it to define time evolution, or in other words to define the Hamiltonian. But from (2.15) we see that at $r = 1$ the norm of $\partial/\partial t$ vanishes, so that it becomes null. In Figure 5 we illustrate what the Killing vector field $\partial/\partial t$ is doing when extended to the various diamonds of the Penrose diagram. In the top and bottom diamonds, $\partial/\partial t$ is spacelike, while in the northern diamond the vector is pointing towards the past! Thus $\partial/\partial t$ in static coordinates can only be used to define a sensible time evolution in the southern diamond of de Sitter space. The absence of a globally timelike Killing vector in de Sitter space has important implications for the quantum theory.

e. Eddington-Finkelstein coordinates (x^+, r, θ_a). This coordinate system is the de Sitter analog of the (outgoing) Eddington-Finkelstein coordinates for a Schwarzschild black hole. Starting from the static coordinates, we define x^+ by the equation

$$\mathrm{d}x^+ = \mathrm{d}t + \frac{\mathrm{d}r}{1-r^2}, \tag{2.16}$$

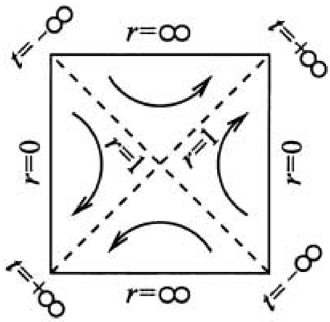


Fig. 5. This Penrose diagram shows the direction of the flow generated by the Killing vector $\partial/\partial t$ in static coordinates. The horizons (dotted lines) are at $r^2 = 1$, and the southern causal diamond is the region with $0 \leq r \leq 1$ on the right hand side. Past and future null infinity $\mathcal{I}^{\pm}$ are at $r = \infty$.

which we can solve to obtain

$$x^+ = t + \frac{1}{2}\ln\frac{1+r}{1-r}\,\cdot \tag{2.17}$$

In these coordinates the metric is

$$\mathrm{d}s^2 = -(1-r^2)(\mathrm{d}x^+)^2 + 2\mathrm{d}x^+\mathrm{d}r + r^2\mathrm{d}\Omega^2_{d-2}. \tag{2.18}$$

The same symmetries are manifest in this coordinate system as in the static coordinates since $\partial/\partial t$ at fixed r is the same as $\partial/\partial x^+$ at fixed r. Lines of constant x^+ are the null lines connecting $\mathcal{I}^-$ with the south pole depicted in Figure 3. These coordinates cover the causal past $\mathcal{O}^-$ of an observer at the south pole while still keeping the symmetry manifest.

We can also define

$$x^- = t - \frac{1}{2}\ln\frac{1+r}{1-r}, \tag{2.19}$$

so that the metric takes the form

$$\mathrm{d}s^2 = -(1-r^2(x^+,x^-))\mathrm{d}x^+\mathrm{d}x^- + r^2\mathrm{d}\Omega^2_{d-2}, \tag{2.20}$$

where $r = \tanh\frac{x^+-x^-}{2}$.

f. Kruskal coordinates (U, V, θ_a). Finally we take

$$x^- = \ln U, \quad x^+ = -\ln(-V), \tag{2.21}$$

in which case

$$r = \frac{1+UV}{1-UV}\,. \tag{2.22}$$

Then the metric takes the form

$$\mathrm{d}s^2 = \frac{1}{(1-UV)^2}(-4\mathrm{d}U\mathrm{d}V + (1+UV)^2\mathrm{d}\Omega^2_{d-2}). \tag{2.23}$$

These coordinates cover all of de Sitter space. The north and south poles correspond to $UV = -1$, the horizons correspond to $UV = 0$, and $\mathcal{I}^{\pm}$ correspond to $UV = 1$. The southern diamond is the region with $U > 0$ and $V < 0$.

Exercise 1. Find $X^0, \ldots, X^d$ as functions of U, V, and θ_a for the Kruskal coordinates.

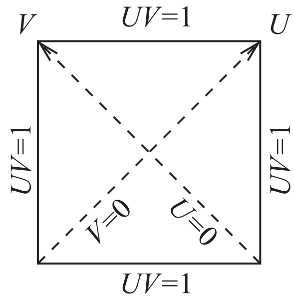


Fig. 6. The Kruskal coordinate system covers all of de Sitter space. In this Penrose diagram the coordinate axes $U = 0$ and $V = 0$ are the horizons, $UV = -1$ are the north and south poles, and $UV = 1$ are $\mathcal{I}^+$ and $\mathcal{I}^-$. The arrows denote the directions of increasing U and V.

g. Hyperbolic coordinates $(\overline{\tau}, \psi, \theta_a)$. Global coordinates foliate de Sitter space with spheres, and planar coordinates foliate de Sitter space with planes. One can also foliate de Sitter space with spaces of constant negative curvature by using the hyperbolic coordinates

$$\begin{aligned} X^0 &= \sinh\overline{\tau}\cosh\psi, \\ X^a &= \omega^a \sinh\overline{\tau}\sinh\psi, \\ X^d &= \cosh\overline{\tau}, \end{aligned} \tag{2.24}$$

in which the metric takes the form

$$\mathrm{d}s^2 = -\mathrm{d}\overline{\tau}^2 + \sinh^2\overline{\tau}(\mathrm{d}\psi^2 + \sinh^2\psi\,\mathrm{d}\Omega^2_{d-2}), \tag{2.25}$$

and surfaces of constant $\overline{\tau}$ are $d{-}1$-dimensional hyperbolic planes.

2.2 Schwarzschild-de Sitter

The simplest generalization of the de Sitter space solution is Schwarzschild-de Sitter, which we abbreviate as SdS. This solution represents a black hole in de Sitter space. In d dimensions in static coordinates the SdS_d metric takes the form

$$\mathrm{d}s^2 = -\left(1 - \frac{2m}{r^{\mathrm{d}-3}} - r^2\right)\,\mathrm{d}t^2 + \frac{1}{1 - \frac{2m}{r^{d-3}} - r^2}\,\mathrm{d}r^2 + r^2\mathrm{d}\Omega^2_{d-2}, \qquad (2.26)$$

where m is a parameter related to the black hole mass (up to some d-dependent normalization constant). In general there are two horizons (recall that these are places where the timelike Killing vector $\partial/\partial t$ becomes null), one of which is the black hole horizon and the other of which is the de Sitter horizon. Note that the two horizons approach each other as m is increased, so that there is a maximum size black hole which can fit inside de Sitter space before the black hole horizon hits the de Sitter horizon.

One reason to introduce SdS is that it plays an important role in the work of Gibbons & Hawking [3] determining the entropy of pure de Sitter space, which will be reviewed in Section 3.3. For this purpose it will be convenient to focus on the three dimensional Schwarzschild-de Sitter solution [49]

$$\mathrm{d}s^2 = -(1 - 8GE - r^2)\mathrm{d}t^2 + \frac{\mathrm{d}r^2}{(1 - 8GE - r^2)} + r^2\mathrm{d}\phi^2, \qquad (2.27)$$

where we have normalized the energy E of the Schwarzschild black hole appropriately for three dimensions. In three dimensions there is only one horizon, at $r_{\mathrm{H}} = \sqrt{1 - 8GE}$, and as E goes to zero this reduces to the usual horizon in empty de Sitter space. The fact that there is only a de Sitter horizon and not a black hole horizon is not surprising in light of the fact that in three dimensional flat space there are no black holes.

We can learn a little more about the solution (2.27) by looking near $r = 0$, where $\mathrm{d}s^2$ behaves like

$$\mathrm{d}s^2 \sim -r_{\mathrm{H}}^2\mathrm{d}t^2 + \frac{\mathrm{d}r^2}{r_{\mathrm{H}}^2} + r^2\mathrm{d}\phi^2. \qquad (2.28)$$

Now we can rescale the coordinates by defining

$$t' = r_{\mathrm{H}}t, \qquad r' = r/r_{\mathrm{H}}, \qquad \phi' = r_{\mathrm{H}}\phi. \qquad (2.29)$$

In the rescaled coordinates the metric (2.28) is simply

$$\mathrm{d}s^2 = -\mathrm{d}t'^2 + \mathrm{d}r'^2 + r'^2\mathrm{d}\phi'^2. \qquad (2.30)$$

This looks like flat space, but it is not quite flat space because while ϕ was identified modulo 2π, ϕ' is identified modulo $2\pi r_{\mathrm{H}}$. Therefore there is a conical singularity with a positive deficit angle at the origin.

You may be familiar with the fact that if you put a point-like mass in flat three dimensional Minkowski space you would also get a conical deficit angle at the location of the particle. Hence we recognize (2.27) as a point-like mass, rather than a black hole, at the south pole of dS_3. If the solution is maximally extended one finds there is also point-like mass of the same size at the north pole [49].

Exercise 2. Show that SdS_3 is a global identification of dS_3.

2.3 Geodesics

Our last topic in the classical geometry of de Sitter space is geodesics. It is clear that if we take two points on the sphere S^n of radius R, then there is only one independent $SO(n+1)$-invariant quantity that we can associate to the two points. That is the geodesic distance D, or equivalently the angle θ between them, which are related by $D = R\theta$. Let us think of the sphere as being embedded in flat Euclidean space, with the embedding equation $\delta_{ij}X^iX^j = R^2$, $i, j = 1, \ldots, n+1$. It is useful to define a quantity P by $R^2P(X,X') \equiv \delta_{ij}X^iX'^j = R^2\cos\theta$.

It is a little harder to visualize, but we can do something similar for dS_d. There we can define

$$P(X,X') = \eta_{ij}X^iX'^j, \qquad \eta_{ij} = \mathrm{diag}(-1,1,\ldots,1) \tag{2.31}$$

(recall that we have set the de Sitter radius ℓ to one). For points in a common causal diamond, this is related to the geodesic distance $D(X,X')$ between X and X' by $P = \cos D$. This quantity P will turn out to be a more convenient invariant to associate to two points in de Sitter space. We can easily write explicit formulas for $P(X,X')$ in the various coordinate systems discussed above. For example, in planar coordinates we have

$$P(t,x^i;t',y^i) = \cosh(t-t') - \frac{1}{2}\mathrm{e}^{-t-t'}\delta_{ij}(x^i-y^i)(x^j-y^j). \tag{2.32}$$

The expression for P is simple in terms of the X's but can get complicated when written in a particular coordinate system.

To conclude, we note a few important properties of P for later use. If $P = 1$, then the geodesic distance is equal to zero, so the two points X and X' coincide or are separated by a null geodesic. We can also consider taking antipodal points $X' = -X$, in which case $P = -1$. In general $P = -1$ when the antipodal point of X lies on the light cone of X'. In general, the geodesic separating X and X' is spacelike for $P < 1$ and timelike for $P > 1$, while for $P < -1$ the geodesic between X and the antipodal point of X' is timelike.

3 Quantum field theory on de Sitter space

Ultimately, a complete understanding of the entropy-area relation (1.1) in de Sitter space will require an understanding of quantum gravity on de Sitter space. In this section we will take a baby step in that direction by considering a single free massive scalar field on a fixed background de Sitter spacetime. This turns out to be a very rich subject which has been studied by many authors [3, 50–60].

3.1 Green functions and vacua

Let us consider a scalar field in dS_d with the action

$$S = -\frac{1}{2}\int \mathrm{d}^d x\sqrt{-g}\left[(\nabla\phi)^2 + m^2\phi^2\right]. \tag{3.1}$$

Since this is a free field theory, all information is encoded in the two-point function of ϕ. We will study the Wightman function

$$G(X,Y) = \langle 0|\phi(X)\phi(Y)|0\rangle, \tag{3.2}$$

which obeys the free field equation

$$(\nabla^2 - m^2)G(X,Y) = 0, \tag{3.3}$$

where ∇^2 is the Laplacian on dS_d.

There are other two point functions that one can discuss: retarded, advanced, Feynman, Hadamard and so on, but these can all be obtained from the Wightman function (3.2), for example by taking the real or imaginary part, and/or by multiplying by a step function in time.

Let us assume that the state $|0\rangle$ in (3.2) is invariant under the $SO(d,1)$ de Sitter group. Then $G(X,Y)$ will be de Sitter invariant, and so at generic points can only depend on the de Sitter invariant length $P(X,Y)$ between X and Y[2]. Writing $G(X,Y) = G(P(X,Y))$, (3.3) reduces to a differential equation in one variable P

$$(1-P^2)\partial_P^2 G - \mathrm{d}P\partial_P G - m^2 G = 0. \tag{3.4}$$

With the change of variable $z = \frac{1+P}{2}$ this becomes a hypergeometric equation

$$z(1-z)G'' + \left(\frac{d}{2} - \mathrm{d}z\right)G' - m^2 G = 0, \tag{3.5}$$

[2] $P(X,Y) = P(Y,X)$ is insensitive to the time ordering between points, which is $SO(d,1)$ (but not $O(d,1)$) invariant. Because of this the $i\epsilon$ prescription for G, as discussed below, can not be written as a function of P alone.

whose solution is

$$G = c_{m,d} F\left(h_+, h_-, \frac{d}{2}, z\right), \tag{3.6}$$

where $c_{m,d}$ is a normalization constant to be fixed shortly, and

$$h_\pm = \frac{1}{2}\left[(d-1) \pm \sqrt{(d-1)^2 - 4m^2}\right]. \tag{3.7}$$

The hypergeometric function (3.6) has a singularity at $z = 1$, or $P = 1$, and a branch cut for $1 < P < \infty$. The singularity occurs when the points X and Y are separated by a null geodesic. At short distances the scalar field is insensitive to the fact that it is in de Sitter space and the form of the singularity is precisely the same as that of the propagator in flat d-dimensional Minkowski space. We can use this fact to fix the normalization constant $c_{m,d}$. Near $z = 1$ the hypergeometric function behaves as

$$F\left(h_+, h_-, \frac{d}{2}, \frac{1+P}{2}\right) \sim \left(\frac{D^2}{4}\right)^{1-d/2} \frac{\Gamma(\frac{d}{2})\Gamma(\frac{d}{2}-1)}{\Gamma(h_+)\Gamma(h_-)}, \tag{3.8}$$

where $D = \cos^{-1} P$ is the geodesic separation between the two points. Comparing with the usual short-distance singularity $\frac{\Gamma(\frac{d}{2})}{2(d-2)\pi^{d/2}}(D^2)^{1-d/2}$ fixes the coefficient to be

$$c_{m,d} = 4^{1-d/2}\frac{\Gamma(h_+)\Gamma(h_-)}{\Gamma(\frac{d}{2})\Gamma(\frac{d}{2}-1)} \times \frac{\Gamma(\frac{d}{2})}{2(d-2)\pi^{d/2}} = \frac{\Gamma(h_+)\Gamma(h_-)}{(4\pi)^{d/2}\Gamma(\frac{d}{2})}\,. \tag{3.9}$$

The prescription for going around the singularity in the complex plane is also the same as in Minkowski space, namely replacing $X^0 - Y^0$ with $X^0 - Y^0 - i\epsilon$.

The equation (3.4) clearly has a $P \to -P$ symmetry, so if $G(P)$ is a solution then $G(-P)$ is also a solution. The second linearly independent solution to (3.4) is therefore

$$F\left(h_+, h_-, \frac{d}{2}, \frac{1-P}{2}\right). \tag{3.10}$$

The singularity is now at $P = -1$, which corresponds to X being null separated from the antipodal point to Y. This singularity sounds rather unphysical at first, but we should recall that antipodal points in de Sitter space are always separated by a horizon. The Green function (3.10) can be thought of as arising from an image source behind the horizon, and (3.10) is nonsingular everywhere within an observer's horizon. Hence the "unphysical" singularity can not be detected by any experiment.

De Sitter space therefore has a one parameter family of de Sitter invariant Green functions G_α corresponding to a linear combination of the solutions (3.6) and (3.10). Corresponding to this one-parameter family of Green functions is a one-parameter family of de Sitter invariant vacuum states $|\alpha\rangle$ such that $G_\alpha(X,Y) = \langle\alpha|\phi(X)\phi(Y)|\alpha\rangle$. These vacua are discussed in detail in [58, 59], but are usually discarded as somehow "unphysical". However, as we try to understand the quantum theory of de Sitter space these funny extra vacua will surely turn out to have some purpose in life.

De Sitter Green functions are often discussed in the context of analytic continuation to the Euclidean sphere. If we work in static coordinates and take $t \to i\tau$, the dS_d metric becomes the metric on the sphere S^d. On the sphere there is a unique Green function, which when analytically continued back to de Sitter space yields (3.6).

Let us say a few more words about the vacuum states. A vacuum state $|0\rangle$ is defined as usual by saying that it is annihilated by all annihilation operators

$$a_n|0\rangle = 0. \tag{3.11}$$

That is, we write an expansion for the scalar field in terms of creation and annihilation operators of the form

$$\phi(X) = \sum_k \left[a_k u_k(X) + a_k^\dagger u_k^*(X)\right], \tag{3.12}$$

where a_k and $a_k^\dagger$ satisfy

$$[a_k, a_l^\dagger] = \delta_{kl}. \tag{3.13}$$

The modes $u_k(X)$ satisfy the wave equation

$$(\nabla^2 - m^2)u_k = 0, \tag{3.14}$$

and are normalized with respect to the invariant Klein-Gordon inner product

$$(u_k, u_l) = -i\int \mathrm{d}\Sigma^\mu \ \left(u_k \overleftrightarrow{\partial}_\mu u_l^*\right) = \delta_{kl}, \tag{3.15}$$

where the integral is taken over a complete spherical spacelike slice in dS_d and the result is independent of the choice of this slice.

The question is, which modes do we associate with creation operators in (3.12) and which do we associate with annihilation operators? In Minkowski space we take positive and negative frequency modes,

$$u \sim \mathrm{e}^{-iEt} f(x), \qquad u^* \sim \mathrm{e}^{iEt} f^*(x), \tag{3.16}$$

respectively to multiply the annihilation and creation operators. But in a general curved spacetime there is no canonical choice of a time variable with respect to which one can classify modes as being positive or negative frequency. If we make a choice of time coordinate, we can get a vacuum state $|0\rangle$ and then the state $(a^\dagger)^n|0\rangle \equiv |n\rangle$ is said to have n particles in it. But if we had made some other choice of time coordinate then we would have a different vacuum $|0'\rangle$, which we could express as a linear combination of the $|n\rangle$'s. Hence the question "How many particles are present?" is not well-defined independently of a choice of coordinates. This is an important and general feature of quantum field theory in curved spacetime.

In order to preserve classical symmetries of dS_d in the quantum theory, we would like to find a way to divide the modes into u and u^* that is invariant under $SO(d,1)$. The resulting vacuum will then be de Sitter invariant. It turns out [51, 58, 59] that there is a family of such divisions, and a corresponding family of Green functions such as G_α.

3.2 Temperature

In this section we will show that an observer moving along a timelike geodesic observes a thermal bath of particles when the scalar field ϕ is in the vacuum state $|0\rangle$. Thus we will conclude that de Sitter space is naturally associated with a temperature [53], which we will calculate.

Since the notion of a particle is observer-dependent in a curved spacetime, we must be careful to give a coordinate invariant characterization of the temperature. A good way to achieve this is to consider an observer equipped with a detector. The detector will have some internal energy states and can interact with the scalar field by exchanging energy, *i.e.* by emitting or absorbing scalar particles. The detector could for example be constructed so that it emits a "bing" whenever its internal energy state changes. All observers will agree on whether or not the detector has binged, although they may disagree on whether the bing was caused by particle emission or absorption. Such a detector is called an Unruh detector and may be modeled by a coupling of the scalar field $\phi(x(\tau))$ along the worldline $x(\tau)$ of the observer to some operator $m(\tau)$ acting on the internal detector states

$$g\int_{-\infty}^{\infty} \mathrm{d}\tau\; m(\tau)\phi(x(\tau)), \tag{3.17}$$

where g is the strength of the coupling and τ is the proper time along the observer's worldline.

Let H denote the detector Hamiltonian, with energy eigenstates $|E_j\rangle$,

$$H|E_j\rangle = E_j|E_j\rangle, \tag{3.18}$$

and let m_{ij} be the matrix elements of the operator $m(\tau)$ at $\tau = 0$:

$$m_{ij} \equiv \langle E_i | m(0) | E_j \rangle \cdot \tag{3.19}$$

We will calculate the transition amplitude from a state $|0\rangle|E_i\rangle$ in the tensor product of the scalar field and detector Hilbert spaces to the state $\langle E_j|\langle\beta|$, where $\langle\beta|$ is any state of the scalar field. To first order in perturbation theory for small coupling g, the desired amplitude is

$$g \int_{-\infty}^{\infty} \mathrm{d}\tau \ \langle E_j | \langle \beta | m(\tau) \phi(x(\tau)) | 0 \rangle | E_i \rangle \cdot \tag{3.20}$$

Using

$$m(\tau) = \mathrm{e}^{iH\tau} m(0) \mathrm{e}^{-iH\tau}, \tag{3.21}$$

this can be written as

$$g m_{ji} \int_{-\infty}^{\infty} \mathrm{d}\tau \ \mathrm{e}^{i(E_j - E_i)\tau} \langle \beta | \phi(x(\tau)) | 0 \rangle \cdot \tag{3.22}$$

Since we are only interested in the probability for the detector to make the transition from E_i to E_j, we should square the amplitude (3.22) and sum over the final state $|\beta\rangle$ of the scalar field, which will not be measured. Using $\sum_\beta |\beta\rangle\langle\beta| = 1$, we find the probability

$$P(E_i \to E_j) = g^2 |m_{ij}|^2 \int_{-\infty}^{\infty} \mathrm{d}\tau \, \mathrm{d}\tau' \ \mathrm{e}^{-i(E_j - E_i)(\tau' - \tau)} G(x(\tau'), x(\tau)), \tag{3.23}$$

where $G(x(\tau'), x(\tau))$ is the Green function (3.2). The Green function is a function only of the geodesic distance $P(x(\tau), x(\tau'))$, and if we consider for simplicity an observer sitting on the south pole, then P is given in static coordinates by $P = \cosh(\tau - \tau')$. Therefore everything inside the integral (3.23) depends only on $\tau - \tau'$ and we get an infinite factor from integrating over $\tau + \tau'$. We can divide out this factor and discuss the transition probability per unit proper time along the detector worldline, which is then given by

$$\dot{P}(E_i \to E_j) = g^2 |m_{ij}|^2 \int_{-\infty}^{\infty} \mathrm{d}\tau \ \mathrm{e}^{-i(E_j - E_i)\tau} G(\cosh \tau). \tag{3.24}$$

The first hint that (3.24) has something to do with a thermal response is that the function G is periodic in imaginary time under $\tau \to \tau + 2\pi i$, and Green functions which are periodic in imaginary time are thermal Green functions.

To investigate the nature of a thermal state, let us suppose it were true (as will be demonstrated shortly) that

$$\dot{P}(E_i \to E_j) = \dot{P}(E_j \to E_i)\mathrm{e}^{-\beta(E_j - E_i)}, \tag{3.25}$$

and that the energy levels of the detector were thermally populated, so that

$$N_i = N\mathrm{e}^{-\beta E_i}, \tag{3.26}$$

where N is some normalization factor. Then it is clear that the total transition rate R from E_i to E_j is the same as from E_j to E_i:

$$R(E_i \to E_j) = N\mathrm{e}^{-\beta E_i}\dot{P}(E_i \to E_j) = R(E_j \to E_i), \tag{3.27}$$

which is the principle of detailed balance in a thermal ensemble. In other words, if the transition probabilities are related by (3.25) and the population of the states is thermal as in (3.26), then there is no change in the probability distribution for the energy levels with time. So (3.25) describes the transition probabilities of a system in a thermal bath of particles at temperature $T = 1/\beta$.

Let us now show that (3.25) holds for the transition probabilities calculated in (3.24). The integrand in (3.24) has singularities in the complex τ-plane at $\tau = 2\pi i n$ for any integer n. Consider integrating the function $\mathrm{e}^{-i(E_j - E_i)\tau}G(\cosh\tau)$ around the contour shown in Figure 7.

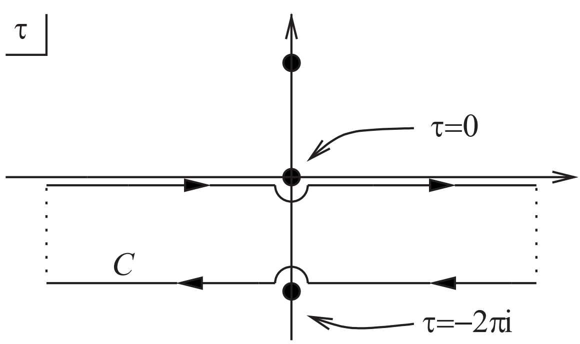


Fig. 7. The integrand in (3.24) has singularities in the complex τ-plane at $\tau = 2\pi i n$ for any integer n. This figure shows the contour C used in the integral (3.28). The dotted lines signify the closure of the contour at infinity.

Since the total integral around this contour is zero, we have

$$\int_{-\infty}^{\infty} \mathrm{d}\tau \; {}^{-i(E_j - E_i)\tau} G(\cosh\tau) + \int_{+\infty - i\beta}^{-\infty - i\beta} \mathrm{d}\tau \; \mathrm{e}^{-i(E_j - E_i)\tau} G(\cosh\tau) = 0, \tag{3.28}$$

where $\beta = 1/2\pi$. The contour in Figure 7 corresponds to the pole prescription for the Wightman function as discussed in Section 3.1. Now redefining the variable of integration in the second integral as $\tau' = -\tau - i\beta$ we get precisely the desired relation (3.25).

Although we performed this calculation only for an observer stationary at the south pole, all timelike geodesics in de Sitter space are related to each other by the $SO(d,1)$ de Sitter isometry group. Since the Green function used in this calculation is de Sitter invariant, the result for the temperature is the same for any observer moving along a timelike geodesic. We conclude that any geodesic observer in de Sitter space will feel that she/he is in a thermal bath of particles at a temperature

$$T_{\rm dS} = \frac{1}{2\pi\ell}, \tag{3.29}$$

where we have restored the factor of the de Sitter radius ℓ by dimensional analysis.

3.3 Entropy

In this subsection we will associate an entropy to de Sitter space. We will restrict our attention to dS_3, where the analysis simplifies considerably.

For the case of black holes one can use similar methods as those in the previous section to calculate the temperature $T_{\rm BH}$ of the black hole. The black hole entropy $S_{\rm BH}$ can then be found by integrating the thermodynamic relation

$$\frac{{\rm d}S_{\rm BH}}{{\rm d}E_{\rm BH}} = \frac{1}{T_{\rm BH}}, \tag{3.30}$$

where $E_{\rm BH}$ is the energy or mass of the black hole. So if you know the value of the temperature just for one value of $E_{\rm BH}$ you will not be able to get the entropy, but if you know it as a function of the black hole mass then you can simply integrate (3.30) to find the entropy. The constant of integration is determined by requiring that a black hole of zero mass has zero entropy.

So for de Sitter space one would expect to use the relation

$$\frac{{\rm d}S_{\rm dS}}{{\rm d}E_{\rm dS}} = \frac{1}{T_{\rm dS}} \tag{3.31}$$

to find the entropy $S_{\rm dS}$. The problem in de Sitter space is that once the coupling constant of the theory is chosen there is just one de Sitter solution, whereas in the black hole case there is a whole one parameter family of solutions labeled by the mass of the black hole, for fixed coupling constant. In other words, what is $E_{\rm dS}$ in (3.31)? One might try to vary the cosmological constant, but that is rather unphysical as it is the coupling constant. One would be going from one theory to another instead of from one configuration in the theory to another configuration in the same theory.

Let us instead follow Gibbons & Hawking [3] and use the one parameter family of Schwarzschild-de Sitter solutions to see how the temperature varies as a function of the parameter E labeling the mass of the black hole.

Exercise 3. The SdS_3 solution in static coordinates is

$$\mathrm{d}s^2 = -(1-8GE-r^2)\mathrm{d}t^2 + \frac{\mathrm{d}r^2}{(1-8GE-r^2)} + r^2\mathrm{d}\phi^2. \tag{3.32}$$

Find a Green function for SdS_3 by analytic continuation from the smooth Euclidean solution. Show that this Green function is periodic in imaginary time with periodicity

$$\tau \to \tau + \frac{2\pi i}{\sqrt{1-8GE}} \cdot \tag{3.33}$$

From the exercise and the discussion in the previous section we conclude that the temperature associated with the Schwarzschild-de Sitter solution is

$$T_{SdS} = \frac{\sqrt{1-8GE}}{2\pi} \cdot \tag{3.34}$$

Using the formula

$$\frac{\mathrm{d}S_{SdS}}{\mathrm{d}E} = \frac{1}{T_{SdS}}, \tag{3.35}$$

and writing the result in terms of the area A_{H} of the de Sitter horizon at $r_{\mathrm{H}} = \sqrt{1-8GE}$ which is given by

$$\sqrt{1-8GE} = \frac{A_{\mathrm{H}}}{2\pi}, \tag{3.36}$$

one finds that the entropy is equal to

$$S_{SdS} = -\frac{A_{\mathrm{H}}}{4G} \cdot \tag{3.37}$$

This differs by a minus sign from the famous formula (1.1)! What did we do wrong? Gibbons and Hawking suggested that to get the de Sitter entropy we should use not (3.35) but instead

$$\frac{\mathrm{d}S_{SdS}}{\mathrm{d}(-E_{dS})} = \frac{1}{T_{SdS}} \cdot \tag{3.38}$$

This looks funny but in fact there is a very good reason for using this new formula.

The de Sitter entropy, although we don't know exactly how to think about it, is supposed to correspond to the entropy of the stuff behind the horizon which we can't observe. Now in general relativity the expression for the energy on a surface is the integral of a total derivative, which reduces to a surface integral on the boundary of the surface, and hence vanishes on any closed surface. Consider a closed surface in de Sitter space such as the one shown in Figure 8. If we put something with positive energy on the south pole, then necessarily there will be some negative energy on the north pole.

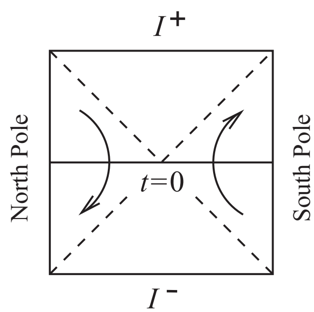


Fig. 8. The energy associated to the Killing vector $\partial/\partial t$ (indicated by the arrows) along the spacelike slice $t = 0$ (solid line) must vanish. If we ascribe positive energy to a positive deficit angle at the south pole, then we must ascribe negative energy to a positive deficit angle at the south pole since the Killing vector $\partial/\partial t$ runs in the opposite direction behind the horizon.

This can be seen quite explicitly in the Schwarzschild-de Sitter solution. With no black hole, the spacelike slice in Figure 8 is an S^2, but we saw in one of the exercises that in the SdS_3 solution there is a positive deficit angle at both the north and south poles. If we ascribe positive energy to the positive deficit angle at the south pole, then because the Killing vector $\partial/\partial t$ used to define the energy changes direction across the horizon, we are forced to ascribe negative energy to the positive deficit angle at the north pole.

Therefore the northern singularity of Schwarzschild-de Sitter behind the horizon actually carries negative energy. In (3.35) we varied with respect to the energy at the south pole, and ended up with the wrong sign in (3.37), but if we more sensibly vary with respect to the energy at the north pole, then we should use the formula (3.38). Then we arrive at the entropy for

Schwarzschild-de Sitter

$$S_{SdS} = \frac{A_{\rm H}}{4G} = \frac{\pi}{2G}\sqrt{1-8GE}. \tag{3.39}$$

The integration constant has been chosen so that the entropy vanishes for the maximal energy $E = \frac{1}{8G}$ at which value the deficit angle is 2π and the space has closed up.

In conclusion we see that the area-entropy law (1.1) indeed applies to three dimensional Schwarzschild-de Sitter.

4 Quantum gravity in de Sitter space

So far we have discussed well established and understood results about classical de Sitter space and quantum field theory in a fixed de Sitter background. Now we turn to the more challenging problem of quantum gravity in de Sitter space, about which little is established or understood.

In this section we will give a pedagogical discussion of several aspects of some recent efforts in this direction [27] (some similar ideas appeared in [7, 11, 14, 26, 47, 48]). We will argue that quantum gravity in dS_3 can be described by a two dimensional conformal field theory, in the sense that correlation functions of an operator ϕ inserted at points x_i on $\mathcal{I}^-$ or $\mathcal{I}^+$ are generated by a two dimensional Euclidean CFT:

$$\langle\phi(x_1)\cdots\phi(x_i)\rangle_{dS_3} \leftrightarrow \langle\mathcal{O}_\phi(x_1)\cdots\mathcal{O}_\phi(x_i)\rangle_{S^2}, \tag{4.1}$$

where $\mathcal{O}_\phi$ is an operator in the CFT associated to the field ϕ. Equation (4.1) expresses the dS/CFT correspondence. The tool which will allow us to reach this conclusion is an analysis of the asymptotic symmetry group for gravity in dS_3. Parallel results pertain in arbitrary dimension, but the three dimensional case is the richest because of the infinite dimensional nature of the $\mathcal{I}^\pm$ conformal group. The results of this section are largely contained in [27] except for the derivation of the asymptotic boundary conditions for dS_3, which were assumed/guessed without derivation in [27].

4.1 Asymptotic symmetries

Consider a simple $U(1)$ gauge theory in flat Minkowski space. A gauge transformation which goes to zero at spatial infinity will annihilate physical states (this is just the statement that a physical state is gauge invariant), while a gauge transformation which goes to a constant at spatial infinity will act nontrivially on the states. In fact the generator will be proportional to the charge operator, by Noether's theorem.

It is useful therefore to consider the so-called asymptotic symmetry group (ASG), which is defined as the set of allowed symmetry transformations modulo the set of trivial symmetry transformations.

$$\text{ASG} = \frac{\text{Allowed Symmetry Transformations}}{\text{Trivial Symmetry Transformations}} \cdot \tag{4.2}$$

Here "allowed" means that the transformation is consistent with the boundary conditions that we have specified for the fields in the theory, and "trivial" means that the generator of the transformation vanishes after we have implemented the constraints–for example asymptotically vanishing gauge transformations in the example of the previous paragraph. The states and correlators of the theory clearly must lie in representations of the ASG. Of course one must know the details of the theory to know which representations of the ASG actually appear, but in some cases a knowledge of the ASG already places strong constraints on the theory.

In this section we will see that the ASG of quantum gravity in dS_3 is the Euclidean conformal group in two dimensions. Since this group acts on $\mathcal{I}^{\pm}$, this means that correlators with points on $\mathcal{I}^{\pm}$ are those of a conformal field theory, and the correspondence (4.1) is simply an expression of diffeomorphism invariance of the theory. Although we will not learn anything about the details of this theory, the fact that the conformal group in two dimensions is infinite dimensional already strongly constrains the physics.

In quantum gravity the relevant gauge symmetry is diffeomorphism invariance, and in de Sitter space the only asymptotia are $\mathcal{I}^{\pm}$. Therefore we need to consider diffeomorphisms in dS_3 which preserve the boundary conditions on the metric at $\mathcal{I}^{\pm}$ but do not fall off so fast that they act trivially on physical states. The analogous problem for three dimensional anti-de Sitter space was solved long ago by Brown & Henneaux [61]. The result for de Sitter differs only by a few signs. However the physical interpretation in the dS_3 case is very different from that of AdS_3, and remains to be fully understood.

4.2 De Sitter boundary conditions and the conformal group

Our first task is to specify the boundary conditions appropriate for an asymptotically dS_3 spacetime. In general specification of the boundary conditions is part of the definition of the theory, and in principle there could be more than one choice. However if the boundary conditions are too restrictive, the theory will become trivial. For example in 4d gravity, one might try to demand that the metric fall off spatially as $\frac{1}{r^2}$. This would allow only zero energy configurations and hence the theory would be trivial. On the other hand one might try to demand that it fall off as $\frac{1}{\sqrt{r}}$. Then the energy and other symmetry generators are in general divergent, and it is unlikely any sense can be made of the theory. So the idea is to make the falloff as weak as possible while still maintaining finiteness of the generators.

Hence we need to understand the surface integrals which generate the diffeomorphisms of dS_3. A convenient and elegant formalism for this purpose was developed by Brown & York [62, 63] (and applied to AdS_3 in [65, 66]). They showed that bulk diffeomorphisms are generated by appropriate moments of a certain stress tensor which lives on the boundary of the spacetime[3]. We will define an asymptotically dS_3 spacetime to be one for which the associated stress tensor, and hence all the symmetry generators, are finite.

The Brown-York stress tensor for dS_3 with $\ell = 1$ is given by[4]

$$T_{\mu\nu} = \frac{1}{4G}\left[K_{\mu\nu} - (K+1)\gamma_{\mu\nu}\right]. \tag{4.3}$$

Here γ is the induced metric on the boundary $\mathcal{I}^-$ and K is the trace of the extrinsic curvature $K_{\mu\nu} = -\nabla_{(\mu}n_{\nu)} = -\frac{1}{2}\mathcal{L}_n\gamma_{\mu\nu}$ with n^μ the outward-pointing unit normal. (4.3) vanishes identically for vacuum dS_3 in planar coordinates

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + \mathrm{e}^{-2t}\mathrm{d}z\mathrm{d}\bar{z}. \tag{4.4}$$

For a perturbed metric $g_{\mu\nu} + h_{\mu\nu}$ we obtain the Brown-York stress tensor

$$\begin{aligned}
T_{zz} &= \frac{1}{4G}\left[h_{zz} - \partial_z h_{tz} + \frac{1}{2}\partial_t h_{zz}\right] + \mathcal{O}(h^2),\\
T_{z\bar{z}} &= \frac{1}{4G}\left[\frac{1}{4}\mathrm{e}^{-2t}h_{tt} - h_{z\bar{z}} + \frac{1}{2}(\partial_{\bar{z}}h_{tz} + \partial_z h_{t\bar{z}} - \partial_t h_{z\bar{z}})\right]\\
&\quad + \mathcal{O}(h^2).
\end{aligned} \tag{4.5}$$

Details of this calculation are given in Appendix A. Requiring the stress tensor to be finite evidently leads to the boundary conditions

$$\begin{aligned}
g_{z\bar{z}} &= \frac{\mathrm{e}^{-2t}}{2} + \mathcal{O}(1),\\
g_{tt} &= -1 + \mathcal{O}(\mathrm{e}^{2t}),\\
g_{zz} &= \mathcal{O}(1),
\end{aligned}$$

[3] Brown and York mainly consider a timelike boundary, but their results can be extended to the spacelike case.

[4] We caution the reader that the generalization of (4.3) to $d > 3$ or to theories with matter is not entirely straightforward [66].

$$g_{tz} = \mathcal{O}(1). \tag{4.6}$$

It is not hard to see that the most general diffeomorphism ζ which preserves the boundary conditions (4.6) may be written as

$$\zeta = U\partial_z + \frac{1}{2}U'\partial_t + \mathcal{O}(\mathrm{e}^{2t}) + \text{complex conjugate}, \tag{4.7}$$

where $U = U(z)$ is holomorphic in z[5]. A diffeomorphism of the form (4.7) acts on the Brown-York stress tensor as

$$\delta_\zeta T_{zz} = -U\partial T_{zz} - 2U'T_{zz} - \frac{1}{8G}U'''. \tag{4.8}$$

The first two terms are those appropriate for an operator of scaling dimension two. The third term is the familiar linearization of the anomalous Schwarzian derivative term corresponding to a central charge

$$c = \frac{3l}{2G}, \tag{4.9}$$

where we have restored the power of ℓ[6]. Note that the $\mathcal{O}(\mathrm{e}^{2t})$ terms in (4.7) do not contribute in (4.8). Therefore they are trivial diffeomorphisms, in the sense described above. We conclude that the asymptotic symmetry group of dS_3 as generated by (4.7) is the conformal group of the Euclidean plane.

The last boundary condition (4.6) differs from the condition $g_{tz} = \mathcal{O}(\mathrm{e}^{2t})$ assumed in [27] and obtained by analytically continuing the AdS_3 boundary conditions of Brown & Henneaux [61] from anti-de Sitter to de Sitter space. The resolution of this apparent discrepancy comes from noting that if $g_{tz} \to f$ on the boundary where $f = f(z, \bar{z})$ is an arbitrary function, then applying the diffeomorphism $\zeta = \mathrm{e}^{2t} f \partial_{\bar{z}}$ gives $\delta_\zeta g_{tz} = \mathcal{O}(\mathrm{e}^{2t})$. Therefore one can always set the component g_{tz} of the metric to be $\mathcal{O}(\mathrm{e}^{2t})$ with a trivial diffeomorphism. In other words, if $g_{tz} = \mathcal{O}(1)$, then in fact one can always choose a gauge in which $g_{tz} = \mathcal{O}(\mathrm{e}^{2t})$. Exploiting this freedom one can impose the asymptotic boundary conditions

$$\begin{aligned} g_{z\bar{z}} &= \frac{\mathrm{e}^{-2t}}{2} + \mathcal{O}(1), \\ g_{tt} &= -1 + \mathcal{O}(\mathrm{e}^{2t}), \\ g_{zz} &= \mathcal{O}(1), \\ g_{tz} &= \mathcal{O}(\mathrm{e}^{2t}), \end{aligned} \tag{4.10}$$

[5]We allow isolated poles in z. In principle this should be carefully justified (as (4.6) is violated very near the singularity), and we have not done so here. A parallel issue arises in AdS_3/CFT$_2$.

[6]Parallel derivations of the central charge for AdS were given in [64, 65].

as given in [27].

A special case of (4.7) is the choice

$$U = \alpha + \beta z + \gamma z^2, \tag{4.11}$$

where $\alpha,\ \beta,\ \gamma$ are complex constants. In this case U''' vanishes, and the dS_3 metric is therefore invariant. These transformations generate the $SL(2,\mathbf{C})$ global isometries of dS_3.

Where do conformal transformations come from? Recall that a conformal transformation in two dimensions is a combination of an ordinary diffeomorphism and a Weyl transformation. In two dimensions a diffeomorphism acts as

$$g_{z\bar{z}} \to \frac{\mathrm{d}z'}{\mathrm{d}z}\frac{\mathrm{d}\bar{z}'}{\mathrm{d}\bar{z}} g_{z'\bar{z}'}, \tag{4.12}$$

and a Weyl transformation acts as

$$g_{z\bar{z}} \to \mathrm{e}^{2\phi} g_{z\bar{z}}. \tag{4.13}$$

A conformal transformation is just an ordinary diffeomorphism (4.12) followed by a Weyl transformation (4.13) with ϕ chosen so that $g_{z\bar{z}} \to g_{z'\bar{z}'}$ under the combined transformation.

Now if we look at what the diffeomorphism ζ defined in (4.7) does, we see that the first term $U\partial_z$ generates a holomorphic diffeomorphism of the plane. Now the form of the metric (4.4) makes it clear that this can be compensated by a shift in t, which from the point of view of the z-plane is a Weyl transformation. This accounts for the second term $\frac{1}{2}U'\partial_t$ in ζ. So a diffeomorphism in dS_3 splits into a tangential piece, which acts like an ordinary diffeomorphism of the complex plane, and a normal piece, which acts like a Weyl transformation. A three dimensional diffeomorphism is thereby equivalent to a two dimensional conformal transformation.

Since $U(z)$ was arbitrary, we conclude that the asymptotic symmetry group of gravity in dS_3 is the conformal group of the complex plane. The isometry group is the $SL(2,\mathbf{C})$ subgroup of the asymptotic symmetry group. In particular, the ASG is infinite dimensional, a fact which highly constrains quantum gravity on dS_3. This is particular to the three dimensional case, since in higher dimensional de Sitter space the ASG is the same as the isometry group $SO(d,1)$.

We conclude these lectures with a last

Exercise 4.

(a) Find an example of string theory on de Sitter space.

(b) Find the dual conformal field theory.

A Calculation of the Brown-York stress tensor

We wish to calculate the Brown-York stress tensor (4.3) for a metric which is a small perturbation of dS_3. We write the metric in planar coordinates $(t, z, \bar{z})$ as

$$\mathrm{d}s^2 = g_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu = -\mathrm{d}t^2 + \mathrm{e}^{-2t}\mathrm{d}z\mathrm{d}\bar{z} + h_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu, \tag{A.1}$$

and we will always drop terms of order $\mathcal{O}(h^2)$. We can put (A.1) into the form

$$\mathrm{d}s^2 = -N^2\mathrm{d}t^2 + \gamma_{ij}(\mathrm{d}x^i + N^i\mathrm{d}t)(\mathrm{d}x^j + N^j\mathrm{d}t), \tag{A.2}$$

where the lapse and shift functions are given by

$$N = 1 - \frac{1}{2}h_{tt}, \qquad N^z = 2\mathrm{e}^{2t}h_{t\bar{z}}, \qquad N^{\bar{z}} = 2\mathrm{e}^{2t}h_{tz}, \tag{A.3}$$

and the induced metric on the boundary $\mathcal{I}^-$ is

$$\gamma_{zz} = h_{zz}, \qquad \gamma_{z\bar{z}} = \frac{1}{2}\mathrm{e}^{-2t} + h_{z\bar{z}}, \qquad \gamma_{\bar{z}\bar{z}} = h_{\bar{z}\bar{z}}. \tag{A.4}$$

The outward pointing unit normal vector to the boundary is

$$n^\mu = \frac{1}{N}\left(-1, N^z, N^{\bar{z}}\right) = \left(-1 - \frac{1}{2}h_{tt}, 2\mathrm{e}^{2t}h_{t\bar{z}}, 2\mathrm{e}^{2t}h_{tz}\right)\cdot \tag{A.5}$$

Upon lowering the indices, we have

$$n_\mu = \left(1 - \frac{1}{2}h_{tt}, 0, 0\right) \tag{A.6}$$

and we use the formula $K_{\mu\nu} = -\frac{1}{2}(\nabla_\mu n_\mu + \nabla_\nu n_\mu)$ to obtain

$$\begin{aligned} K_{zz} &= -\partial_z h_{tz} + \frac{1}{2}\partial_t h_{zz}, \\ K_{z\bar{z}} &= -\frac{1}{2}\mathrm{e}^{-2t}(1 + \frac{1}{2}h_{tt}) - \frac{1}{2}\left(\partial_{\bar{z}}h_{tz} + \partial_z h_{t\bar{z}} - \partial_t h_{z\bar{z}}\right). \end{aligned} \tag{A.7}$$

The trace is

$$K = g^{\mu\nu}K_{\mu\nu} = \gamma^{ij}K_{ij} = -2 - h_{tt} + 4\mathrm{e}^{2t}h_{z\bar{z}} - 2\mathrm{e}^{2t}\left(\partial_{\bar{z}}h_{tz} + \partial_z h_{t\bar{z}} - \partial_t h_{z\bar{z}}\right). \tag{A.8}$$

Plugging (A.7) and (A.8) into (4.3) gives the desired result (4.5).

References

[1] J.D. Bekenstein, *Phys. Rev. D* **7** (1973) 2333.

[2] S.W. Hawking, *Commun. Math. Phys.* **43** (1975) 199.

[3] G.W. Gibbons and S.W. Hawking, *Phys. Rev. D* **15** (1977) 2738.

[4] A. Strominger and C. Vafa, *Phys. Lett. B* **379** (1996) 99 [hep-th/9601029].

[5] J. Maldacena, *Adv. Theor. Math. Phys.* **2** (1998) 231; *Int. J. Theor. Phys.* **38** (1998) 1113 [hep-th/9711200].

[6] K. Pilch, P. van Nieuwenhuizen and M.F. Sohnius, *Commun. Math. Phys.* **98** (1985) 105.

[7] C.M. Hull, *JHEP* **9807** (1998) 021 [hep-th/9806146].

[8] N.C. Tsamis and R.P. Woodard, *Annals Phys.* **253** (1997) 1 [hep-ph/9602316].

[9] J. Maldacena and A. Strominger, *JHEP* **9802** (1998) 014 [gr-qc/9801096].

[10] M.I. Park, *Phys. Lett. B* **440** (1998) 275 [hep-th/9806119].

[11] M. Banados, T. Brotz and M.E. Ortiz, *Phys. Rev. D* **59** (1999) 046002 [hep-th/9807216].

[12] W.T. Kim, *Phys. Rev. D* **59** (1999) 047503 [hep-th/9810169].

[13] F. Lin and Y. Wu, *Phys. Lett. B* **453** (1999) 222 [hep-th/9901147].

[14] R. Bousso, *JHEP* **9906** (1999) 028 [hep-th/9906022].

[15] C.M. Hull and R. R. Khuri, *Nucl. Phys. B* **575** (2000) 231 [hep-th/9911082].

[16] S. Hawking, J. Maldacena and A. Strominger, *JHEP* **0105** (2001) 001 [hep-th/0002145].

[17] T. Banks, *Cosmological breaking of supersymmetry or little Lambda goes back to the future. II* [hep-th/0007146].

[18] R. Bousso, *JHEP* **0011** (2000) 038 [hep-th/0010252].

[19] R. Bousso, *JHEP* **0104** (2001) 035 [hep-th/0012052].

[20] A. Volovich, *Discreteness in deSitter space and quantization of Kaehler manifolds* [hep-th/0101176].

[21] T. Banks and W. Fischler, *M-theory observables for cosmological space-times* [hep-th/0102077].

[22] V. Balasubramanian, P. Horava and D. Minic, *JHEP* **0105** (2001) 043 [hep-th/0103171].

[23] S. Deser and A. Waldron, *Nucl. Phys. B* **607** (2001) 577 [hep-th/0103198].

[24] S. Deser and A. Waldron, *Phys. Lett. B* **513** (2001) 137 [hep-th/0105181].

[25] E. Witten, *Quantum gravity in de Sitter space* [hep-th/0106109].

[26] E. Witten, *Quantum gravity in de Sitter space*, Strings 2001 online proceedings http://theory.tifr.res.in/strings/Proceedings.

[27] A. Strominger, *The dS/CFT correspondence* [hep-th/0106113].

[28] M. Li, *Matrix model for de Sitter* [hep-th/0106184].

[29] S. Nojiri and S. D. Odintsov, *Conformal anomaly from dS/CFT correspondence* [hep-th/0106191].

[30] E. Silverstein, *(A)dS backgrounds from asymmetric orientifolds* [hep-th/0106209].

[31] D. Klemm, *Some aspects of the de Sitter/CFT correspondence* [hep-th/0106247].

[32] A. Chamblin and N. D. Lambert, *Zero-branes, quantum mechanics and the cosmological constant* [hep-th/0107031].

[33] Y. Gao, *Symmetries, matrices, and de Sitter gravity* [hep-th/0107067].

[34] J. Bros, H. Epstein and U. Moschella, *The asymptotic symmetry of de Sitter spacetime* [hep-th/0107091].

[35] S. Nojiri and S. D. Odintsov, *Quantum cosmology, inflationary brane-world creation and dS/CFT correspondence* [hep-th/0107134].

[36] E. Halyo, *De Sitter entropy and strings* [hep-th/0107169].

[37] I. Sachs and S. N. Solodukhin, *Horizon holography* [hep-th/0107173].

[38] A.J. Tolley and N. Turok, *Quantization of the massless minimally coupled scalar field and the dS/CFT correspondence* [hep-th/0108119].

[39] T. Shiromizu, D. Ida and T. Torii, *Gravitational energy, dS/CFT correspondence and cosmic no-hair* [hep-th/0109057].

[40] L. Dolan, C.R. Nappi and E. Witten, *Conformal operators for partially massless states* [hep-th/0109096].

[41] R. Kallosh, *$N = 2$ supersymmetry and de Sitter space* [hep-th/0109168].

[42] C.M. Hull, *De Sitter Space in Supergravity and M Theory* [hep-th/0109213].

[43] B.P. Schmidt *et al.*, *ApJ* **507** (1998) 46 [astro-ph/9805200].

[44] A.G. Riess *et al.*, *ApJ* **116** (1998) 1009 [astro-ph/9805201].

[45] S. Perlmutter *et al.*, *ApJ* **517** (1999) 565 [astro-ph/9812133].

[46] S. Perlmutter, *Int. J. Mod. Phys. A* **15S1** (2000) 715 [eConfC **990809** (2000) 715].

[47] I. Antoniadis, P.O. Mazur and E. Mottola, *Comment on Nongaussian isocurvature perturbations from inflation* [astro-ph/9705200].

[48] P.O. Mazur and E. Mottola, *Phys. Rev. D* **64** (2001) 104022 [hep-th/0106151].

[49] S. Deser and R. Jackiw, *Annals Phys.* **153** (1984) 405.

[50] N.D. Birrell and P.C.W. Davies, *Quantum Fields in Curved Space* (Cambridge University Press, Cambridge, 1982).

[51] N.A. Chernikov and E.A. Tagirov, *Ann. Poincare Phys. Theor. A* **9** (1968) 109.

[52] E.A. Tagirov, *Ann. Phys.* **76** (1973) 561.

[53] R. Figari, R. Hoegh-Krohn and C.R. Nappi, *Commun. Math. Phys.* **44** (1975) 265.

[54] H. Rumpf and H.K. Urbantke, *Ann. Phys.* **114** (1978) 332.

[55] L.F. Abbott and S. Deser, *Nucl. Phys. B* **195** (1982) 76.

[56] A.H. Najmi and A.C. Ottewill, *Phys. Rev. D* **30** (1984) 1733.

[57] L.H. Ford, *Phys. Rev. D* **31** (1985) 710.

[58] E. Mottola, *Phys. Rev. D* **31** (1985) 754.

[59] B. Allen, *Phys. Rev. D* **32** (1985) 3136.

[60] B. Allen and A. Folacci, *Phys. Rev. D* **35** (1987) 3771.

[61] J.D. Brown and M. Henneaux, *Commun. Math. Phys.* **104** (1986) 207.

[62] J.D. Brown and J.W. York, *Phys. Rev. D* **47** (1993) 1407.

[63] J.D. Brown, S.R. Lau and J.W. York, *Action and Energy of the Gravitational Field* [gr-qc/0010024].

[64] M. Henningson and K. Skenderis, *JHEP* **9807** (1998) 023 [hep-th/9806087].

[65] V. Balasubramanian and P. Kraus, *Commun. Math. Phys.* **208** (1999) 413 [hep-th/9902121].

[66] S. de Haro, S.N. Solodukhin and K. Skenderis, *Commun. Math. Phys.* **217** (2001) 595 [hep-th/0002230].

LECTURE 7

STRING COMPACTIFICATION WITH $\mathcal{N} = 1$ SUPERSYMMETRY

M.R. DOUGLAS

Rutgers and IHES,
35 route de Chartres,
91440 Bures-sur-Yvette, France

Contents

STRING COMPACTIFICATION WITH $\mathcal{N} = 1$ SUPERSYMMETRY

M.R. Douglas

Abstract

We give a very broad overview of problems and recent work in the study of string compactification with $\mathcal{N} = 1$ supersymmetry, following the paradigms of heterotic string compactification on a Calabi-Yau manifold, and the more recent Dirichlet brane paradigm.

1 Introduction and the question of vacuum selection

The last Les Houches summer school which treated the fundamental theories of physics: gauge theory, gravity and the string and M theories which contain both of these, was held in 1995 [29].

In the six years since, the subject has changed dramatically, in what is often called the "second superstring revolution". Many of the central topics of our 2001 school, were in 1995 either curiosities known to few (such as Dirichlet branes) or had not been conceived of at all. Although string theory was by then widely accepted as the leading candidate theory of quantum gravity, it had said little about the main physical questions identified in previous work. This changed at the end of 1995 with the Strominger-Vafa derivation of the Bekenstein-Hawking entropy formula from string theory. Even more surprising discoveries were to follow, such as M(atrix) theory, F theory, solutions of gauge theory using brane techniques, the AdS/CFT correspondence, and the relevance of noncommutative field theory. We now have some non-perturbative understanding of the theory, and some new and radically different pictures of space-time.

By 1997 the main ideas of this second revolution had emerged. Although compared to 1994–97, string theory today appears relatively stable, there is still much to do. The new ideas have still not been completely understood and absorbed, and the field remains very active and lively; it is not in obvious crisis. Much of the attention has shifted towards possible application to the real world. This situation is very good from the point of view of the organizers of a school, as it allowed us to make a well motivated choice of topics. Just as importantly, the speakers could feel that their time was well spent, in introducing material of lasting importance (we think).

On the other hand, students (and everyone else) also come looking for vision, and from this point of view one might find the school a bit too cut and dried. So let us try to look ahead a bit to the "third superstring revolution".

Perhaps the most basic question from the first revolution left untouched by the developments of the second, is the problem of vacuum selection. The original scenario of $E_8 \times E_8$ heterotic string theory compactified on a six dimensional Calabi-Yau manifold led quite directly to theories resembling the standard model, but attempts to make more specific statements foundered on the problem that these depend on the choice of manifold, and there are thousands if not millions (if indeed the number is finite!) of such manifolds.

Far from helping, the second superstring revolution only seems to make this problem worse. Many new ways were found to get gauge symmetry, which clearly enlarge the number of models. For example, the heterotic bound on the rank of the gauge group is easy to exceed. Early hopes that many compactifications would be inconsistent at the nonperturbative level now seem ill-founded; we know weak coupling nonperturbative consistency conditions for backgrounds, and we have strong coupling duals for enough examples which satisfy these conditions to believe they are sufficient. At present there are not even good speculations as to how many four-dimensional compactifications of string/M theory there might be.

It seems to me that the third superstring revolution will have to address the problem of vacuum selection, as the outstanding barrier to contact with observation. I think many would agree with this; where opinions differ is on how.

At one extreme is the idea is that there is some *a priori* "Vacuum Selection Principle" waiting to be discovered, which will uniquely fix four macroscopic dimensions, a particular theory and internal space, and all of the couplings.

At the other extreme is the null hypothesis, which states that if choices enter in deriving predictions from string/M theory, one should not rule out possibilities *a priori*, but instead check whether observations can be fit by any of the possible choices. From this point of view, after the primary test of consistency, each check for a possible conflict with observation could be regarded as a vacuum selection principle. But there need be no single principle which in itself governs the problem; one simply needs to sort through all possibilities and apply a series of tests to each.

Of course these are two extreme points of view and the truth is likely to lie in between. It does seem to me however that too much importance is given to the first point of view, which is not at all necessary for string theory to make contact with reality.

There are various ways that the hypothesis of a selection principle might be realized in string theory. The only generally accepted selection principle is the following: candidate vacua must be local minima of the vacuum energy, or almost minima. In any case, any instabilities or time dependence must operate on cosmological time scales. This is an extremely powerful principle which one expects will stabilize all of the continuous parameters in the standard model Lagrangian, since the vacuum energy depends on all of these. Of course one does not expect a single local minimum. Other problems with this principle are that testing it requires both fairly detailed understanding of the set of candidate vacua, and extremely precise computational ability to get the vacuum energy.

The resulting vacuum energy is the cosmological constant, and it seems *a priori* implausible that computing it in this way would lead to an acceptably small result. Clearly this is an outstanding problem, and it might be that a mechanism fixing this will only work in one or a few backgrounds. See [7, 41] for a recent review of some proposals. Many of these do not fit well with our present understanding of string/M theory, while many more seem too far beyond our understanding to evaluate at present, but this would certainly be an attractive outcome.

The closest thing one has to a "null hypothesis" in this area is the idea that the four dimensional cosmological constant, while appearing as a parameter in the four dimensional Lagrangian, can appear as a boundary condition in some larger framework. One way this might be realized is by having many quantized and independent contributions to the vacuum energy; the cosmological constant is then not continuously adjustible but rather lies in a "discretuum" of closely spaced possible values [10, 11, 23]. A different realization of this idea emerges in the context of higher dimensional theories with gravity confined to a four-dimensional brane [37].

Having the cosmological constant as a boundary condition facilitates fixing it by quantum or even anthropic considerations. Even without these further considerations, as soon as one has found consistent models which can match observation, including the known bounds on the cosmological constant, one might regard the problem as solved. Such mechanisms seem likely to work in a broad class of compactifications but not all, so this would fall closer to the weaker, uncapitalized, type of principle.

Another likely arena for a selection principle to operate is cosmology. It is certainly clear that one has the "antiselection" principle that putative vacua can be unstable to tunnelling to other nearby vacua, and that this generally becomes more important at high temperature. It also seems reasonable to assume (meaning that there is no good evidence against the idea) that the configuration space of compactifications is connected, especially if we allow Planck scale barrier energies and temperatures, so that early dynamics would select the relatively stable vacua out of the entire set.

It would be interesting to have some rough, qualitative model of the configuration space and effective potential, to judge how useful any of these considerations are. In the absence of other evidence, the most plausible scenario is that the effective potential takes a "spin glass" form, with a very large number of minima. Generally, cooling such a system does not lead to a global minimum, but rather local minima with some distribution depending on the annealing process. So cosmological evolution might well lead to a selection principle, but again of the weaker type.

All this leads us to the conclusion that there may be no way around searching through a list of candidate vacua to find the right one. Should this prospect depress us? It depends very much on whether we can construct such a list and how long it is. Even the question of what form such a "list" would take requires discussion. It is not at all clear how to work on such a hard problem.

In the face of such a hard problem, one needs to propose simpler analog problems for which we at least know how to start. As the most basic problem which has any resemblance to this one, let me suggest the following: to determine whether there are finitely or infinitely many vacua of string/M theory with four macroscopic dimensions and $\mathcal{N} = 1$ supersymmetry. Of course, if there are finitely many, one could go on to ask how many.

The main simplification one has imposed is $\mathcal{N} = 1$ supersymmetry. Of course the primary motivation for this is the many reasons to believe that supersymmetry will be discovered in the coming years. But there is also a good theoretical motivation: namely, in recent years there has been quite a lot of success in finding $\mathcal{N} = 1$ vacua of field theory and string theory, and relatively little in understanding supersymmetry breaking or vacua of non-supersymmetric theories. There are many reasons for this, among which the two most important are the prevalence of phase transitions in more general field theory, and the role of holomorphy in finding the $\mathcal{N} = 1$ solutions.

As stated, the problem is ill-posed, because generic supersymmetric theories have flat directions and moduli spaces of vacua. A rough definition which should suffice for the finiteness problem is to count each irreducible subvariety of the moduli space (*i.e.* a component defined by a set of finitely many non-redundant F-flatness conditions) as a vacuum. Better definitions can be made, as will be discussed elsewhere.

2 From ten dimensional geometry to four dimensional effective field theory

Let us give a (very incomplete) overview of $\mathcal{N} = 1$ compactification, starting with the traditional heterotic string on a large (compared to α') Calabi-Yau at weak coupling, and then discussing stringy and quantum corrections, mostly in the context of Dirichlet branes in type II on CY_3[1]. The main point we will emphasize is the close link between geometry in ten dimensions, and the structure of $\mathcal{N} = 1$ effective field theory in four dimensions.

One should start the discussion in the large compactification limit, which in this case is to obtain $\mathcal{N} = 1$ in four dimensions by compactifying $\mathcal{N} = 1$, $d = 10$ supergravity coupled to super Yang-Mills (with the stringy anomaly cancelling term) on a Calabi-Yau threefold, as discussed at length in Green, Schwarz and Witten's well known textbook, which we will assume is familiar. To summarize, one obtains an $\mathcal{N} = 1$ model by postulating a Ricci-flat six manifold M with a covariantly constant spinor, and a gauge bundle V on M with structure group $G \in E_8 \times E_8$ (or $Spin(32)/Z_2$) satisfying a condition on the second Chern class, $c_2(V) = c_2(TM)$, and with connection satisfying the hermitian Yang-Mills equations. The resulting low energy theory has as gauge group $\hat{G}$ the commutant of G in $E_8 \times E_8$; and massless matter spectrum corresponding to the various infinitesimal deformations one can make of this structure. One would choose $G = SU(3), SU(4)$ or $SU(5)$ to obtain the standard grand unified groups $\hat{G} = E_6, SO(10)$ and $SU(5)$ respectively. The case of $G = SU(3)$ and $V = TM$ is simpler than the others and it played an important role in the early discussions, but there seems to be no other reason to prefer it.

This story rests on some rather sophisticated mathematics, developed in the late 70's and early 80's. The most famous part of this is Yau's theorem and the Donaldson-Uhlenbeck Yau theorems, which reduce the problem of solving the physical Einstein-Yang-Mills equations with a covariantly constant spinor, to problems of pure algebraic geometry, namely finding a three complex dimensional manifold $\mathcal{M}$ with $c_1(\mathcal{M}) = 0$ and a holomorphic vector bundle $\mathcal{V}$ satisfying the constraints above, and a condition called μ-stability which is generally (not always) satisfied.

In a deep sense [14], this reduction of the problem has a direct counterpart in the language of $d = 4$, $\mathcal{N} = 1$ effective field theory, in the familiar statement [6] that given a solution of the F-flatness conditions of supersymmetric gauge theory, one can always solve the D-flatness equations. This statement is only strictly true with zero Fayet-Iliopoulos terms;

[1]The lectures at the school went into much more depth, and most of this material will appear in the proceedings of the 2001 Trieste spring school and the 2002 CMI spring school.

with non-zero FI terms there is a necessary and sufficient condition called θ-stability [19, 32], very analogous to μ-stability.

Similarly, the problem of finding holomorphic $\mathcal{M}$ and $\mathcal{V}$ corresponds to solving the F-flatness conditions. These are not easy to solve for a model with a general superpotential and correspondingly one should expect that this type of algebraic geometry is somewhat intricate. One can show that the superpotential vanishes if $\mathcal{V} = T\mathcal{M}$, but this does not get one too far.

One of the main observables of interest in the early discussions was the "Yukawa couplings", the cubic term in the superpotential between charged matter multiplets, which must reproduce the matter–Higgs couplings in the supersymmetric standard model. This very generally arises from the holomorphic Chern-Simons action,

$$W = \int_{\mathcal{M}} \bar{\Omega} \wedge \mathrm{Tr} \left(A\mathrm{d}A + \frac{2}{3} A^3 \right) \tag{2.1}$$

by perturbing around the background connection $A^{(V)}$; writing

$$A = A^{(V)} + \sum_i \phi_i \delta A^{(i)}$$

with $\delta A^{(i)}$ a basis of solutions of the linearized equation of motion $D^{(V)} \delta A = 0$, one gets a cubic term

$$W = \frac{2}{3} \phi_i \phi_j \phi_k \int \bar{\Omega} \wedge \delta A^{(i)} \wedge \delta A^{(j)} \wedge \delta A^{(k)}.$$

Although this is the piece which is most relevant for phenomenology, this coupling is only the tip of the iceberg. In fact the entire holomorphic Chern-Simons action 2.1 can be identified with the classical superpotential at arbitrary points in configuration space. This follows from the equation

$$0 = W' = \frac{\delta W}{\delta A_i} = \Omega^{ijk} F_{jk} = F^{2,0} \tag{2.2}$$

which is precisely the condition for A to be the connection of a holomorphic bundle. Thus the correspondence we mentioned between this condition and F-flatness is quite precise.

This global picture shows us that many different vacua are in fact connected by moving through configuration space. A simple example of this is that given $G \subset E_8$, the number of generations minus antigenerations of $\hat{G}$ is determined by the index theorem to be half the third Chern class $c_3(V)$. One also has physical grounds to expect this number to be invariant in four dimensions, at least at weak coupling.

On the other hand, within the larger class of E_8 bundles, c_3 is not conserved (one cannot even define c_3, which requires having a third rank invariant symmetric tensor). This means that one can find transitions (by varying from one G-bundle $\mathcal{V}$ to another $\mathcal{V}'$) in which the number of generations changes in four dimensions. The resolution of the apparent contradiction is that, although one finds the gauge group $\hat{G}$ on both sides, the gauge symmetry is broken to a smaller group along the way, and the two gauge groups are not the same but just isomorphic[2].

Getting concrete models out of this construction now requires building concrete CY_3 manifolds and bundles. A lot of work was devoted to this, and a lot of CY_3 manifolds are known. The most general systematic construction known is to realize them as hypersurfaces in toric varieties. The simplest case of this is the familiar quintic hypersurface in $\mathbb{CP}^4$. The generalization is best described using a linear sigma model, which embeds the CY in a higher dimensional linear space, as the string world-sheet theory. [27] This theory must have $(2,2)$ world-sheet supersymmetry and the possibilities are (essentially) what one can get by dimensionally reducing from $d = 4$, $\mathcal{N} = 1$ to $d = 2$. Thus the vacua, which now are interpreted as the target space of the string, are again determined by gauge symmetry, D-flatness and F-flatness. A toric variety is now a space of vacua of $U(1)^r$ gauge theory with no superpotential, and the hypersurface is realized as $W' = 0$.

Mathematicians do not believe all CY_3's can be constructed this way, although apparently they do not have a specific counterexample. One reason behind this belief is that, except in simple cases, not all of the geometric moduli of the CY are manifest in this description. In world-sheet terms, the $U(1)$ quotients lead to orbifold singularities and associated twisted sector moduli. This does not directly contradict the weaker hypothesis that one gets a representative of each birational equivalence class of CY, but this hypothesis is less compelling.

The idea of "birational transformations" will come back so let us describe it physically. Mathematically, two complex manifolds are birationally equivalent if one can find a complex map between them which is one-to-one except for some lower dimensional subvarieties. This allows topology change of a mild type (for example the flop). In the toric construction, varying FI terms at fixed complex structure very generally leads to birationally equivalent spaces.

This provides a very large set of examples which has in a sense been classified [34], showing that their number is finite and less than 4.8×10^8 (one expects many duplications appear on their list). It includes natural

[2]There is another, apparently different mechanism for changing the number of generations, which passes through a strong coupling transition [31].

CY's with maximal and minimal Euler number, and all of the examples which have been suggested by duality arguments seem to be present as well, so there is as yet no physical evidence that there are non-toric CY's. It is also known that this set is connected in the sense of [36].

The linear sigma model construction also enables us to go on and study stringy corrections. This is a long story, but the well-understood part of it is pretty well summarized in the following result of mirror symmetry [27]: while the complex structure moduli space of $\mathcal{M}$ is classically exact, the Kähler moduli space is drastically modified by world-sheet instanton corrections, to become isomorphic to the complex structure moduli space of a mirror manifold $\mathcal{W}$. This is a very attractive description as it means we only need to know about complex structure moduli spaces, which have been much studied. In particular, the resulting moduli space can be described without even mentioning α'. It also leads directly to the idea that string theory provides a smooth interpolation between birationally equivalent CY's.

The story with the bundles is somewhat less satisfactory. The linear sigma model construction can be directly generalized to the heterotic string, by realizing the gauge symmetry with $(32+k,k)$ left and right-moving fermions, and pairing these in a position-dependent way. The result reproduces the "monad" construction of mathematics, which constructs a bundle V in terms of two maps f and g as

$$V = \frac{\text{kernel } g}{\text{image } f}\,. \tag{2.3}$$

In world-sheet terms, $g = W'$ and f comes from the $U(1)^r$ action. This construction produces lots of examples. Still, mathematicians know more about bundles than about CY's, and in this case they can prove that it does not produce all of them[3].

There is another feature lacking which a really good construction should have, namely it should provide a one-to-one correspondence between the objects being constructed, and the data of the construction, up to some simple gauge equivalences. This allows one to directly interpret the data of the construction as fields in the effective theory, pulling back the various ten-dimensional quantities (such as 2.1) to obtain the corresponding four-dimensional quantities. This makes the relation between ten and four dimensions completely transparent.

This might seem like too much to ask for, but indeed it was realized in some of the early mathematical applications of 2.3. In particular, the

[3]The main point is that on a d-dimensional complex manifold, one needs d maps in this type of construction to get everything. This generalization can be made in open string linear sigma models, and attempts have been made for the heterotic string [28].

ADHM construction of instantons on $\mathbb{R}^4$ works precisely this way. Although one can verify this explicitly, the mathematical intuition for this point was for a long time unclear to physicists, who regarded the whole thing as a sort of black art.

This changed with a paper of Witten [39], which appeared almost immediately after Polchinski's famous paper on Dirichlet branes [35], showing that the ADHM construction for $Sp(n)$ instantons followed almost trivially from the physics of D-branes in the type I string. This was quickly generalized to other groups in [15] and to orbifolds of $\mathbb{R}^4$ in [21].

To put the argument in a nutshell, consider dimensional reduction of the type I string on a very large torus T^4[4]. In a background with k instantons this leads to a low energy theory with $\mathcal{N}s = 2 \times 4$ supercharges, *i.e.* the equivalent of $\mathcal{N} = 2$ supersymmetry in $d = 4$. The instanton moduli become massless fields and their number is determined at large volume by an index theorem. To discuss the stringy regime, one relies on supersymmetry arguments. $\mathcal{N} = 2$ supersymmetry implies that the number of moduli cannot change, except at points where the unbroken gauge group changes. There is no evidence for this in the problem at hand, rather the only simple picture is that the exact moduli space has a limit describing zero size instantons.

The limiting zero size instantons are in fact D5-branes in type I. In the more general context of instantons on Dp-branes in type II, they are D$p-4$-branes. This is motivated by the following world-sheet coupling on the Dp-branes [15, 30]

$$\int C \wedge \mathrm{Tr} e^{F-B} \sqrt{\hat{A}}$$

which one can derive by explicit world-sheet computation or by anomaly inflow arguments. In particular, the instanton number or second Chern class $\mathrm{Tr} F^2$ becomes the D$p - 4$ charge.

The test of this hypothesis is that, expanding around this zero size limit, BPS (supersymmetric) brane configurations correspond to supersymmetric gauge field configurations. These are supersymmetric vacua of the coupled Dp-D$p-4$ system, whose world-volume Lagrangian is determined by the massless spectrum and $d = 4$, $\mathcal{N} = 2$ supersymmetry. Indeed, the resulting equations are precisely the ADHM equations. One can even go on and derive the explicit gauge field configuration by introducing a D$p-8$ brane, which serves as a probe.

To connect this with our previous discussion, one can use S-duality a small instanton in the $SO(32)$ heterotic string is the same object. In fact, the heterotic five-brane solution has the property that the dilaton runs off

[4]We take a compact background to provide a clear separation between massless and massive KK modes.

to strong coupling in the core, so the type I description is more appropriate. This justifies the use of weakly coupled world-volume theory in the problem.

On the surface, what stands out in this argument is the role of supersymmetry in deriving the world-volume theory. But although this greatly simplifies the discussion, it is not really the most important point. Rather, it is the one-to-one equivalence between configurations of the brane world-volume theory and configurations of the original higher dimensional theory. Having checked that dimensions of moduli spaces agree, there are two further claims which go into this.

The first and easier point to check is that, for a given topological class of configuration, there is a specific combination of branes whose world-volume theory must represent it. This is obvious here but only because we know that self-dual instantons have $c_2(V) > 0$. If there had been configurations preserving the same supersymmetry with both signs of c_2, one would need to use both D$p-4$-branes and their antibranes, and one would have lost this specificity.

The second point is that, as we move out from the zero configuration of the world-volume theory, we do not have identifications between naively distinct configurations. In fact this is false on T^4 as taking a D5-brane around a noncontractible cycle leads to such an identification. We should thus not take it for granted on $\mathbb{R}^4$.

The signal of this in D-brane physics is that, if we keep a reference brane at a point p in configuration space, moving a second brane to an image point p' will lead to $U(2)$ gauge symmetry and thus to additional massless states. In the problem at hand, this clearly must come from the higher dimensional gauge symmetry. In fact, the principle is more general than this – any discrete identifications on four dimensional field space must be the lower dimensional remnant of some higher dimensional gauge symmetry.

3 D-branes with stringy corrections

A Lagrangian with $\mathcal{N} = 1$ supersymmetry is not simply determined by the massless spectrum; explicit computation will be required to determine the superpotential. This opens the possibility that the moduli space of vacua can be modified by α' corrections.

A primary question is to what extent the results are geometric or not. The null hypothesis, called the "geometric hypothesis" in [16], is that despite all the apparent differences between string theory and the mathematics of vector bundles, the results will turn out to be the same, with α' corrections either absent or having only a quantitative effect, not changing the topology and structure of the space of supersymmetric vacua. This was true in the $\mathcal{N} = 2$ examples and the possibility should not be dismissed out of hand.

A more subtle possible outcome is that the stringy corrections can be summarized in terms of an *alternate* geometric picture. As we recalled above, mirror symmetry is the primary example of this. The stringy CY_3 moduli space is defined in terms of closed string theory, but mirror symmetry acts on the open strings and Dirichlet branes as well, turning the "B type" branes we have been discussing into "A type" branes on the mirror manifold. In the large volume limit, A branes are special Lagrangian submanifolds. Although not as well studied as holomorphic bundles, their theory is rapidly developing. We should thus ask: are special Lagrangian manifolds an alternate description of the configurations we are discussing, more suited to the stringy regime?

Stringy corrections to "protected" quantities come from world-sheet instantons (here with disk topology) and as such necessarily depend on the Kähler form of the CY_3. This allows us to summarize the stringy corrections in the two pictures as follows. In the B picture, the holomorphic bundle condition is uncorrected, and remains 2.2. On the other hand, the hermitian YM equations, which in $d = 4$ language led to the D-flatness conditions, get stringy corrections. Conversely, in the A picture, the Lagrangian condition and the F-flatness conditions must get corrections, while the "special" condition (the holomorphic 3-form Ω has constant phase) and D-flatness conditions are uncorrected. Indeed, one can show that the entire superpotential comes from disk instantons in this picture.

This leads one to an even more subtle *modified geometric hypothesis* [16], in which BPS branes and the corresponding $\mathcal{N} = 1$ supersymmetric vacua can be described entirely in geometric terms, in other words without needing to explicitly mention α' in the final results, but only by *combining* geometric input from the two mirror CY_3's.

This somewhat abstract statement was fleshed out in an example in [20]. One can understand almost all of the important points in the simplest stringy examples, orbifolds of flat space. Here computations are not hard, and much work has been done on $\mathbb{C}^3/\Gamma$ orbifolds for this reason. For compact CY, one can get a foothold in the stringy regime by considering Gepner models. These are LG orbifold models, with many similarities.

The world-volume theories of collections of D-branes in any of these spaces are supersymmetric quiver gauge theories. This is reviewed in detail in [20] and appears implicitly or explicitly in all of the D-brane literature, so we now assume the reader has some acquaintance with this.

Using these simple theories as a starting point, one can go on to get a picture of the set of BPS branes for general CY_3. The development of this picture is another long story, summarized in [18]. But the most important points are the answers to the questions I raised earlier in the ADHM discussion: namely, to what extent is supersymmetric field theory a valid description, and to what extent can one get a set of low energy theories whose vacua are in one-to-one correspondence to the higher dimensional vacua. Let us discuss what is understood about this.

The most important difference between the problems of finding BPS branes on T^4 and CY_3 is that it is no longer true that one can construct all of the BPS objects by combining branes from some chosen "basis" in a unique way. Let $K_0(\mathcal{M})$ be the additive group of D-brane charges, and K_{BPS} be the subset of these realized by BPS branes. Then, one would require that K_{BPS} be contained in some cone K_+ generated as non-negative combinations of rank $K_0(\mathcal{M})$ vectors, the charges of the basis branes, for this to work. However, even in the simplest non-trivial large volume examples, such as the one considered in [20], this is false.

This forces one to consider more general combinations of branes and antibranes. At this point one might despair of getting a description purely in low energy terms, as the description of brane-antibrane annihilation in such terms is notoriously difficult. Nevertheless one can get remarkably far by adding one new idea to supersymmetric field theory, namely a description of brane-antibrane annihilation in terms of the "derived category".

The entry of this admittedly rather abstract construction was presaged by Kontsevich's homological description of mirror symmetry [33]. It was justified physically in [5, 17], where it was argued that (B type) boundary conditions in topological open string theory on a CY_3 $\mathcal{M}$, which are the most general objects one might want to call "BPS D-branes", are precisely objects in the derived category of coherent sheaves on $\mathcal{M}$.

In fact the basic ideas behind this construction and even its detailed implementation do not require talking about sheaves or even Calabi-Yaus, but can be done purely in gauge theory terms; one can define the derived category of the configurations of a general supersymmetric quiver theory. To explain this, one needs to first explain the category of such configurations. Let us introduce the notation "object" for a complex gauge equivalence class of solutions to the F-flatness conditions of a quiver theory with specified ranks of the gauge groups. One then systematically studies the combined gauge theory of a pair of objects with gauge groups $U(N_i)$ and $U(N_i')$, a $U(N_i+N_i')$, to find its massless states. This is the data of the "category" of brane configurations which will enter the construction.

In particular, one has a notion of partial gauge equivalence or homomorphism between two objects E and E' whose fields are matrices we denote as $\phi_{i,j}$ and $\phi'_{i,j}$; these are off-diagonal gauge transformations ϵ_i in the (N_i, N_i') satisfying

$$\phi_{i,j}\epsilon_j = \epsilon_i\phi'_{i,j}.$$

This is a linear condition, so the homomorphisms form a linear space denoted $\mathrm{Hom}(E, E')$.

The role of partial gauge equivalence is the following: if there is a Hom(A, B), then A and B must each have some identical constituent, call it C. Therefore, if one combines the antibrane $\bar{A}$ with the brane B, one should see a tachyon whose condensation allows the constituents $\bar{C}$ and C to annihilate.

Going from the category of brane configurations to the derived category then amounts to a systematic way to identify all such processes which lead to the same result, independent of what $C\bar{C}$ might be. Its objects are "complexes", sequences of objects and homomorphisms

$$\ldots \longrightarrow E_0 \longrightarrow E_1 \longrightarrow E_2 \longrightarrow \ldots \, ,$$

which can be thought of as describing explicit series of tachyon condensation processes. Two such complexes which produce the same result are then identified, by regarding the partial gauge invariance which relates them as an actual gauge invariance.

One can then work with configurations containing both branes and antibranes, and this solves the problem we mentioned: any set of branes whose charges form a basis of the K theory becomes a reasonable candidate basis, and indeed necessary and sufficient conditions are known for a basis to work. If the basis branes appear with non-negative multiplicity, one can reduce the discussion back to ordinary supersymmetric gauge theory. But one can choose various different bases, and thereby get a set of different supersymmetric gauge theories which can each describe part of the full spectrum of BPS branes.

In general, the same brane configuration can typically be represented by non-negative combinations of branes from more than one basis, and thus has more than one supersymmetric gauge theory realization. However, this is not really an ambiguity in the description; the relation between these different gauge theories is Seiberg duality [8, 9, 12, 22]. In the larger CY_3 context, the same ideas lead to Fourier-Mukai transforms and other operations which act on the BPS brane spectrum, but only become symmetries when acting on the full derived category.

All this is still just talking about the F-flatness conditions; to fully solve the problem of characterizing supersymmetric vacua or BPS branes one must also solve the D-flatness conditions. By analogy to the DUY theorem, the first step towards this is to find the necessary and sufficient conditions for a solution to exist. A proposal has been made for these conditions [17, 19] called Π-stability, which incorporates the idea that these conditions are exact in the A brane picture. It reduces to μ-stability and θ-stability in the appropriate limits, and has allowed rederiving a number of the known results on the BPS spectrum and monodromy actions [3, 4].

To summarize, it is not true that one can translate holomorphic bundles on CY_3 and their stringy D-brane generalizations into the language of a single $\mathcal{N} = 1$ effective field theory. However, much progress has been made towards an enhanced $\mathcal{N} = 1$ effective field theory language, which provides a complete description by combining various Seiberg dual $\mathcal{N} = 1$ theories, all with the same derived category of supersymmetric configurations.

4 Quantum corrections

Clearly even a complete understanding of the spectrum and world-volume theories of BPS branes in weakly coupled type II is only a start on the general problem of $\mathcal{N} = 1$ string compactification. A direct way to continue would be to use these models as a starting point for defining type I and orientifold models, by implementing a world-sheet orientation reversal operator Ω and projecting on states satisfying $\Omega = 1$ in a way consistent with RR tadpole cancellation. This type of construction has been much studied in the language of conformal field theory, especially for orbifolds (see [2] for a review).

An important next step is to reformulate the Ω projection in the more geometric language we just outlined. Presumably, by suitably interpreting the projections that embed $SO(N)$ and $Sp(N)$ bundles in $U(N)$ bundles, one would get the general result. In terms of a bundle E and its dual bundle E^*, such a projection is an element $\Omega \in \mathrm{Hom}(E, E^*)$ which satisfies the condition $\Omega^*\Omega = \pm 1$. Reinterpreting this in terms of the sheaf or derived category languages should lead to a general prescription from which special cases would follow, such as the quotient procedures used in orbifold constructions [2, 21, 24].

Such a description might well exhibit nontrivial monodromies and dualities, including some subset of the type II results. But from the point of view of quantum corrections, this is just a starting point. If we restrict attention to the superpotential, such corrections will come from D-branes wrapping holomorphic cycles. At large volume, one can distinguish D1 brane and D5-brane corrections. The former are the S-dual of heterotic string world-sheet instanton corrections, while the latter are space-time instanton corrections. Whether this distinction makes sense in the stringy regime is at present not clear.

By now many nontrivial exact superpotentials are known, both in $\mathcal{N} = 1$ field theory and in compactifications of string and M theory. Generally arguments leading to these have two ingredients. One first needs some understanding of the leading quantum contribution, usually an instanton, to determine that it is nonzero and to see what symmetries it respects. Such computations are difficult to press beyond leading order however and at present all exact results come from other arguments. In field theory, one typically continues following Seiberg's approach using holomorphy and symmetry [38].

In string/M theory, one can also use duality to another theory with a classical superpotential. A prototypical example is the "MQCD" construction which realizes $\mathcal{N} = 1$ SYM using D4-branes suspended between NS5-branes in IIa theory [40]. This can be reinterpreted as a smooth 5-brane configuration in M theory, with 2 real dimensions embedded in 6 flat dimensions a non-trivial way. Now $\mathbb{R}^6 \cong \mathbb{C}^3$ is a CY_3 and there is a general formula for the classical superpotential in this situation: let the 2 dimensions embed into a subspace Σ, and let B be a 3-ball with boundary Σ; then

$$W = \int_B \Omega.$$

In the D-brane context this is a special case of 2.1, but using duality we have reinterpreted it as a quantum superpotential.

Another example is open string mirror symmetry, which reexpresses a sum over disk instantons on special Lagrangian branes as a classical B brane superpotential of the sort we discussed [1]. A further example is an open-closed string duality proposed by Gopakumar and Vafa, which allows understanding certain space-time instanton corrections [26]. The basic example is the theory of N D5-branes wrapping an isolated two-cycle in a CY_3; this is pure $\mathcal{N} = 1$ $U(N)$ SYM theory, which has N vacua distinguished by the phase of the gaugino condensate. They argue that this model is dual to a theory on a related CY obtained by a "conifold transition", *i.e.* blowing down the two-cycle and deforming to obtain a three-cycle, carrying N units of quantized flux. This theory has a "flux superpotential" which reproduces the quantum vacuum structure. Many generalizations of this result are known, involving branes at complicated singularities and product gauge theories. If it can be generalized to compactifications, this would clearly be major progress.

At present our understanding of the physics of superpotentials, especially in supergravity, is still somewhat primitive, and much can be learned by studying these classical examples. However, it would be truly amazing if the superpotential in a general string and M theory compactification could be determined by duality to a classical theory. A more realistic hope is that given enough results on "intrinsically quantum" superpotentials with some subset of the instantons treated exactly (one of the few results of this type is the F theory superpotential of [13]), combining these with global arguments of the sort used in field theory might give us general superpotentials.

In general, string world-sheet instantons still seem simpler than others, so an attractive laboratory for this problem is $(0,2)$ compactifications of the heterotic string at finite α'. This is S-dual to keeping the D1 corrections to type I discussed above, but perhaps combining space-time brane considerations and the heterotic world-sheet point of view will lead to further progress.

5 Towards the low energy theory of everything

Although the discussion we gave is a patchwork (and with important patches such as G_2 compactification completely left out), reflecting our primitive understanding of $\mathcal{N} = 1$ compactification, it does point towards a logical conclusion. Namely, that further progress along these lines, assuming only that we can solve problems comparable to those which have been solved in simpler theories, should someday allow us to find a complete description of *all* $d = 4$, $\mathcal{N} = 1$ vacua of string/M theory.

What would such a description look like? The simplest idea one might come up with is that there is some appropriate type of geometric object, say Calabi-Yau fivefolds carrying an X structure, which is in one-to-one correspondence with supersymmetric vacua. Or maybe one is looking for objects in the "ultra-derived category" of these, *i.e.* something which is not literally geometric but is so close that the existing methods of algebraic geometry would apply. This is an attractive prospect which should not be dismissed out of hand, and might even be true for $\mathcal{N} = 2$ compactifications [18]. For $\mathcal{N} = 1$ this hope seems at present to founder on the lack of a geometric picture for a general superpotential.

Barring something like this, the most concrete idea one can form at present is a "low energy effective field theory of everything", namely an $\mathcal{N} = 1$ effective supergravity Lagrangian with an infinite number of fields and (presumably) infinite rank gauge group, but with a definite superpotential and other structure, and rules which divide the configuration space into "regions" in each of which only a finite number of fields and gauge symmetries need to be considered. Supersymmetric vacua would then be those satisfying the standard F-flatness and D-flatness conditions[5]. There would of course be much to say about the problem of finding these vacua, but at this point the problem would have been turned into mathematics.

The D-brane discussion already points to one generalization of this picture which appears necessary. Namely, it is not natural to restrict attention to a single effective Lagrangian, but rather use a collection of Lagrangians, with certain regions of configuration space described by multiple Seiberg

[5]The closest thing to this in the literature is the $\mathcal{N} = 4$ effective theory of [25], which is a valid description of the moduli space and gauge symmetry (of course, there are simpler ways to give this data given $\mathcal{N} = 4$ supersymmetry). Thinking about this example is one way to see why one needs an infinite number of fields.

dual theories. This is important to understand in detail, but on the grand level of our present discussion amounts to more or less the same thing.

Is such a thing possible? The answer to this question should be given, at least in some mathematical sense, by the answer to the question I asked in Section 2, namely is the number of supersymmetric vacua finite or infinite. The reason is the general theorem that the union of a finite set of varieties, namely the various moduli spaces of vacua, is itself a variety, defined by a finite number of equations in a finite number of variables. These finite number of variables and equations would then be the finite number required in a specific region, while the "transition functions" between regions would just reflect the fact that going around non-trivial paths in configuration space can come back to a dual description of the same vacuum.

Another perspective on this problem is that it forms the primary application for one of the holy grails of this subject, namely the background independent formulation of string/M theory. Presumably this would play the role of all of the various ten and eleven dimensional theories of the present discussions, and its symmetries and configuration space would be directly reflected in the low energy theory in analogy to what we discussed.

These goals may seem hopelessly ambitious, and even if such a description existed, too complicated to ever make concrete and too unwieldy to be practically useful. But my reading of the history of string theory suggests the opposite point of view, that precise and sensible questions do have simple and pretty answers, even if we may not have the language to express them. We do know the language of $\mathcal{N} = 1$ effective field theory, and we are talking about a problem which (given finiteness) differs only in scale from those we can solve, so I conjecture that such a description will be found someday.

Given low energy supersymmetry, such a description might have some physical relevance, but of course it would not be the low energy theory of our world. Nevertheless it seems to me that this is the most natural (in the mathematical sense, *i.e.* involving the fewest arbitrary choices) problem coming out of string compactification which it seems reasonable to hope will someday have a complete solution.

References

[1] M. Aganagic, A. Klemm and C. Vafa, *Z. Naturforsch. A* **57** (2002) 1 [hep-th/0105045].

[2] C. Angelantonj and A. Sagnotti, Open strings, [hep-th/0204089].

[3] P.S. Aspinwall, *A point's point of view of stringy geometry* [hep-th/0203111].

[4] P.S. Aspinwall and M.R. Douglas, *D-brane stability and monodromy* [hep-th/0110071].

[5] P.S. Aspinwall and A.E. Lawrence, *JHEP* **0108** (2001) 004 [hep-th/0104147].

[6] J. Bagger and J. Wess, *Supersymmetry and Supergravity* (Princeton Univ. Press, 1990).
[7] T, Banks, M. Dine and L. Motl, *JHEP* **0101** (2001) 031 [hep-th/0007206].
[8] C.E. Beasley and M.R. Plesser, *JHEP* **0112** (2001) 001 [hep-th/0109053].
[9] D. Berenstein and M.R. Douglas, *Classical Seiberg Duality for Quiver Theories*, to appear.
[10] R. Bousso and J. Polchinski, *JHEP* **0006** (2000) 006 [hep-th/0004134].
[11] J.D. Brown and C. Teitelboim, *Phys. Lett. B* **195** (1987) 177.
[12] F. Cachazo, B. Fiol, K. Intriligator, S. Katz and C. Vafa, *A Geometric Unification of Dualities* [hep-th/0110028].
[13] R. Donagi, A. Grassi and E. Witten, *Mod. Phys. Lett. A* **11** (1996) 2199 [hep-th/9607091].
[14] S.K. Donaldson and P.B. Kronheimer, *The Geometry of Four-Manifolds* (Oxford Univ. Press, 1990).
[15] M.R. Douglas, Branes within branes, in *Cargese 1997, Strings, branes and dualities*, 267 [hep-th/9512077].
[16] M.R. Douglas, *Class. Quant. Grav.* **17** (2000) 1057 [hep-th/9910170].
[17] M.R. Douglas, *J. Math. Phys.* **42** (2001) 2818 [hep-th/0011017].
[18] M.R. Douglas, D-branes and $N = 1$ supersymmetry, to appear in the proceedings of Strings 2001: International Conference, Mumbai, India, 5–10 Jan. 2001 [hep-th/0105014].
[19] M.R. Douglas, B. Fiol and C. Römelsberger, *Stability and BPS branes* [hep-th/0002037].
[20] M.R. Douglas, B. Fiol and C. Römelsberger, *The spectrum of BPS branes on a noncompact Calabi–Yau* [hep-th/0003263].
[21] M.R. Douglas and G. Moore, *D-branes, Quivers, and ALE Instantons* [hep-th/9603167].
[22] B. Feng, A. Hanany, Y.-H. He and A.M. Uranga, *Toric Duality as Seiberg Duality and Brane Diamonds* [hep-th/0109063].
[23] J.L. Feng, J. March-Russell, S. Sethi and F. Wilczek, *Nucl. Phys. B* **602** (2001) 307 [hep-th/0005276].
[24] E.G. Gimon and J. Polchinski, *Phys. Rev. D* **54** (1996) 1667 [hep-th/9601038].
[25] A. Giveon and M. Porrati, *Nucl. Phys. B* **355** (1991) 422.
[26] R. Gopakumar and C. Vafa, *Adv. Theor. Math. Phys.* **2** (1998) 413 [hep-th/9802016].
[27] B.R. Greene, Lectures on the quantum geometry of string theory, in [29], 126.
[28] S. Hellerman and J. McGreevy, *JHEP* **0110** (2001) 002 [hep-th/0104100].
[29] *Les Houches 1995, Quantum symmetries,* edited by A. Connes and K. Gawędzki (North Holland, 1998), p. 519.
[30] M.B. Green, J.A. Harvey and G.W. Moore, *Class. Quant. Grav.* **14** (1997) 47 [hep-th/9605033].
[31] S. Kachru and E. Silverstein, *Nucl. Phys. B* **504** (1997) 272 [hep-th/9704185].
[32] A.D. King, *Quart. J. Math.* Oxford (2), **45** (1994) 515.
[33] M. Kontsevich, *Homological algebra of mirror symmetry* [alg-geom/9411018].
[34] M. Kreuzer and H. Skarke, *Complete classification of reflexive polyhedra in four-dimensions* [hep-th/0002240].
[35] J. Polchinski, *Phys. Rev. Lett.* **75** (1995) 4724 [hep-th/9510017].
[36] M. Reid, *Math. Ann.* **278** (1987) 329.

[37] V.A. Rubakov and M.E. Shaposhnikov, *Phys. Lett. B* **125** (1983) 139.

[38] N. Seiberg, The power of holomorphy: Exact results in 4-D SUSY field theories, in PASCOS 1994, 357 [hep-th/9408013].

[39] E. Witten, *Nucl. Phys. B* **460** (1996) 541 [hep-th/9511030].

[40] E. Witten, *Nucl. Phys. B* **507** (1997) 658 [hep-th/9706109].

[41] E. Witten, The Cosmological constant from the viewpoint of string theory, in *Marina del Rey 2000, Sources and detection of dark matter and dark energy in the universe*, 27 [hep-ph/0002297].

LECTURE 8

LECTURES ON OPEN STRINGS, AND NONCOMMUTATIVE GAUGE THEORIES

N.A. NEKRASOV

Institut des Hautes Études Scientifiques, Le Bois-Marie, Bures-sur-Yvette, F-91440 France
Institute for Theoretical and Experimental Physics, 117259 Moscow, Russia

Contents

LECTURES ON OPEN STRINGS, AND NONCOMMUTATIVE GAUGE THEORIES

N.A. Nekrasov

Abstract

In this notes the background independent formulation of the gauge theories on D-branes in flat space-time is considered, some examples of the solutions of their equations of motion are presented, the solutions of Dirac equation in these backgrounds are analyzed, and the generalizations to the orbifolded spaces are looked upon.

1 Introduction

We suggest to look for a generalized version of the gauge fields/strings correspondence [1] which may prove easier to establish (as sometimes general problems are easier then the particular ones). In the past few years a (new) connection has been (re-)discovered between noncommutative geometry [2] and string theory [3–6]. The previous understanding [7] of intrinsic noncommutativity of open string theory was supplemented by a vast number of examples stemming from the studies of D-branes, which allowed to make the noncommutativity manifest already at the field theory (or zero slope) limit. This connection may prove useful both ways. On the one hand, the noncommutative geometry is a deeply studied subject, thanks to the work of A. Connes and his followers. On the other hand, using the intuition/results from D-brane physics one can come up with new solutions/ideas for theories on noncommutative spaces.

These lecture notes should not be considered as an introduction into the noncommutative field theories and their relation to string theories. We refer the interested reader to [8]. Instead, we shall expand on several points not covered in [8].

We shall start with a unified construction of the worldvolume theories of D-branes of unspecified dimensionality. We shall discuss mostly the flat Minkowski space closed string background. In the concluding section we present a few results on curved backgrounds: the orbifolds of flat space, their deformations, the conifolds. The more abstract approach to classification of D-brane states on rational conformal field theories will not be covered here (see *cf.* [9, 10])

In the mean time we shall discuss several classical solutions of the noncommutative gauge theory. They correspond to flat D-branes, D-branes at angles, D-branes with different magnetic fields turned on, D-branes of various dimensions. Generically these solutions are unstable (open string spectrum contains a tachyon). In the noncommutative gauge theory the instability is reflected by the negative modes in the expansion around the solution. In principle one should be able to study the decay of the solution towards the stable ones. However the application of this analysis to string theory is limited, as in the majority of the cases the α'-corrections to the flow may not be negligible.

We shall also consider in some detail the case of four dimensional instanton/monopole solutions. In each case we shall construct the solutions of the Dirac equation in the background of the instanton/monopole. In the case of instantons the analysis of the solutions of the Dirac equation permits to establish the completeness of the noncommutative version of ADHM construction [11]. This is a noncommutative version of the reciprocity of [12].

In the course of these lectures we shall try to make clear that the noncommutative algebras, which should be thought of the algebras of functions on the noncommutative manifold, are not fixed, but rather arise upon a choice of classical solution in the background independent formulation of the gauge theory. We shall show that this background independent formulation is on the one hand related to the background independent open string field theory [13]. On the other hand it is related to Matrix theory.

2 Background independence

It was observed by many authors, see *cf.* [14, 15] that the Matrix theory action (or its euclidean version [16]) provides a background independent formulation of the large class of gauge theories. We shall now state this more precisely, and at the same time we shall fix our notations. Consider an abstract Hilbert space $\mathcal{H}$ and a collection of d Hermitian operators Y^i, $i = 1, \ldots, d$, acting there. We shall sometimes use the notation $\mathbf{Y}$ for the vector in $V = \mathbf{R}^d$ with values in the space $\mathbf{H}(\mathcal{H})$ of Hermitian operators in $\mathcal{H}$. Let g_{ij} be a Euclidean metric on V, Define a formal action:

$$S = -\frac{1}{4g_{\mathrm{YM}}^2} \sqrt{g}\; g_{ik}\; g_{jl}\; \mathrm{Tr}_{\mathcal{H}}[Y^i, Y^j][Y^k, Y^l]. \tag{2.1}$$

We said (2.1) is formal because it may happen that for a physically sensible choice of $\mathbf{Y}$ the value of S is infinite. We shall consider $\mathbf{Y}$ such that with appropriate choice of a constant μ, the action becomes finite upon subtraction of μ $\mathrm{Tr}_{\mathcal{H}}\mathbf{1}$ (more precisely, we subtract $\mu\mathbf{1}$ from $[Y,Y]^2$ before calculating the trace).

The action (2.1) has an obvious gauge symmetry (which is consistent with the infrared regularization above):

$$Y^i \mapsto g^\dagger Y^i g,\ g^\dagger g = gg^\dagger = \mathbf{1}. \tag{2.2}$$

The equations of motion following from (2.1) are:

$$g_{kl}[Y^k,[Y^l,Y^j]] = 0,\ j = 1,\dots d. \tag{2.3}$$

The operators Y^i generate some associative algebra $\mathcal{A}$. Let us define the so-called *Yang-Mills algebra* $\mathcal{A}_{\mathrm{YM,d}}$ to be the associative algebra generated by Y^i, subject to the relations (2.3).

We shall now pause to establish the meaning of the operators Y^i in the open string theory context.

To this end let us consider open superstring propagating in the flat $\mathbf{R}^{9,1}$. Consider the worldsheet of the disk topology. Let z, $|z|\leq 1$ be the holomorphic coordinate on this disk D. Let $z = \mathrm{e}^{-r+i\sigma}$. The fields X^μ will have the following boundary conditions:

$$\begin{aligned} &\partial_{\mathrm{r}}X^\mu = 0, \qquad \mu = 0,1\dots,p, \qquad\qquad r = 1\\ &\partial_\sigma X^\mu = 0, \qquad \mu = p+1,\dots,25, \qquad r = 1. \end{aligned} \tag{2.4}$$

These boundary conditions describe the single flat Dp-brane. If we want to have several parallel Dp-branes then we should allow for the Chan-Paton factors, say $i = 1,\dots,N$, so that the (constant) value of $X^\mu(r=1,\sigma) = \varphi^\mu$, $\mu = p+1,\dots 25$ may depend on i: φ_i^μ.

The canonical way of setting up a string calculation in these circumstances is to consider the dimension 1 vertex operators on the boundary, evaluate their correlation function, and integrate it over the moduli space of points on the boundary of the disk. One can also add the closed string vertex operators into the interior of the disk. These operators should have the dimension $(1,1)$. The open string vertex operators in general change the boundary conditions, *i.e.* the boundary conditions corresponding to the Chan-Paton index i to the left of the vertex operator may be followed by the j'th boundary condition to the right of the operator. This is, of course, reflected by the contribution to the dimension of the operator of the mass squared of the stretched string: $m_{ij}^2 = \|\varphi^i - \varphi^j\|^2$.

We would like to have a setup in which there is no need to specify in advance neither the values of φ^i, nor p or N. All this data will be encoded in the properties of the operators Y^μ, $\mu = 0,\dots,9$, acting in some auxiliary Hilbert space $\mathcal{H}$.

Consider the correlation function of a closed string vertex operator $\mathcal{O}$ inserted at the center $z = 0$ of the disk D, and the boundary operator, generalizing the usual supersymmetric Wilson loop:

$$\mathcal{Z}\left[\mathcal{O}|Y\right] = \left\langle \mathcal{O} \exp\left(-\frac{1}{4\pi\alpha'}\int_D g_{ij}\left(\partial x^i\bar{\partial}x^j + \psi^i\bar{\partial}\psi^j + \tilde{\psi}^i\partial\tilde{\psi}^j\right)\right)\right.$$
$$\left.\times\ \mathrm{Tr}_{\mathcal{H}}\left(P\exp\oint_{\partial D} ik_i\left(Y^i - x^i\mathbf{1}\right) + \vartheta_i\Psi^i + \vartheta_i\vartheta_j[Y^i,Y^j]\right)\right\rangle \cdot \qquad (2.5)$$

Here k_i, ϑ_i are the momenta conjugate to x^i, $\Psi^i = \psi^i + \tilde{\psi}^i$: $k_i = g_{ij}\partial_n x^j$, $\vartheta_i = \psi^i - \tilde{\psi}^i$. For example, for the graviton [17]: $\mathcal{O}_{\mathrm{h}} = h_{ij}(p) : \psi^i\tilde{\psi}^j \mathrm{e}^{ip\cdot x}$: (times the ghosts and superghosts):

$$\mathcal{Z}\left[\mathcal{O}_{\mathrm{h}}|Y\right] = g_{kl}\int \mathrm{d}p\mathrm{e}^{-ip\cdot x}h_{ij}(p)$$
$$\times\int_0^1 \mathrm{d}s\ \mathrm{Tr}_{\mathcal{H}}\mathrm{e}^{isp\cdot Y}[Y^i,Y^k]\mathrm{e}^{i(1-s)p\cdot Y}[Y^j,Y^l] + o(\alpha').$$

It can be shown that the condition that the boundary interaction (2.5) is consistent with the conformal invariance of the worldsheet sigma model reads as:

$$g_{ij}[Y^i,[Y^j,Y^k]] = 0(\alpha'). \qquad (2.6)$$

We can think of (2.6) as defining a one-parametric family of the associative algebras, generated by Y^i. We shall call them the *algebras of functions on the D-brane worldvolume.*

It should be straightforward to establish a direct relation between the background independent formulation of the noncommutative gauge theories *via* Y^i's, and the background independent open string field theory [13].

In the sequel we shall study the case $\alpha' \to 0$, *i.e.* Yang-Mills algebras.

An obvious class of *Yang-Mills algebras* is provided by the Heisenberg-Weyl algebras, where the generators Y^i obey the stronger condition:

$$[Y^i,Y^j] \in \mathrm{center}(\mathcal{A}_{\mathrm{YM,d}}). \qquad (2.7)$$

Clearly, for $d = 2$ all Yang-Mills algebras are at the same time Heisenberg-Weyl. For generic Heisenberg-Weyl solution let us denote

$$Z^{ij} = [Y^i,Y^j], \qquad (2.8)$$

we have $[Z^{ij},Y^k] = 0$ for any i,j,k. Let us diagonalize Z^{ij}:

$$\mathcal{H} = \bigoplus_A H_A, \qquad Z^{ij}|_{H_A} = i\theta_A^{ij}\ \cdot 1_{H_A}. \qquad (2.9)$$

Then each subspace H_A is an irreducible representation of the Heisenberg algebra $[x^i_A, x^j_A] = i\theta^{ij}_A$.

We now wish to describe the spectrum of fluctuations around the solution (2.8, 2.9). To this end we can decompose the fluctuation $y^i = \delta Y^i$ as follows:

$$y^i = \sum_{A,B} y^i_{AB}, \qquad y^i_{AB} \colon H_B \to H_A. \tag{2.10}$$

We shall also impose the following gauge condition:

$$g_{ij}[Y^i, y^j] = 0 \leftrightarrow \sum_{A,B} g_{ij} \left(x^i_A y^j_{AB} - y^i_{AB} x^j_B \right) = 0. \tag{2.11}$$

The linearized fluctuations are governed by the quadratic approximation to the action:

$$\mathcal{K} y^j = g_{ik}[x^i, [x^k, y^j]] + 2g_{ik}[y^i, Z^{kj}] \tag{2.12}$$

which leads to the following eigenvalue problem for the spectrum of masses:

$$\begin{aligned} \omega^2 \; y^j_{AB} &= \left(\Delta_A y^j + y^j \Delta_B - 2g_{ik} x^i_A y^j_{AB} x^k_B + 2i y^i_{AB} \left(T^j_{i,B} - T^j_{i,A} \right) \right) \\ \Delta_A &= g_{ij} x^i_A x^j_A, \qquad T^j_{i,A} = \theta^{kj}_A g_{ik}. \end{aligned} \tag{2.13}$$

We now proceed with some examples.

2.1 Dolan-Nappi solutions

This solution describes two branes which could sit on top of each other, and have different magnetic fields turned on.

Set $d = 2$, $\theta^{ij}_{A,B} = \theta_{A,B}\epsilon^{ij}$, $A, B = 1, 2, i, j = 1, 2$, $g_{ij} = \delta_{ij}$. The operators $\mathbf{Y}$ corresponding to this solution are given by:

$$Y^1 + iY^2 = \begin{pmatrix} \sqrt{2\theta_1} a_1 & 0 \\ 0 & \sqrt{2\theta_2} a_2 \end{pmatrix} \tag{2.14}$$

where a_1, a_2 are the annihilation operators acting in two (isomorphic) copies $\mathcal{H}_1, \mathcal{H}_2$ of the Hilbert space $\mathcal{H}$.

To facilitate the analysis let us map the operators $L_{AB}\colon H_A \to H_B$ to the vectors in the tensor product: $H_B \otimes H^*_A \approx H \otimes H$:

$$L_{AB} \mapsto \sum_{n_1, n_2} \langle n_2 | L_{AB} | n_1 \rangle \quad |n_2, n_1\rangle \cdot$$

Also, introduce the notation $\zeta = y^1_{AB} + iy^2_{AB}, \bar{\zeta} = y^1_{AB} - iy^2_{AB}$. First, assume that $\theta_{A,B} > 0$ and introduce the creation-annihilation operators:

$$x^1_{A,B} + ix^2_{A,B} = \sqrt{2\theta_{A,B}}a_{1,2}$$

and the number operators $n_\alpha = a^+_\alpha a_\alpha$. Then (upon the identification $\zeta \in H_1 \otimes H_2$)

$$\mathcal{K}\frac{\zeta}{\bar{\zeta}} = \\ 2\theta_A(n_1 + \tfrac{1}{2}) + 2\theta_B(n_2 + \tfrac{1}{2}) - 2\sqrt{\theta_A\theta_B}\left(a^+_1 a^+_2 + a_1 a_2\right) \pm 2(\theta_A - \theta_B)\frac{\zeta}{\bar{\zeta}}\,. \tag{2.15}$$

This operator is conveniently diagonalized by introduction of the $SU(1,1)$ generators:

$$L_+ = a^+_1 a^+_2,\ L_- = a_1 a_2, L_0 = \tfrac{1}{2}\left(n_1 + n_2 + 1\right) \tag{2.16}$$

and the operator

$$M = \tfrac{1}{2}\left(n_1 - n_2\right).$$

Now, the spectrum of

$$\mathcal{K} = 2(\theta_A - \theta_B)(M \pm 1) + 2(\theta_A + \theta_B)L_0 - 2\sqrt{\theta_A\theta_B}(L_+ + L_-). \tag{2.17}$$

depends on whether θ_A equals θ_B or not. If $\theta_A - \theta_B \neq 0$ then upon a Bogolyubov $SU(1,1)$ transformation we can bring $\mathcal{K}$ to the form:

$$2(\theta_A - \theta_B)(M \pm 1) + 2|\theta_A - \theta_B|L_0 \tag{2.18}$$

whose spectrum is (remember that the spectrum of L_0 is given by: $L_0 = \frac{1}{2} + |M| + k, \quad k \in \mathbf{Z}_+$):

$$2(\theta_A - \theta_B)(\pm 1 + M) + |\theta_A - \theta_B|(1 + 2|M| + 2k) \tag{2.19}$$

which contains a tachyonic mode (for ζ or $\bar{\zeta}$ depending on the sign of $\theta_A - \theta_B$).

For $\theta_A = \theta_B = \theta$ the operator $\mathcal{K}$ is manifestly positive definite:

$$\mathcal{K} = 2\theta bb^\dagger, \qquad b = a_1 - a^+_2,\ b = a^+_1 - a_2$$

and its spectrum is continuous:

$$\mathcal{K} = 2\theta|\kappa|^2$$

with the eigenvectors:

$$e^{i(\kappa a_1^+ + \bar{\kappa} a_2^+) + a_1^+ a_2^+} |0,0\rangle$$

which in the ordinary, operator, representation correspond to the plane waves:

$$e^{i(\kappa a^\dagger + \bar{\kappa} a)}.$$

2.2 Intersecting branes

As another interesting example of the solution (2.7) we shall look at the case $d = 4$:

$$\begin{aligned} \theta_A^{12} &= \theta_1 \delta_{A,1} \\ \theta_A^{34} &= \theta_2 \delta_{A,2}. \end{aligned} \tag{2.20}$$

This solution describes two branes, each having two dimensions transverse to another. The operators Y^i corresponding to this solution have the following block-diagonal form:

$$Y^1 + iY^2 = \sqrt{2\theta_1}\begin{pmatrix} a_1 & 0 \\ 0 & 0 \end{pmatrix}, \; Y^3 + iY^4 = \sqrt{2\theta_2}\begin{pmatrix} 0 & 0 \\ 0 & a_2 \end{pmatrix} \tag{2.21}$$

where we denote by a_1, a_2 the annihilation operators acting in two copies $\mathcal{H}_1, \mathcal{H}_2$ of the Hilbert space $\mathcal{H}$.

In this case the spectrum of fluctuations contains in the AB sector the discrete modes, corresponding to the strings localized at the intersection of the branes, and, if $\theta_1 \neq \theta_2$, starts off with the tachyonic mode:

$$\omega^2 = 2\theta_1 n_1 + 2\theta_2 n_2 \pm |\theta_1 - \theta_2|, \; n_{1,2} \in \mathbf{Z}_+. \tag{2.22}$$

2.3 T-duality

The configuration from the previous example is closely related to the "piercing string" solution of [18], which in our present notation is described as follows: let as before $\mathcal{H}_1, \mathcal{H}_2$ denote two copies of the Hilbert space $\mathcal{H}$. Now, let $\mathcal{H}_3 = \mathcal{H}_1 \otimes \mathcal{H}_2$. Then the operators $\mathbf{Y}$ will be acting in $\mathcal{H}_3 \oplus \mathcal{H}_2$:

$$Y^1 + iY^2 = \sqrt{2\theta_1}\begin{pmatrix} a_1 \otimes 1 & 0 \\ 0 & 0 \end{pmatrix}, \; Y^3 + iY^4 = \sqrt{2\theta_2}\begin{pmatrix} 1 \otimes \frac{1}{2}(a_2 + a_2^+) & 0 \\ 0 & a_2 \end{pmatrix}. \tag{2.23}$$

The analysis of fluctuations around this solution is similar, it also exibiths a tachyonic mode for $\theta_1 \neq \theta_2$.

3 BPS algebras

Less trivial is the case of the so-called *BPS algebras* (not to be confused with the algebras of BPS states), whose generators obey the following relations: let $S_\pm$ be the spaces of chiral spinors of $SO(d)$, let γ_i be the generators of the Clifford algebra,

$$\{\gamma_i, \gamma_j\} = g_{ij}\ \mathbf{1}, \tag{3.1}$$

then the relations state that there exist two spinors $\epsilon_1, \epsilon_2 \in S_+$ such that

$$[Y^i, Y^j][\gamma_i, \gamma_j]\ \epsilon_1 + \mathbf{1}\ \epsilon_2 = 0. \tag{3.2}$$

If $d = 4$ then these relations have the following simple form:

$$[Y^i, Y^j] \pm \tfrac{1}{2}\varepsilon_{mnkl}\sqrt{g}\ g^{mi}\ g^{nj}[Y^k, Y^l] \in \mathrm{center}(\mathcal{A}_{YM,4}). \tag{3.3}$$

The operators Y^i solving (3.3) considered up to a gauge transformation (2.2), define the *noncommutative instanton.*

3.1 Noncommutative $U(1)$ instantons

The construction of noncommutative instantons [11] is a generalization of the famous ADHM procedure, which produces an anti-self-dual gauge field on $\mathbf{R}^4$ (or its one-point compactification, $\mathbf{S}^4$, given a solution to a finite-dimensional version of the anti-self-duality condition. The latter is imposed on the set of complex matrices, B_α, $\alpha = 1, 2$, and (in the $U(1)$ case) I, where B_α are the operators in a vector space $V = \mathbf{C}^k$, where k is the instanton charge, while I is a vector in V. The conditions, imposed on (B_α, I) are:

$$\begin{aligned}&[B_1, B_2] = 0\\ &[B_1, B_1^\dagger] + [B_2, B_2^\dagger] + II^\dagger = 2 \cdot \mathbf{1}_V.\end{aligned} \tag{3.4}$$

The 2 in the right-hand side of (3.4) is a convenient choice of normalization, which could be altered – what matters is whether there stands 0 or something positive. Given a solution to (3.4), one can generate another one by applying a $U(k)$ transformation $(B_\alpha, I) \mapsto (g^{-1}B_\alpha g, g^{-1}I)$ with $g \in U(k)$. Such solutions will be considered as equivalent ones.

The purpose of this note is to elucidate the meaning of the space V from the point of view of four dimensional noncommutative gauge theory. We shall see that V is nothing but the space of normalizable solutions to Dirac equation for the spinor field in the fundamental representation, in the instanton background.

We now recall the construction of the instanton gauge field. Consider the associative algebra $\mathbf{R}^4_\theta$ of operators in the Hilbert space $\mathcal{H} = L^2(\mathbf{R}^2)$, which we identify with the space of states of a two-dimensional harmonic oscillator. Let a_α, $a^\dagger_\alpha$, $\alpha = 1, 2$ be the annihilation and creation operators, respectively, which obey the algebra:

$$[a_\alpha, a^\dagger_\beta] = \delta_{\alpha\beta}. \tag{3.5}$$

Consider the $\mathbf{R}^4_\theta$-module $M = \mathcal{H} \otimes V$, and let Δ, $\tilde{\Delta}$ denote the operators in M:

$$\begin{aligned}
\Delta &= \sum_\alpha (B_\alpha - a^\dagger_\alpha)(B^\dagger_\alpha - a_\alpha) \\
\tilde{\Delta} &= \sum_\alpha (B^\dagger_\alpha - a_\alpha)(B_\alpha - a^\dagger_\alpha) \\
\tilde{\Delta} - \Delta &= II^\dagger.
\end{aligned} \tag{3.6}$$

Let $\mathcal{D}^\dagger$: $M \otimes \mathbf{C}^2 \oplus \mathcal{H} \to M \otimes \mathbf{C}^2$ be a morphism of $\mathbf{R}^4_\theta$ modules, given by:

$$\mathcal{D}^\dagger = \begin{pmatrix} B_1 - a^\dagger_1 & B_2 - a^\dagger_2 & I \\ -B^\dagger_2 + a_2 & B^\dagger_1 - a_1 & 0 \end{pmatrix}. \tag{3.7}$$

It follows from (3.4) that $\mathcal{D}^\dagger \mathcal{D} = \tilde{\Delta} \otimes \mathbf{1}_{\mathbf{C}^2}$: $M \otimes \mathbf{C}^2 \to M \otimes \mathbf{C}^2$. One shows [19] that $\tilde{\Delta}$ is a positive definite Hermitian operator in M. Hence, the following operator is a well-defined projector in $M \otimes \mathbf{C}^2 \oplus \mathcal{H}$:

$$\frac{\Pi_1 = \mathcal{D}1}{\mathcal{D}^\dagger \mathcal{D}} \mathcal{D}^\dagger . \tag{3.8}$$

Let Ψ denote the fundamental solution to the equation $\mathcal{D}^\dagger \Psi = 0$, *i.e.* a morphism of $\mathbf{R}^4_\theta$ modules:$\Psi : \mathcal{H} \to M \otimes \mathbf{C}^2 \oplus \mathcal{H}$. One shows that Ψ can be normalized so as to define a unitary isomorphism between $\mathcal{H}$ and the kernel of $\mathcal{D}^\dagger$: $\Psi^\dagger \Psi = \mathbf{1}_{\mathcal{H}}$:

$$\begin{aligned}
&\Psi = \begin{pmatrix} \psi_1 \\ \psi_2 \\ \xi \end{pmatrix}, \\
&\psi_\alpha = (B^\dagger_\alpha - a_\alpha) v \\
&\Delta v = \quad -I\xi \\
&\xi = \Lambda^{-\frac{1}{2}} S^\dagger \\
&\Lambda = 1 + I^\dagger \frac{1}{\Delta} I = \frac{I^\dagger \frac{1}{\Delta} I}{I^\dagger \frac{1}{\tilde{\Delta} I}} \\
&\Lambda^{-1} = 1 - I^\dagger \tilde{\Delta}^{-1} I \\
&S^\dagger : \mathcal{H} \to \mathcal{H}, \quad SS^\dagger = 1, \quad S^\dagger S = 1 - P
\end{aligned} \tag{3.9}$$

where P is a orthogonal projection in $\mathcal{H}$ onto a subspace, isomorphic to V, which is spanned by the elements η which are in the image of the operator $I^\dagger \exp\left(\sum_\alpha B_\alpha^\dagger a_\alpha^\dagger\right)|0,0\rangle\colon V \to \mathcal{H}$, where $|0,0\rangle$ is the vacuum state in $\mathcal{H}$.

It is instructive to show that this image can be also characterized as the kernel of the operator Λ^{-1}. Indeed, let $\lambda \in \mathcal{H}$ be such that $\lambda = I^\dagger \tilde{\Delta}^{-1} I\lambda$. It follows:

$$\begin{aligned} I\lambda = II^\dagger\tilde{\Delta}^{-1}I\lambda \ \Rightarrow\ & \Delta\tilde{\Delta}^{-1}I\lambda = \left(\tilde{\Delta} - II^\dagger\right)\tilde{\Delta}^{-1}I\lambda = 0 \\ & \Rightarrow \tilde{\Delta}^{-1}I\lambda = \exp\left(\sum_\alpha B_\alpha^\dagger a_\alpha^\dagger\right)|0,0\rangle \otimes \nu, \qquad \nu \in V \\ & \Rightarrow I\left(\lambda - I^\dagger \exp\left(\sum_\alpha B_\alpha^\dagger a_\alpha^\dagger\right)|0,0\rangle \otimes \nu\right) = 0. \end{aligned} \tag{3.10}$$

In writing the formulae (3.9) we only had to use the operator Λ^{-1} which is an element of $\mathbf{R}_\theta^4$. However, for computational purposes it is useful to work with Λ as well. Technically one has to localize $\mathbf{R}_\theta^4$ over Λ, *i.e.* consider the formal polynomials of the form $\sum_n a_n\Lambda^n$ where a_n are the elements of $\mathbf{R}_\theta^4$.

One defines the second projector:

$$\Pi_2 = \Psi\Psi^\dagger. \tag{3.11}$$

It is clear from the positivity of $\tilde{\Delta}$, that:

$$\Pi_1 + \Pi_2 = \mathbf{1}_{M\otimes\mathbf{C}^2\oplus\mathcal{H}}. \tag{3.12}$$

This relation implies that the following identities hold:

$$\begin{aligned} &(B_1^\dagger - a_1)\frac{1}{\Delta}(B_1 - a_1^\dagger) + (B_2 - a_2^\dagger)\frac{1}{\tilde{\Delta}}(B_2^\dagger - a_2) = \mathbf{1}_k \\ &(B_2^\dagger - a_2)\frac{1}{\Delta}(B_2 - a_2^\dagger) + (B_1 - a_1^\dagger)\frac{1}{\tilde{\Delta}}(B_1^\dagger - a_1) = \mathbf{1}_k \\ &\left(I^\dagger\frac{1}{\tilde{\Delta}} - \Lambda^{-1}I^\dagger\frac{1}{\Delta}\right)(B_\alpha - a_\alpha^\dagger) = 0. \end{aligned} \tag{3.13}$$

Now, define operators in $\mathcal{H}$

$$A_\alpha = \Psi^\dagger a_\alpha \Psi, \qquad A_\alpha^\dagger = \Psi^\dagger a_\alpha^\dagger \Psi. \tag{3.14}$$

One shows [11, 19] that

$$\begin{aligned} &[A_1, A_2] = 0, \quad [A_1^\dagger, A_2^\dagger] = 0 \\ &[A_1, A_1^\dagger] + [A_2, A_2^\dagger] = 2. \end{aligned} \tag{3.15}$$

3.2 Higher dimensional instantons

We now proceed with discussing BPS solutions involving more then four Y's. In general, for the $U(N)$ gauge theory on a p complex dimensional Kähler manifold X the natural analogues of the instanton equations are the so-called Hermitian Yang-Mills equations, which state that the curvature of the gauge field A is of the type $(1,1)$ and that its nonabelian part is primitive, that is orthogonal to the Kähler form ω:

$$\begin{aligned} &F^{(2,0)} = 0 \\ &F \wedge \omega^{p-1} = \lambda\, \mathbf{1}\, \omega^{p} \end{aligned} \tag{3.16}$$

where λ is a constant, which can be computed from the first Chern class of the gauge bundle.

We shall now consider the noncommutative analogues of the equation (3.16). Let as introduce a complex structure on $\mathbf{R}^d$ such that the noncommutativity tensor θ is of the type $(1,1)$. We shall for simplicity assume that it actually related to the Kähler form: $\theta = \omega^{-1}$. The equation (3.16) will now read:

$$\begin{aligned} &[Y^\alpha, Y^\beta] = 0,\ \alpha,\beta = 1,\dots,p \\ &\sum_\alpha [Y^\alpha, Y^{\alpha,\dagger}] = p\cdot \mathbf{1} \end{aligned} \tag{3.17}$$

(we have normalized things in such a way that $Y^\alpha = a_\alpha$, with $[a_\alpha, a^\dagger_\beta] = \delta_{\alpha\beta}$ is a solution to (3.17)). Then it is relatively easy to produce a $U(p)$ invariant solution to (3.17) with positive action:

$$\begin{aligned} &Y^\alpha = S a_\alpha \left(1 - \frac{p!}{(N)_p}\right)^{\frac{1}{2}} S^\dagger \\ &N = \sum_\alpha a_{\alpha,\dagger} a_\alpha \\ &(N)_p = N(N+1)\dots(N+p-1) \\ &SS^\dagger = \mathbf{1},\ S^\dagger S = \mathbf{1} - |0\rangle\langle 0|. \end{aligned} \tag{3.18}$$

The action on this solution is finite:

$$S_p = \mathrm{Tr}_{\mathcal{H}} \left([Y^\alpha, Y^{\beta,\dagger}] - \delta^{\alpha\beta}\right)\left([Y^\beta, Y^{\alpha,\dagger}] - \delta^{\alpha\beta}\right) = p(p-1). \tag{3.19}$$

The topological charges associated with this solution are:

$$\begin{aligned} &ch_r = \mathrm{Tr}_{\mathcal{H}} F^{\wedge r} \wedge \omega^{p-r} \\ &F_{ij} = \omega_{ik}\omega_{jl}\left([Y^k, Y^l] - i\theta^{kl}\right) \\ &ch_1 = 0 \\ &ch_2 = S_p = p(p-1). \end{aligned} \tag{3.20}$$

4 Fermions in the $\mathbf{Y}$ background

Given a generic $\mathbf{Y}$ we define *Dirac operator* $\not{D}$: $S_+ \otimes \mathcal{A} \longrightarrow S_- \otimes \mathcal{A}$ by the formula:

$$\not{D}\psi = \gamma_i [Y^i, \psi]. \tag{4.1}$$

Now consider a background $\mathbf{Y} = \underline{Y} \oplus \underline{Y}'$, which splits as a direct sum of two independent Y-backgrounds. The Dirac operator, as defined by (4.1) splits as a sum of four independent operators. The most interesting for us is the "off-diagonal" one:

$$\not{D}\chi = \gamma_i \left(\underline{Y}^i \chi - \chi \underline{Y}'^{,i} \right). \tag{4.2}$$

Now suppose that $\underline{Y}'$ is a frozen vacuum solution, while $\underline{Y}$ is a dynamical gauge field. The off-diagonal component χ of the fermion ψ which enters (4.2) is what is sometimes called the fermion in the *fundamental* representation (as opposed to the *adjoint* fermion in (4.1)).

4.1 Fermions in the instanton background

We are now interested in solution of the Dirac equation in the instanton background for the fermions in the fundamental representation:

$$\not{D}\chi = \begin{pmatrix} \nabla_1 & -\nabla_2 \\ \nabla_{\bar{1}} & \nabla_{\bar{2}} \end{pmatrix} \begin{pmatrix} \chi_1 \\ \chi_2 \end{pmatrix} = 0 \tag{4.3}$$

where the covariant derivative of a field χ in the (right) fundamental representation is given by:

$$\nabla_\alpha \chi = -A_\alpha^\dagger \chi + \chi a_\alpha^\dagger, \quad \nabla_{\bar{\alpha}} \chi = A_\alpha \chi - \chi a_\alpha. \tag{4.4}$$

We claim that the fundamental solution χ to (4.3) (which should be thought as of the morphism of the $\mathbf{R}^4_\theta$ modules:

$$\chi: M \to \mathcal{H} \otimes \mathbf{C}^2$$

upon identification of the space of solutions with V, which is done essentially in the same way as in [12]) is given by:

$$\chi_\alpha = v^\dagger (B_\alpha - a_\alpha^\dagger) \tilde{\Delta}^{-1} = -S \Lambda^{\frac{1}{2}} \partial_\alpha \left(I^\dagger \tilde{\Delta}^{-1} \right). \tag{4.5}$$

It is easy to check (4.3) using the identities (3.13). Notice the following normalization of the solution (4.5):

$$\sum_\alpha \chi_\alpha^\dagger \chi_\alpha = \sum_\alpha [a_\alpha^\dagger, [a_\alpha, \tilde{\Delta}^{-1}]] \Rightarrow \mathrm{Tr}_{\mathcal{H}} \sum_\alpha \chi_\alpha^\dagger \chi_\alpha = \mathbf{1}_V \tag{4.6}$$

(in the last equality we used the fact that the trace of the double commutator will not change if we set $B_\alpha = B_\alpha^\dagger = 0$). By similar calculations one shows, that:

$$\mathrm{Tr}_{\mathcal{H}} \sum_\alpha \chi_\alpha^\dagger \chi_\alpha \begin{pmatrix} a_\beta \\ a_\beta^\dagger \end{pmatrix} = \begin{pmatrix} B_\beta^\dagger \\ B_\beta \end{pmatrix}. \tag{4.7}$$

Now, given an instanton gauge field $Y^i = (A_\alpha, A_\alpha^\dagger)$ we consider the associated Dirac operator $\not{D}$ and the space V of its normalizable zero-modes. We shall require that Y^i obey the following asymptotics:

$$Y^i = Sx^i S^\dagger + y^i \tag{4.8}$$

where $SS^\dagger = \mathbf{1}$, and the eigenvalues λ_n^i of y^i decay faster then $\frac{1}{n}$ (after some reshuffling). Similarly, we are interested in the solutions χ to the Dirac equation, which have the form:

$$\chi_\alpha = S a_\alpha^\dagger \frac{1}{(aa^\dagger)^2} + \ldots \tag{4.9}$$

where $\ldots$ denote subleading (in the same sense as for the gauge field, except that with $\frac{1}{n^2}$ instead of $\frac{1}{n}$ asymptotics) terms. One shows, analogously to [12] that the space of these solutions is of the dimension k, given by the instanton charge, and that the formulae (4.7) produce ADHM matrices $B, B^\dagger$, while the asymptotics

$$\chi_\alpha = -ISa_\alpha^\dagger \frac{1}{(aa^\dagger)^2} \tag{4.10}$$

gives I. By the completely analogous calculations to [12] one proves that $B, B^\dagger, I$ obey (3.4). This concludes the reciprocity.

4.2 The Dirac field in the monopole background

We now present the solution $\chi = \begin{pmatrix} \chi_+ \\ \chi_- \end{pmatrix}$ to the Dirac equation

$$2\, D_c \chi_+ - (D_3 + \Phi - z)\chi_- = 0 \tag{4.11}$$

$$2\, D_c^\dagger \chi_- - (D_3 - \Phi + z)\chi_+ = 0 \tag{4.12}$$

in the NC monopole background:

$$\chi_\pm = \pm \Psi_\mp^\dagger \frac{1}{\Delta} \tag{4.13}$$

where

$$\begin{aligned}\Delta &= bb^\dagger + c^\dagger c\\ b^\dagger\Psi_+ + c\Psi_- &= 0\\ -c^\dagger\Psi_+ + b\Psi_- &= 0\end{aligned}\tag{4.14}$$

and the notations are from [18]. First of all, we need to invert Δ. This is easy, for (4.2) implies:

$$\begin{aligned}0 = \left(bb^\dagger + cc^\dagger\right)\Psi_+ = (\Delta+1)\Psi_+ \quad &\Rightarrow \quad \Delta\Psi_+ = -\Psi_+\\ 0 = \left(b^\dagger b + c^\dagger c\right)\Psi_- = (\Delta-1)\Psi_- \quad &\Rightarrow \quad \Delta\Psi_- = \Psi_-.\end{aligned}\tag{4.15}$$

However, it is early to assume that $\chi_\pm = \Psi_\mp^\dagger$ since Δ has a kernel:

$$\Delta f = 0 \Leftrightarrow f = vK$$

where v is from [18], $v = \sum_{n=0}^{\infty} \nu_n\, b^n\varphi\, |n\rangle\langle n|$, and K is an arbitrary x_3-dependent operator in the Fock space $\mathcal{H}$.

Thus, $\chi_\pm = \Psi_\mp^\dagger - K_\pm^\dagger v^\dagger$, and the operators K must be chosen in such a way, that $\chi_\pm\,(z=0) = 0$, which implies:

$$K_\pm^\dagger = \Psi_\mp^\dagger(z=0)\xi.\tag{4.16}$$

Recall that $\Psi_+ = cv, \quad \Psi_- = -b^\dagger v$. Hence[1]

$$K_+^\dagger = \xi^{-1}c^\dagger\xi, \quad K_-^\dagger = \xi^2.\tag{4.17}$$

5 Non-trivial backgrounds

So far our discussion concerned the flat closed string background. We shall now generalize the discussion to cover some non-trivial string backgrounds.

The obvious starting point is to consider orbifolds. So, let Γ be a discrete subgroup of $Spin(d) \times \mathbf{R}^d$ – the group of isometries of $\mathbf{R}^d$. For $g \in \Gamma$ let

[1] *Proof*: use the identities: $\xi v = \frac{b^n\varphi}{\zeta_n}, \frac{1}{2}\partial_3\,(\xi v) = \left(\partial_z + x_3 - \frac{\zeta_{n+1}}{\zeta_n}\right)(\xi v), \zeta_{n+1} = 2x_3\zeta_n + n\zeta_{n-1}$. Then: $\chi_+ = \left(\partial_z - x_3 - \frac{1}{2}z - \xi_n^2\right)v = \frac{1}{2}\xi^{-1}\left(\partial_3 - z\right)(\xi v) = \frac{1}{2}\left(\partial_3 + \Phi - z\right)v, \chi_- = vc^\dagger - \xi^{-1}c^\dagger\xi v = \xi^{-1}\partial_{\mathrm{c}}\,(\xi v) = \left(\partial_{\mathrm{c}} + A_{\mathrm{c}}\right)v$. From this (4.11) follows, with the help of the Bogomolny equations obeyed by A_{c}, Φ in [20]. It remains to check (4.12). Indeed, if we denote $\eta = \xi^2, \lambda_n(z) = \frac{1}{\zeta_n}b^n\varphi$, then: $\left(\partial_3 - z\right)\lambda_n = 2\eta_n\left(\lambda_{n-1} - \lambda_n\right), \quad \eta_n\left(\partial_3 + z\right)\frac{1}{\eta_n}\left(\partial_3 - z\right)\lambda_n = 4\eta_n\left[\eta_{n-1}\left(\lambda_{n-2} - \lambda_{n-1}\right) + \left(z - \eta_n\right)\left(\lambda_{n-1} - \lambda_n\right)\right]$ and at the same time: $\frac{1}{4}\eta\partial_{\mathrm{c}^\dagger}\left(\eta^{-1}\left(\partial_{\mathrm{c}}\lambda\right)\right) = \frac{\eta_n}{\eta_{n+1}}(n+1)\left(\lambda_{n+1} - \lambda_n\right) - n\left(\lambda_n - \lambda_{n-1}\right)$, which implies (4.12).

$\gamma(g) \in Spin(d)$ be the corresponding rotation matrix, and $l(g) \in \mathbf{R}^d$ the corresponding shift vector. We have the following composition rule:

$$\begin{aligned}(l(g_1), \gamma(g_1)) \times (l(g_2), \gamma(g_2)) &= (l(g_1) + \gamma(g_1)l(g_2), \gamma(g_1)\gamma(g_2)) \\ &= (l(g_1g_2), \gamma(g_1g_2)).\end{aligned} \tag{5.1}$$

We wish to consider D-branes living in the background obtained by taking the quotient of the Minkowskian space $\mathbf{R}^d$ by Γ.

The prescription for modifying the action (2.1) to reflect the orbifolded nature of the ambient space-time is the following. We demand that the Hilbert space $\mathcal{H}$ forms a representation of Γ. Let $\Omega\colon \Gamma \to \mathrm{End}(\mathcal{H})$ be the corresponding homomorphism. If $\mathcal{R}_i$, $i = 0, \ldots, r$, are irreducible unitary representations of Γ, $\mathcal{R}_0 = \mathbf{C}$ being the trivial representation, then

$$\mathcal{H} = \bigoplus_i \mathcal{H}_i \otimes \mathcal{R}_i$$

is the decomposition of $\mathcal{H}$. We demand that the operators $\mathbf{Y}$ are equivariant with respect to Γ in the sense that:

$$\Omega^{-1}(g)Y^i\Omega(g) = \gamma^i_j(g)Y^j + l^i(g). \tag{5.2}$$

In practice it is convenient to introduce the quiver diagram. We shall skip some standard issues like the reduced gauge invariance, and as a consequence more freedom in the action, coming from the twisted sector fields couplings. The latters are enumerated by the tensors

$$\theta^{ij}(g) = \gamma^i_{i'}(h)\gamma^j_{j'}(h)\theta^{i'j'}(h^{-1}gh) \tag{5.3}$$

which will alter the ground state solutions as follows:

$$[Y^i, Y^j] = i\frac{1}{\#\Gamma}\sum_{g\in\Gamma}\theta^{ij}(g)\Omega(g) \tag{5.4}$$

(for infinite Γ the normalization factor should be discussed separately).

5.1 *Example:* $\Gamma = \mathbf{Z}_2$

Let $\Gamma = \mathbf{Z}_2$ act on $\mathbf{R}^d$ as $y \mapsto -y$. We have:

$$\mathcal{H} = \mathcal{H}_+ \oplus \mathcal{H}_-. \tag{5.5}$$

The solution to (5.2) reads:

$$Y^i = \begin{pmatrix} 0 & Y^i_{+-} \\ Y^i_{-+} & 0 \end{pmatrix} \tag{5.6}$$

where $Y^i_{\pm\mp}$: $\mathcal{H}_\mp \to \mathcal{H}_\pm$. To simplify our problem let us assume very special form of the twisted sector field:

$$\theta^{ij}(-1) = \zeta\theta^{ij}(+1).$$

Then, by going to the complex notations we can rewrite (5.4) as follows:

$$[A_\alpha, A^\dagger_\beta] = \tfrac{1}{2}\delta_{\alpha\beta}(1+\zeta F), \qquad [A_\alpha, A_\beta] = 0, \alpha, \beta = 1, \ldots, d/2 \tag{5.7}$$

where F is the parity operator: $F|_{\mathcal{H}_\pm} = \pm 1$.

We shall now study the representation theory of the algebra (5.4). Clearly, it is sufficient to study the case $d = 2$. Let us rewrite (5.6) in terms of $A, A^\dagger$:

$$A_\alpha = \begin{pmatrix} 0 & b_\alpha \\ a_\alpha & 0 \end{pmatrix}, \tag{5.8}$$

then (5.7) becomes the condition (we now drop the index α):

$$aa^\dagger - b^\dagger b = \tfrac{1}{2}(1-\zeta), \quad bb^\dagger - a^\dagger a = \tfrac{1}{2}(1+\zeta). \tag{5.9}$$

We can assume $\zeta \geq 0$ (otherwise we exchange a and b). Notice that if $\zeta = 0$ we can take as a solution

$$b = \frac{1}{\sqrt{2}}(1-P)\mathbf{a}P, a = \frac{1}{\sqrt{2}}P\mathbf{a}(1-P), \qquad P = \tfrac{1}{2}(1+F), \quad F = (-1)^{\mathbf{a}^\dagger\mathbf{a}}$$

and $\mathbf{a}, \mathbf{a}^\dagger$ are the standard creation-annihilation operators.

For $0 \leq \zeta < 1$ we have the following (essentially unique) representation:

$$\begin{aligned} be^+_n &= \sqrt{n}e^-_{n-1}, & a^\dagger e^+_n &= \sqrt{n+\tfrac{1}{2}(1-\zeta)}e^-_n \\ b^\dagger e^-_n &= \sqrt{n+1}e^+_{n+1}, & ae^-_n &= \sqrt{n+\tfrac{1}{2}(1-\zeta)}e^+_n. \end{aligned} \tag{5.10}$$

For $\zeta \geq 1$

$$\begin{aligned} be^+_n &= \sqrt{n+\tfrac{1}{2}(\zeta-1)}e^-_{n-1}, & a^\dagger e^+_n &= \sqrt{n}e^-_n, \\ b^\dagger e^-_n &= \sqrt{n+\tfrac{1}{2}(\zeta+1)}e^+_{n+1} & ae^-_n &= \sqrt{n}e^+_n, \end{aligned} \tag{5.11}$$

and the vector e^+_0 should be dropped from the representation.

6 Conclusions and outlook

The gauge fields/strings duality is a fascinating long-standing problem [1]. We have considered a slightly generalized version of this duality, which includes noncommutative gauge fields in various dimensions, and, as a limit, the ordinary gauge theories. We have defined Yang-Mills algebras and considered several interesting examples of their representations. These arise as $\alpha' \to 0$ limits of the algebras of functions on the D-branes in flat space-time. We have also presented an analysis of the D-branes in curved backgrounds, namely in those, obtained by orbifolding from the flat space-time.

The topics left outside of this short note include the generalizations to the non-trivial H-fields (in which case any universal enveloping algebra of a simple Lie algebra may arise as an example of Yang-Mills algebra, at least if the dimension of the latter permits), time-dependent backgrounds (with applications to cosmology), explicit examples of instantons on the curved spaces, and, most interestingly, the consequences of the decoupling of the null-vectors on the closed string side for the open string gauge invariant quantities. We plan to present some of these considerations elsewhere.

References

[1] A. Polyakov [hep-th/0110196].

[2] A. Connes, *Noncommutative geometry* (Academic Press, 1994).

[3] A. Connes, M. Douglas and A. Schwarz, *JHEP* **9802** (1998) 003.

[4] M. Douglas and C. Hull, *JHEP* **9802** (1998) 008 [hep-th/9711165].

[5] V. Schomerus, *JHEP* **9906** (1999) 030.

[6] N. Seiberg and E. Witten, *JHEP* **9909** (1999) 032 [hep-th/9908142].

[7] E. Witten, *Nucl. Phys. B* **268** (1986) 253.

[8] M. Douglas and N. Nekrasov [hep-th/0106048].

[9] A. Alekseev and V. Schomerus [hep-th/9812193].

[10] A. Recknagel and V. Schomerus [hep-th/9712186]; G. Felder, J. Frölich, J. Fuchs and C. Schweigert [hep-th/9909140, hep-th/9912239].

[11] N. Nekrasov and A.S. Schwarz, *Comm. Math. Phys.* **198** (1998) 689 [hep-th/9802068].

[12] E. Corrigan and P. Goddard, *Ann. Phys.* **154** (1984) 253.

[13] S. Shatashvili [hep-th/0105076].

[14] N. Seiberg [hep-th/0008013].

[15] A. Polychronakos [hep-th/0007043].

[16] N. Ishibashi, H. Kawai, Y. Kitazawa and A. Tsuchiya, *Nucl. Phys. B* **498** (1997) 467 [hep-th/9612115].

[17] H. Ooguri and Y. Okawa [hep-th/0103124].

[18] D. Gross and N. Nekrasov [hep-th/9907204].

[19] N. Nekrasov [hep-th/0010017].

[20] D. Gross and N. Nekrasov, *JHEP* **0007** (2000) 034 [hep-th/0005204].

LECTURE 9

CONDENSATES NEAR THE ARGYRES-DOUGLAS POINT IN $SU(2)$ GAUGE THEORY WITH BROKEN $\mathcal{N} = 2$ SUPERSYMMETRY

A. GORSKY

Institute of Experimental and Theoretical Physics, Moscow 117259, Russia

Contents

CONDENSATES NEAR THE ARGYRES-DOUGLAS POINT IN $SU(2)$ GAUGE THEORY WITH BROKEN $\mathcal{N} = 2$ SUPERSYMMETRY

A. Gorsky

Abstract

The behaviour of the chiral condensates in the $SU(2)$ gauge theory with broken $\mathcal{N} = 2$ supersymmetry is reviewed. The calculation of monopole, dyon, and charge condensates is described. It is shown that the monopole and charge condensates vanish at the Argyres-Douglas point where the monopole and charge vacua collide. This phenomenon is interpreted as a deconfinement of electric and magnetic charges at the Argyres-Douglas point.

1 Introduction

This talk is based on the paper [1] where the behaviour of the supersymmetric gauge theories near the Argyres-Douglas point was considered. The main question discussed concerned the nature of the hypothetical phase transition occured at the Argyres-Douglas point. It is widely believed that the theory near this point flows into the superconformal point in the infrared however the physics of this critical point was unclear. Since the results presented below are based on the exact statements concerning $\mathcal{N} = 1$ and $\mathcal{N} = 2$ supersymmetric theories the identification of the phase transition at the Argyres-Douglas point as a deconfainment phase transition is rigorous.

The derivation of exact results in $\mathcal{N} = 1$ supersymmetric gauge theories based on low energy effective superpotentials and holomorphy was pioneered in [2, 3] and then strongly developed, mostly by Seiberg, see [4] for review. An extra input was provided by the Seiberg-Witten solution of $\mathcal{N} = 2$ supersymmetric gauge theories with and without matter [5]. It was also clarified that Seiberg-Witten solution amounts the set of vacua in the corresponding $\mathcal{N} = 1$ theory [5, 7–10]. Different vacua are distinguished by values of chiral condensates, such as gluino condensate $\langle \mathrm{Tr}\, \lambda\lambda \rangle$ and the condensate of the fundamental matter $\langle \tilde{Q} Q \rangle$. Recently some points concerning the formation of the condensate and the identification of the relevant field configurations were clarified in [12–15].

We compare then the condensate of the adjoint matter with the discriminant locus defined by Seiberg-Witten solution in $\mathcal{N} = 2$ theory and find a complete matching. Our results for matter and gaugino condensates are consistent with those obtained by "integrating in" method [8, 16, 17] and can be viewed as an independent confirmation of the method. What is specific for our approach is that we start from weak coupling regime where notion of effective Lagrangian is well defined and then use holomorphy to extend results for chiral condensates into strong coupling.

Then we determine monopole, dyon and charge condensates following to the Seiberg-Witten approach, *i.e.* considering effective superpotentials near singularities on the Coulomb branch in $\mathcal{N} = 2$ theory. Again, holomorphicity allows us to extend results to the domain of strong $\mathcal{N} = 2$ breaking.

Our next step is study of chiral condensates in the Argyres-Douglas (AD) points. These points were originally introduced in moduli/parameter space of $\mathcal{N} = 2$ theories as points where two singularities on the Coulomb branch collide [18–20]. It is believed that the theory at the AD point flows in infrared to a nontrivial superconformal theory. The notion of AD point continue to make sense even when the $\mathcal{N} = 2$ theory is broken to $\mathcal{N} = 1$ by nonzero μ, in the $\mathcal{N} = 1$ theory it is the point in parameter space where two vacua collide.

Particularly, we consider collision of monopole and charge vacua at certain value of the mass of the fundamental flavor. Our key result is that both monopole and charge condensates vanish at the AD point. We interpret this as deconfinement of both electric and magnetic charges at the AD point.

Let us remind that the condensation of monopoles ensures confinement of quarks in the monopole vacuum [5], while the condensation of charges provides confinement of monopoles in the charge vacuum. As it was shown by 't Hooft [21] it is impossible for these two phenomena to coexist. This leads to a paradoxical situation in the AD point where the monopole and charge vacua collide. Our result resolves this paradox.

This paradox is a part of more general problem: whether there is a uniquely defined theory in the AD point. Indeed, when two vacua collide the Witten index of the emerging theory is 2, *i.e.* there are two bosonic vacuum states. The question is if there is any physical quantity which could serve as an order parameter differentiating these two vacua. The continuity of chiral condensates in the AD point we found shows that these condensates are not playing this role. The same continuity leads also to vanishing of tension of domain walls interpolating between colliding vacua when we approach the AD point.

2 Matter and gaugino condensates

Let us consider $\mathcal{N}=1$ theory with $SU(2)$ gauge group where the matter sector consists of the adjoint field $\Phi^\alpha_\beta = \Phi^a(\tau^a/2)^\alpha_\beta$ ($\alpha,\beta=1,2;\ a=1,2,3$) and two fundamental fields Q^α_f ($f=1,2$) describing one flavor. The most general renormalizable superpotential for this theory has the form,

$$\mathcal{W} = \mu\,\mathrm{Tr}\,\Phi^2 + \frac{m}{2}\,Q^\alpha_f Q^f_\alpha + \frac{1}{\sqrt{2}}\,h^{fg}\,Q_{\alpha f}\Phi^\alpha_\beta Q^\beta_g. \tag{2.1}$$

Here parameters μ and m are related to masses of the adjoint and fundamental fields, $m_\Phi = \mu/Z_\Phi$, $m_Q = m/Z_Q$, by corresponding Z factors in kinetic terms. Having in mind normalization to the $\mathcal{N}=2$ case we choose for bare parameters $Z^0_\Phi = 1/g_0^2$, $Z^0_Q = 1$. The matrix of Yukawa couplings h^{fg} is the symmetric, summation over color indices $\alpha,\beta=1,2$ is explicit. Unbroken $\mathcal{N}=2$ SUSY appears when $\mu=0$ and $\det h = -1$.

To get an effective theory similar to SQCD we integrate out the adjoint field Φ implying that $m_\Phi \gg m_Q$. In classical approximation this integration reduces to to the substitution

$$\Phi^\alpha_\beta = -\frac{1}{2\sqrt{2}\,\mu}\,h^{fg}\left(Q_{\beta f}Q^\alpha_g - \frac{1}{2}\delta^\alpha_\beta Q_{\gamma f}Q^\gamma_g\right), \tag{2.2}$$

which follows from $\partial\mathcal{W}/\partial\Phi = 0$. It is well known from the study of SQCD that perturbative loops do not contribute and nonperturbative effects are exhausted by the Affleck-Dine-Seiberg (ADS) superpotential generated by one instanton [2]. The effective superpotential then is

$$\mathcal{W}_{\rm eff} = m\,V - \frac{(-\det h)}{4\mu}\,V^2 + \frac{\mu^2\Lambda_1^3}{4\,V} \tag{2.3}$$

where the gauge and subflavor invariant chiral field V is defined as

$$V = \frac{1}{2}\,Q^\alpha_f Q^f_\alpha. \tag{2.4}$$

The third nonperturbative term in equation (2.3) is the ADS superpotential. The coefficient $\mu^2\Lambda_1^3/4$ in the ADS superpotential is an equivalent of $\Lambda^5_{\rm SQCD}$ in SQCD. The factor μ^2 in the coefficient reflects four zero modes of the adjoint field, see *e.g.* references [14, 22] for details.

When $\det h$ is nonvanishing we have three vacua, marked by vevs of the lowest component of V,

$$v = \langle\, V\,\rangle\,\cdot \tag{2.5}$$

These vevs are roots of the algebraic equation $\mathrm{d}\mathcal{W}_{\rm eff}/\mathrm{d}v = 0$ which looks as

$$m - \frac{(-\det h)}{2}\,\frac{v}{\mu} - \frac{\Lambda_1^3}{4}\left(\frac{\mu}{v}\right)^2 = 0. \tag{2.6}$$

This equation shows, in particular, that although the second term in the superpotential (2.3) looks as suppressed at large μ it is of the same order as the ADS term. From equation (2.6) it is also clear that the dependence on μ is given by scaling $v \propto \mu$.

To see dependence on other parameters let us substitute v by the dimensionless variable κ as

$$v = \mu\sqrt{\frac{\Lambda_1^3}{4m}}\,\kappa. \tag{2.7}$$

Then equation (2.6) in terms of κ

$$1 - \sigma\,\kappa - \frac{1}{\kappa^2} = 0 \tag{2.8}$$

is governed by the dimensionless parameter σ,

$$\sigma = \frac{(-\det h)}{4}\left(\frac{\Lambda_1}{m}\right)^{3/2}\cdot \tag{2.9}$$

To verify this interesting mapping we need to find out vevs for

$$u = \langle U\rangle = \langle \mathrm{Tr}\,\Phi^2\rangle\cdot \tag{2.10}$$

This can be done using set of Konishi anomalies. Generic equation for arbitrary matter field Q looks as follows (we are using notations of the review [11]):

$$\frac{1}{4}\,\bar{D}^2 J_Q = Q\,\frac{\partial\mathcal{W}}{\partial Q} + T(R)\,\frac{\mathrm{Tr}\,W^2}{8\pi^2}, \tag{2.11}$$

where $T(R)$ is the Casimir in the matter representation. The left hand side is the total derivative in superspace so its average over supersymmetric vacuum vanishes. In our case it results in two relations for condensates,

$$\begin{aligned}
&\left\langle \frac{m}{2}\,Q_f^\alpha Q_\alpha^f + \frac{1}{\sqrt{2}}\,h^{fg}\,Q_{\alpha f}\Phi^\alpha_\beta Q^\beta_g + \frac{1}{2}\,\frac{\mathrm{Tr}\,W^2}{8\pi^2}\right\rangle = 0\\
&\left\langle 2\,\mu\mathrm{Tr}\,\Phi^2 + \frac{1}{\sqrt{2}}\,h^{fg}\,Q_{\alpha f}\Phi^\alpha_\beta Q^\beta_g + 2\,\frac{\mathrm{Tr}\,W^2}{8\pi^2}\right\rangle = 0.
\end{aligned} \tag{2.12}$$

From the first relation after substitution (2.2) and comparison with equation (2.6) we find the expression for gluino condensate s

$$s = \frac{\langle \mathrm{Tr}\, \lambda^2 \rangle}{16\pi^2} = -\frac{\langle \mathrm{Tr}\, W^2 \rangle}{16\pi^2} = \frac{\mu^2 \Lambda_1^3}{4\, v} \cdot \tag{2.13}$$

This is consistent with the general expression $[T_{\mathrm{G}} - \sum T(R)]\langle \mathrm{Tr}\lambda^2 \rangle / 16\pi^2$ for the nonperturbative ADS piece of the superpotential (2.3) [24]. Combining then two relations (2.12) we express the condensate value of u *via* v,

$$u = \frac{1}{2\mu}\,(m\, v + 3\, s) = \frac{1}{2\mu}\left(m\, v + \frac{3}{4}\,\frac{\mu^2 \Lambda_1^3}{v}\right) = \frac{\sqrt{m\Lambda_1^3}}{4}\left(\kappa + \frac{3}{\kappa}\right) \cdot \tag{2.14}$$

Now we see that at the limit of large m two vacua $\kappa = \pm 1$ are in perfect correspondence with $u = \pm\, \Lambda_0^2$ for the monopole and dyon vacua of SYM. Indeed, $\Lambda_0^4 = m\Lambda_1^3$ is a correct relation between scale parameters of the theories.

For the third vacuum at large m the value $u = m^2/(-\det h)$ corresponds on the Coulomb branch to the so called charge vacuum, where some fundamental fields become massless. Moreover, the correspondence with $\mathcal{N} = 2$ results can be demonstrated for three vacua at any value of m. To this end we use the relation (2.14) and equation (2.8) to derive the following equation for u,

$$(-\det h)\, u^3 - m^2\, u^2 - \frac{9}{8}\,(-\det h)\, m\Lambda_1^3\, u + m^3 \Lambda_1^3 + \frac{27}{2^8}\,(-\det h)^2 \Lambda_1^3 = 0. \tag{2.15}$$

Three roots of this equation are vevs of Tr Φ^2 in the corresponding vacua.

How does it look from $\mathcal{N} = 2$ side? The Riemann surface governing the Seiberg-Witten solution is given by the curve [5]

$$y^2 = x^3 - u\, x^2 + \frac{1}{4}\Lambda_1^3 m\, x - \frac{1}{64}\Lambda_1^6. \tag{2.16}$$

Singularities of the metric, *i.e.* the discriminant locus of the curve, is defined by two equations, $y^2 = 0$ and $\mathrm{d}y^2/\mathrm{d}x = 0$,

$$x^3 - u\, x^2 + \frac{1}{4}\Lambda_1^3 m\, x - \frac{1}{64}\Lambda_1^6 = 0, \quad 3x^2 - 2u\, x + \frac{1}{4}\Lambda_1^3 m = 0, \tag{2.17}$$

which lead to

$$u^3 - m^2\, u^2 - \frac{9}{8}\, m\Lambda_1^3\, u + m^3 \Lambda_1^3 + \frac{27}{2^8}\, \Lambda_1^3 = 0. \tag{2.18}$$

We see that this is a particular case of the $\mathcal{N} = 1$ equation (2.15) at $\det h = -1$.

The point in the parameter manifold where two vacua coincide is the AD point [18]. In $SU(2)$ theory these points were studied in [19]. Mutually non-local states, say charges and monopoles becomes massless at these points. On the Coulomb branch of $\mathcal{N}=2$ theory these points correspond to non-trivial conformal field theory [19]. Here we study the $\mathcal{N}=1$ SUSY theory, where $\mathcal{N}=2$ is broken down by the mass term for the adjoint matter as well as by the difference of the Yukawa coupling from its $\mathcal{N}=2$ value. But collisions of two vacua still occur in the theory. In this subsection we find the values of m at which AD points appear and calculate values of condensates at this point. In the next section we study what happen to the confinement of charges in the monopole point at non-zero μ once we approach AD point.

First let us work out the AD values of m, generalizing the consideration [19]. Collision of two roots for v means that together with equation (2.6) the derivative of its left-hand-side should also vanish,

$$m-\frac{(-\det h)}{2}\,\frac{v}{\mu}-\frac{\Lambda_1^3}{4}\left(\frac{\mu}{v}\right)^2=0,\qquad -(-\det h)+\Lambda_1^3\left(\frac{\mu}{v}\right)^3=0. \quad (2.19)$$

This system is consistent only at three values of $m=m_{\mathrm{AD}}$,

$$m_{\mathrm{AD}}=\frac{3}{4}\,\omega\,\Lambda_1\,(-\det h)^{2/3},\qquad \omega=\mathrm{e}^{2\pi i n/3}\quad (n=0,\pm 1), \qquad (2.20)$$

related by Z_3 symmetry. The condensates at the AD vacuum are

$$\begin{aligned} v_{\mathrm{AD}}&=\omega\,\frac{\mu\,\Lambda_1}{(-\det h)^{1/3}},\\ u_{\mathrm{AD}}&=\omega^{-1}\,\frac{3}{4}\,\Lambda_1^2\,(-\det h)^{1/3},\\ s_{\mathrm{AD}}&=\omega^{-1}\,\frac{1}{4}\,\mu\Lambda_1^2\,(-\det h)^{1/3}. \end{aligned} \qquad (2.21)$$

3 Dyon condensates

In this section we calculate various dyon condensates at three vacua of the theory. As it was discussed above holomorphicity allows us to find these condensates starting from consideration on the Coulomb branch in $\mathcal{N}=2$ near the singularities associated with given massless dyon. Namely, we calculate the monopole condensate near the monopole point, the charge condensate near the charge point and the dyon $(n_{\mathrm{m}},n_{\mathrm{e}})=(1,1)$ condensate near the point where this dyon is light. Although we start with small value of adjoint mass parameter μ, our results for condensates are exact for any μ.

3.1 Monopole condensate

Let us start with calculation of the monopole condensate near the monopole point. Near this point the effective low energy description of our theory can be given in terms of $\mathcal{N}=2$ dual QED [5]. It includes light monopole hypermultiplet interacting with vector (dual) photon multiplet in the same way as electric charges interact with ordinary photons. Following Seiberg and Witten [5] we write down the effective superpotential in the following form.

$$W=\sqrt{2}\,\tilde{M}MA_{\mathrm{D}}+\mu U, \tag{3.1}$$

where A_{D} is a chiral neutral field (it is a part of $\mathcal{N}=2$ dual photon multiplet in $\mathcal{N}=2$ theory) and $U=\mathrm{Tr}\,\Phi^2$. The second term breaks $\mathcal{N}=2$ supersymmetry down to $\mathcal{N}=1$.

Variating this superpotential with respect to A_{D}, M and $\tilde{M}$ we find that $A_{\mathrm{D}}=0$, *i.e.* the monopole mass vanishes, and

$$\langle\tilde{M}M\rangle=-\frac{\mu}{\sqrt{2}}\frac{\mathrm{d}u}{\mathrm{d}a_{\mathrm{D}}}\cdot \tag{3.2}$$

The condition $A_{\mathrm{D}}=0$ means that the Coulomb branch near the monopole point, where the monopole mass vanishes, shrinks to the single vacuum state at the singularity while equation (3.2) together with D flatness condition (up to gauge transformation) $\tilde{M}=M$ determines the value of monopole condensate.

The non-zero value of monopole condensate ensures the $U(1)$ confinement for charges *via* the formation of Abrikosov-Nielsen-Olesen vortices. Let us work out the r.h.s. of equation (3.2) to determine the μ and m dependence of the monopole condensate. From exact Seiberg-Witten solution [5] we have

$$\frac{\mathrm{d}a_{\mathrm{D}}}{\mathrm{d}u}=\frac{\sqrt{2}}{8\pi}\oint_\gamma\frac{\mathrm{d}x}{y(x)}\cdot \tag{3.3}$$

Here for $y(x)$ given by equation (2.16) we use the form

$$y^2=(x-e_0)(x-e_-)(x-e_+). \tag{3.4}$$

We get finally

$$\langle\tilde{M}M\rangle=2i\mu\left(u_{\mathrm{M}}^2-\frac{3}{4}m\Lambda_1^3\right)^{1/4}. \tag{3.5}$$

Now let us address the question: what happens with the monopole condensate when we reduce m and approach the AD point. The AD point corresponds to particular value of m which ensures colliding of monopole and charge singularities in the u plane. Near the monopole point we have condensation of monopoles and confinement of charges while near the charge point we have condensation of charges and confinement of monopoles. As it was shown by 't Hooft these two phenomena cannot happen simultaneously [21]. The question is: what happen when monopole and charge points collide in the u plane?

The monopole condensate at the AD point is given by equation (3.5) when $m_{\rm AD}$ and $u_{\rm AD}$ from equations (2.20) and (2.21) are substituted,

$$\langle \tilde{M} M \rangle_{\rm AD} = 0. \tag{3.6}$$

We see that monopole condensate goes to zero at the AD point. Our derivation above makes clear why it happens. At the AD point all three roots of y^2 become degenerate, $e_+ = e_- = e_0$, so the monopole condensate which is proportional to $\sqrt{e - e_0}$ naturally vanishes.

In the next subsection we calculate the charge condensate in the charge point and show that it is also goes to zero as m approaches its AD value (2.20). Thus we interpret the AD point as a deconfinement point for both monopoles and charges.

3.2 Charge and dyon condensates

In this subsection we use the same method to calculate values of charge and dyon condensate near charge and dyon points respectively. We first consider m above AD value (2.20) and then continue our results to values of m below $m_{\rm AD}$. In particular in the limit $m = 0$ we recover Z_3 symmetry.

Let us start with the charge condensate. At $\mu = 0$, $\det h = -1$ and large m the effective theory near the charge point

$$a = -\sqrt{2}\, m \tag{3.7}$$

on the Coulomb branch is $\mathcal{N} = 2$ QED. The half of degrees of freedom in color doublets becomes massless whereas the other half acquire large mass $2m$. These massless fields form one hypermultiplet $\tilde{Q}_+, Q_+$ of charge particle in the effective electrodynamics. Once we add the mass term for the adjoint matter the effective superpotential near the charge point becomes

$$\mathcal{W} = \frac{1}{\sqrt{2}}\, \tilde{Q_+} Q_+ A + m\, \tilde{Q_+} Q_+ + \mu\, U. \tag{3.8}$$

Minimizing this superpotential we get condition (3.7) as well as

$$\langle \tilde{Q_+} Q_+ \rangle = -\sqrt{2}\, \mu\, \frac{\mathrm{d}u}{\mathrm{d}a}\,\cdot \tag{3.9}$$

Now following the same steps which led us from (3.2) to (3.5) we get

$$\langle \tilde{Q_+} Q_+ \rangle = 2\,\mu\,(u_{\rm C}^2 - \frac{3}{4} m\,\Lambda_1^3)^{1/4}. \tag{3.10}$$

Here $u_{\rm C}$ is the position of charge point in the u plane, $u_{\rm C} = m^2$ at large m. Thus, at large m

$$\langle \tilde{Q_+} Q_+ \rangle = 2\,\mu m. \tag{3.11}$$

Holomorphicity allows us to extend the result (3.10) to arbitrary m and $\det h$. So we can use equation (3.10) to find the charge condensate at the AD point. Using equations (2.20) and (2.21) we see that the charge condensates vanishes in the AD point the same way the monopole one does. As it was mentioned we interpret this as deconfinement for both charges and monopoles.

Similarly to the monopole condensate we can relate the charge condensate with the quark one v,

$$\langle \tilde{Q}_+ Q_+ \rangle^2 = v^2 - \frac{\mu^3 \Lambda_1^3}{v} = v^2 - 4\;\mu s. \tag{3.12}$$

This expression differs from the one for the monopole condensate only by sign. The coincidence of the charge condensate with the quark one at large v, *i.e.* at weak coupling is natural. The difference is due to nonperturbative effects and similar to the difference between $a^2/2$ and u on the Coulomb branch of the $\mathcal{N} = 2$ theory. In strong coupling the difference is not small, in particular, the charge condensate vanishes in the AD point while the quark condensate remains finite.

Note that near the AD point we can consider an effective superpotential which includes both light monopole and charge fields simultaneously. Such consideration leads to the same results for condensates.

Now let us work out the dyon condensate. More generally let us introduce the dyon field D_i , $i = 1, 2, 3$, which stands for charge, monopole and $(1, 1)$ dyon, $D_i = (Q_+, M, D)$. The arguments of the previous subsection which led us to the result (3.5) for monopole condensate gives for $\langle \tilde{D}_i\, D_i \rangle$

$$\langle \tilde{D}_i\, D_i \rangle = 2\, i\, \zeta_i\, \mu \left(u_i^2 - \frac{3}{4}\, m\, \Lambda_1^3 \right)^{1/4}, \tag{3.13}$$

where u_i is the position of the i-th point in the u plane and ζ_i are phase factors.

For the monopole condensate at real values of m larger than $m_{\rm AD} = (3/4)\Lambda_1(-\det h)^{2/3}$ equation (3.5) gives

$$\zeta_{\rm M} = 1, \tag{3.14}$$

while for charge from equation (3.10)

$$\zeta_{\mathrm{C}} = -i. \tag{3.15}$$

In fact one can fix the phase factor for charge imposing the condition that the charge condensate should approach the value 2 $m\mu$ in the large m limit. For dyon the phase factor is

$$\zeta_{\mathrm{D}} = i. \tag{3.16}$$

At the AD point monopole and charge condensates go to zero, while the dyon one remains non-zero, see (3.13). Below the AD point condensates are given by the same equation (3.13), but the phase factors for charge and monopole can change its values[1]. The dyon phase factor (3.16) is not changing when we move through the AD point because the dyon condensate does not vanish at this point.

4 The Argyres-Douglas point: How well the theory is defined

As we discussed in Introduction in the AD point we encounter the problem of not uniquely defined vacuum state. Indeed, when the mass parameter m approaches its AD value m_{AD} we deal with two vacuum states which can be distinguished by values of chiral condensates. It is unlikely that the number of states with zero energy will change when we reach the AD point, it is very much similar to Witten index. However, the continuity of chiral condensates we obtained above shows that they are no longer parameters which differentiate two states once we reach the AD point.

A natural possibility to consider is domain walls interpolating between colliding vacua. In case of BPS domain walls their tension is given by central charges,

$$T_{ab} = 2\,|\mathcal{W}_{\mathrm{eff}}(v_a) - \mathcal{W}_{\mathrm{eff}}(v_b)| \tag{4.1}$$

where a, b label colliding vacua. The central charge here is expressed *via* values of exact superpotential (2.3) in corresponding vacua. The continuity of the condensate v shows that the domain wall becomes tensionless in the AD point. If such domain wall were observable it could serve as a signal of two vacua.

Let us note one more interesting question. Namely the BPS tension should obey the Picard-Fuchs equation providing the dependence on the quark mass. The mass corresponding to the position of the Argyres-Douglas

[1]Note that quantum numbers of "charge" and "monopole" are also changed, see [25].

point plays the role of the "strong coupling singularity" like the monopole singularity in the Seiberg-Witten solution. The generic structure of the monodromies at the complex m plane should provide the unique solution for the multiplet of tensions. It would be very interesting to compare the solutions of Picard-Fuchs equations with the tensions followed from the exact superpotentials.

Recently a method of a calculation of the values of the exact superpotential at minima has been suggested [27] using the auxiliary Riemann surfaces. It matches by a kind of duality the calculation *via* the degenerate curves and effective description *via* the effective composite fields. It seems that the behaviour near the Argyres-Douglas point as well as the Konishi type relations provides the additional tests of the duality proposed.

5 Conclusions

We analyze monopole, charge and dyon condensates departing from the Coulomb branch of the $\mathcal{N}=2$ theory. It results in the explicit relations between these condensates and those of the fundamental matter. The most interesting phenomenon occurs in the AD point: when the monopole and charge vacua collide both the monopole and charge condensates vanish. We interpret this as a deconfinement of electric and magnetic charges in the AD point.

Let us mention a relation to finite-dimensional integrable systems. It was recognized that $\mathcal{N}=2$ theories are governed by finite-dimensional integrable systems. The integrable system responsible for $\mathcal{N}=2$ SQCD was identified with the nonhomogenious XXX spin chain [28]. After perturbation to the $\mathcal{N}=1$ theory the Hamiltonian of the integrable system is expected to coincide with the superpotential of corresponding $\mathcal{N}=1$ theory. This has been confirmed by direct calculation in the pure $\mathcal{N}=2$ gauge theory [29] as well in the theory with massive adjoint multiplet [30, 31]. It would be very interesting to find a similar connection between spin chain Hamiltonians and superpotentials in the $\mathcal{N}=1$ SQCD. One more point to be clarified is a meaning of the AD point within approach based on integrability. Since the quark mass is identified as a value of spin [28] one could expect that at particular values of spins corresponding to the AD mass XXX spin chain has additional symmetries similar to superconformal ones.

I am grateful to A. Vainshtein and A. Yung for the collaboration on this issue. The work was partially supported by the grants INTAS-00-00334 and CRDF-RP1-2108.

References

[1] A. Gorsky, A. Vainshtein and A. Yung, *Nucl. Phys. B* **584** (2000) 197 [hepth/0004087].

[2] I. Affleck, M. Dine and N. Seiberg, *Phys. Lett. B* **137** (1984) 187.

[3] V.A. Novikov, M.A. Shifman, A.I. Vainshtein, and V.I. Zakharov, *Nucl. Phys. B* **229** (1983) 407 [Reprinted in *Supersymmetry*, edited by S. Ferrara (North Holland/World Scientific, Amsterdam – Singapore, 1987), Vol. 1, p. 606]; M. Shifman and A. Vainshtein, *Nucl. Phys. B* **296** (1988) 445.

[4] K. Intriligator and N. Seiberg, *Nucl. Phys. Proc. Suppl. BC* **45** (1996) 1 [hep-th/9509066].

[5] N. Seiberg and E. Witten, *Nucl. Phys. B* **426** (1994) 19; (E) B **430** (1994) 485 [hep-th/9407087]; B **431** (1994) 484 [hep-th/9408099].

[6] D. Kutasov, A. Schwimmer and N. Seiberg, *Nucl. Phys. B* **459** (1996) 455.

[7] K. Intrilligator and N. Seiberg, *Nucl. Phys. B* **431** (1994) 551 [hep-th/9408155].

[8] S. Elitzur, A. Forge, A. Giveon, and E. Rabinovici, *Phys. Lett. B* **353** (1995) 79 [hep-th/9504080]; *Nucl. Phys. B* **459** (1996) 160 [hep-th/9509130].
S. Elitzur, A. Forge, A. Giveon, K. Intrilligator, and E. Rabinovici, *Phys. Lett. B* **379** (1996) 121 [hep-th/9603051].

[9] S. Terashima and S. Yang, *Phys. Lett. B* **391** (1997) 107 [hep-th/9607151].

[10] K. Konishi and H. Terao, *Nucl. Phys. B* **511** (1998) 264 [hep-th/9707005].

[11] M. Shifman and A. Vainshtein, In M.A. Shifman: *ITEP lectures on particle physics and field theory,* Vol. 2, p. 485 (World Scientific, Singapore, 1999) [hep-th/9902018].

[12] N.M. Davies, T.J. Hollowood, V.V. Khoze and M.P. Mattis, *Nucl. Phys. B* **559** (1999) 123 [hep-th/9905015].

[13] N.M. Davies and V.V. Khoze, *JHEP* **0001** (2000) 015 [hep-th/9911112].

[14] A. Ritz and A. Vainshtein, *Nucl. Phys. B* **566** (2000) 311 [hep-th/9909073].

[15] G. Carlino, K. Konishi and H. Murayama, *JHEP* **0002** (2000) 004 [hep-th/0001036].

[16] K. Intriligator, R.G. Leigh and N. Seiberg, *Phys. Rev. D* **50** (1994) 1092 [hep-th/9403198].

[17] K. Intriligator, *Phys. Lett. B* **336** (1994) 409 [hep-th/9407106].

[18] P.C. Argyres and M.R. Douglas, *Nucl. Phys. B* **448** (1995) 93 [hep-th/950506].

[19] P.C. Argyres, M.R. Plesser, N. Seiberg and E. Witten, *Nucl. Phys. B* **461** (1996) 71 [hep-th/9511154].

[20] T. Eguchi, K. Hori, K. Ito and S. Yang, *Nucl. Phys. B* **471** (1996) 430 [hep-th/9603002].

[21] G.'t Hooft, *Nucl. Phys. B* **138** (1978) 1; *Nucl. Phys. B* **153** (1979) 141.

[22] A. Yung, *Nucl. Phys. B* **485** (1997) 38 [hep-th/9604096].

[23] N. Seiberg, *Phys. Rev. D* **49** (1994) 6857 [hep-th/9402044].

[24] V.A. Novikov, M.A Shifman, A.I. Vainshtein and V.I. Zakharov, *Nucl. Phys. B* **260** (1985) 157.

[25] A. Bilal and F. Ferrari, *Nucl. Phys. B* **516** (1998) 175 [hep-th/9706145].

[26] A. Hanany, M. Strassler and A. Zaffaroni, *Nucl. Phys. B* **513** (1998) 87 [hep-th/9707244].

[27] F. Cachazo, K.A. Intriligator and C. Vafa, *Nucl. Phys. B* **603** (2001) 3 [hep-th/0103067].

[28] A. Gorsky, A. Marshakov, A. Mironov and A. Morozov, *Phys. Lett. B* **380** (1996) 75 [hep-th/9603140].
A. Gorsky, S. Gukov and A. Mironov, *Nucl. Phys. B* **517** (1998) 409 [hep-th/9707120].
A. Gorsky and A. Mironov, *Nucl. Phys. B* **550** (1999) 513 [hep-th/9902030].

[29] S. Katz and C. Vafa, *Nucl. Phys. B* **497** (1997) 196 [hep-th/9611090].

[30] N. Dorey, *JHEP* **9907** (1999) 021 [hep-th/9906011].

[31] S.P. Kumar and J. Troost, *JHEP* **0201** (2002) 020 [hep-th/0112109].

SEMINAR 1

QUANTUM FIELD THEORY WITH EXTRA DIMENSIONS

L. BAULIEU

LPTHE,
Université Pierre et Marie Curie,
4 place Jussieu,
75252 Paris Cedex 05, France

Contents

QUANTUM FIELD THEORY WITH EXTRA DIMENSIONS

L. Baulieu

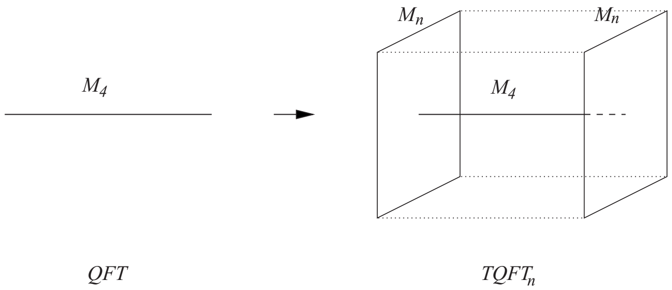
Abstract

We explain that a bulk with arbitrary dimensions can be added to the space over which a quantum field theory is defined. This gives a TQFT such that its correlation functions in a slice are the same as those of the original quantum field theory. This generalizes the stochastic quantization scheme, where the bulk is one dimensional.

1 Introduction

In recent works [1, 2], the basic ideas of stochastic quantization [3, 4] have been elaborated in a systematic approach, called bulk quantization. These papers give a central role to the introduction of a symmetry of a topological character. The correlation functions for equal values of the bulk time define the correlations of the physical theory. Pertubatively, bulk quantization and the usual quantization method are equivalent because the observables satisfy the same Schwinger-Dyson equations in both approaches; basic concepts such as the definition of the S-matrix (in the LSZ sense) and the Cutkowski rules can be also directly addressed in 5 dimensions [1]. We believe that the difficulty of giving a consistent stochastic interpretation to all details of the formalism, especially in the case of gauge theories, justifies to directly postulate that bulk quantization is a particular type of a topological field theory. Moreover, there is an interesting geometrical interpretation for many of the ingredients that are needed in bulk quantization. The idea of a topological field theory is actually relevant, since one wishes to define observable that are independent of most of the details of the bulk, such as the metrics components g_{tt} and $g_{\mu t}$. Quite interestingly, the interpretation of anomalies in gauge theories is that, the limit of taking the limit of an equal bulk time can be ambiguous.

As shown in [1], the additional dimension t does not take part in the Poincaré group of symmetries for the case of an additional non compact



dimension. The homogeneity of the Lagrangian requires that t has the dimension of the square of the ordinary coordinates. We investigated how this method can be applied to gravity and supersymmetric theories in [5,6]. Here, we will show that bulk quantization can be generalized to the case where the bulk has an arbitrary dimension $n \geq 1$, according to the following picture: The interest of this generalization, when n is larger than one, is yet to be discovered, but we find that the aesthetics of the whole construction makes it worth being presented.

2 The fields

We consider for simplicity the case of a commuting scalar field $\phi(x)$ in 4 dimensions, with a Lagrangian $L_0(\phi)$ and action $I_0[\phi] = \int \mathrm{d}^4 x L_0(\phi)$. We want to find another formulation of this theory with an action in a space with dimensions $4+n$, where n denotes generically the dimension of the bulk, with observables that are defined in a slice of dimension 4. The number of bosonic fields that are needed depends on n, and increases as 2^n, according to:

$$
\begin{aligned}
&\text{level } n=0, \quad && \Phi_0 = \phi_1(x) = \phi(x) \\
&\text{level } n=1, \quad && \Phi_1 = \phi_1(x,t^1), \phi_2(x,t^1) \\
& && .... \\
&\text{level } n, && \Phi_n = \phi_1(x,t^1,\dots,t^n), \phi_2(x,t^1,\dots,t^n), \dots \phi_{2^n}(x,t^1,\dots,t^n) \\
& && ....
\end{aligned}
\tag{2.1}
$$

Indeed, if we apply from level n to $n+1$ the process explained in [1], the number of field degrees of freedom that are needed doubles. Thus, we have a multiplet with 2^n components when the bulk has dimension n. For $n=1$, $\phi_2(x,t^1)$ can be heuristically interpreted as the Gaussian noise of a stochastic process. A more fundamental interpretation is that $\phi_2(x,t^1)$ is the canonical moment with respect to the bulk time of the field $\phi_1 = \phi$, and so on.

At every level n, there is a hidden BRST symmetry, defined by the graded differential operator s_n:

$$\begin{array}{ll} s_n\phi_p = \psi_p & s_n\bar{\Psi}_{p+2^{n-1}} = \phi_{p+2^{n-1}} \\ s_n\Psi_p = 0 & s_n\phi_{p+2^{n-1}} = 0, \end{array} \tag{2.2}$$

where, $1 \leq p \leq 2^{n-1}$. The Ψ's and $\bar{\Psi}$'s are ghosts and antighosts, with the opposite statistics to the ϕ's. The set of fields ϕ, Ψ and $\bar{\Psi}$ determine a BRST topological quartet. We can define an anti-BRST operator, which merely interchanges the ghosts and antighosts:

$$\begin{array}{ll} \bar{s}_n\phi_p = \bar{\Psi}_{p+2^{n-1}} & \bar{s}_n\Psi_p = -\phi_{p+2^{n-1}} \\ \bar{s}_n\bar{\Psi}_{p+2^{n-1}} = 0 & \bar{s}_n\phi_{p+2^{n-1}} = 0. \end{array} \tag{2.3}$$

The action that describes the formulation of the theory with a bulk of dimension n must be s_n-exact, that is, it must be of the form:

$$\int \mathrm{d}^4x\mathrm{d}t^1\ldots\mathrm{d}t^n\ s_n\left(\sum_{p=1}^{2^{n-1}}\bar{\Psi}_{p+2^{n-1}}Z_p\right), \tag{2.4}$$

that is,

$$\int \mathrm{d}^4x\mathrm{d}t^1\ldots\mathrm{d}t^n\left(\sum_{p=1}^{2^{n-1}}\phi_{p+2^{n-1}}Z_p - \sum_{p=1}^{2^{n-1}}\bar{\Psi}_{p+2^{n-1}}s_nZ_p\right)\cdot \tag{2.5}$$

To determine the action, it is sufficient to determine the expression of the functionals $Z_p[\Phi_n]$, for every level n. At level n, the ghosts have parabolic propagators along t_n. Thus, they can be integrated out exactly when one computes correlations functions of the Φ's. Indeed, the latter cannot contain closed loops of the ghosts. It follows that it is sufficient to determine the following part of the action:

$$I_n = \int \mathrm{d}^4x\mathrm{d}t^1\ldots\mathrm{d}t^n\sum_{p=1}^{2^{n-1}}\phi_{p+2^{n-1}}Z_p. \tag{2.6}$$

The rest of the action will be determined by the requirement of BRST symmetry. The determination of the factors Z_p is restricted by power counting and by symmetries. The latter include the parity in the bulk, that is, the invariance under $t_p \to -t_p$, for $1 \leq p \leq n$, and by translation invariance. The covariance of the fields under this parity will be shortly determined.

We define the following self-consistent assignments for the dimensions of the bulk coordinates and of the fields:

$$[t^n]^{-1} = 2^n \qquad [\phi_p] = 2p - 1. \tag{2.7}$$

With this assignments, the dimension of the Lagrangian at level n must be $2 + 2^{n+1}$. This ensures that the theory generated by the action I_n is renormalizable by power counting, as a generalization of [1]. It is of relevance to note that:

$$[\phi_1] + [\phi_{2^{n-1}}] + [t^n]^{-1} = 2 + 2^{n+1}. \tag{2.8}$$

This will imply that, in the formulation at level n, $\phi_{2^{n-1}}$ is the conjugate momentum of ϕ_1 with respect to t^n. This turns out to be one of the key facts for proving by induction that the physical content of the theory at level n is the same physics as for the theory at level $n-1$, and so on, down to the ordinary formulation with the action I_0.

3 The action at level n

At level zero, the theory is defined by the standard action $I_0[\phi] = \int \mathrm{d}^4 x L_0$ $(\phi(x))$. At level $n = 1$, it is defined by:

$$I_1[\phi_1, \phi_2] = \int \mathrm{d}^4 x \mathrm{d}t^1 \left(\phi_2 \partial_1 \phi_1 + \phi_2 \left(\phi_2 + \frac{\delta I_0}{\delta \phi_1} \right) \right). \tag{3.1}$$

The exponential of this action must be inserted in the path integral with measure $[\mathrm{d}\phi_1]_{x,t}[\mathrm{d}\phi_2]_{x,t}$. I_1 satisfies power counting according to equation (2.7) (here $[t^1]^{-1} = 2$, $[\phi_1] = 1, [\phi_2] = 3$) and the bulk-parity symmetry P_1 is:

$$\begin{aligned} t^1 &\to -t^1 \\ \phi_1 &\to \phi_1 \\ \phi_2 &\to -\phi_2 - \frac{\delta I_0}{\delta \phi_1}. \end{aligned} \tag{3.2}$$

The action I_1 and its symmetry P_1 have been discussed in details in [1], where we have also shown that it describes the same physics as the action I_0. Notice that the existence of the symmetry P_1 is obvious after the elimination of ϕ_2 by its equation of motion.

At level $n = 2$, the action is:

$$I_2[\phi_1, \phi_2, \phi_3, \phi_4] = \int \mathrm{d}^4 x \mathrm{d}t^1 \mathrm{d}t^2 \left(\phi_3 \partial_2 \phi_1 + \phi_3 \left(\phi_3 + \frac{\delta I_1}{\delta \phi_1} \right) \right.$$
$$\left. + \phi_4 \left(\phi_2 + \frac{\delta I_0}{\delta \phi_1} + \partial_1 \phi_1 \right) \right) \tag{3.3}$$

I_2 is invariant under the bulk-parity transformations P_2 and P_1. P_2 is defined as:

$$\begin{array}{rcl} t_2 & \rightarrow & -t_2 \\ \phi_1, \phi_2 & \rightarrow & \phi_1, \phi_2 \\ \phi_3 & \rightarrow & -\phi_3 - \dfrac{\delta I_1}{\delta \phi_1} \\ \phi_4 & \rightarrow & \phi_4 + \partial_2 \phi_2 . \end{array} \tag{3.4}$$

The action of the symmetry P_2 on ϕ_4 is such that $\delta I_2 = -\int \partial_2(I_1)$. P_1 is defined as:

$$\begin{array}{rcl} t_1 & \rightarrow & -t_1 \\ \phi_1, \phi_3 & \rightarrow & \phi_1, \phi_3 \\ \phi_2 & \rightarrow & -\phi_2 - \dfrac{\delta I_0}{\delta \phi_1} \\ \phi_4 & \rightarrow & -\phi_4 + \phi_3 \dfrac{\delta^2 I_0}{\delta \phi_1 \delta \phi_1} \end{array} \tag{3.5}$$

P_1 transforms ϕ_4 in such a way that the variation of the term $\phi_4(\delta I_2/\delta\phi_2)$ compensates that of $\phi_3(\delta I_1/\delta\phi_1)$.

The parity symmetry under P_1 and P_2 implies that a I_2 has the form displayed in equation (3.3). In this action, power conting implies that $I_0[\phi]$ is a local functional, which can be identified as an action that is renormalizable in 4 dimensions. Thus, no new parameter of physical relevance can be introduced when one switches from the ordinary formulation to the formulation with a bulk. By a straightforward generalization of [1], one can than prove that the correlations functions, computed from I_2 at a given point of the two-dimensional bulk, satisfy the same Dyson–Schwinger equations as those computed from I_1, at a given point of the one-dimensional bulk. In turn, there is the equivalence of the physics computed either from I_1 from I_0, which gives the desired result that we can use a two-dimensional bulk to compute physical quantities with the same result as in the ordinary formulation.

We can now give the general expression of the action at level n, which satisfies power counting and is invariant under all parity transformations in the bulk, $t_p \to -t_p$, for $1 \leq p \leq n$. It reads:

$$\begin{aligned} I_n \;&=\; \int \mathrm{d}^4x\mathrm{d}t^1\mathrm{d}t^2\ldots\mathrm{d}t^n \left(\phi_{1+2^{n-1}}\partial_n\phi_1 + \phi_{1+2^{n-1}}\left(\phi_{1+2^{n-1}} + \frac{\delta I_{n-1}}{\delta\phi_1}\right)\right. \\ &\left. + \sum_{p=2+2^{n-1}}^{2^n} \phi_p \frac{\delta I_{n-1}[\phi_1 \ldots, \phi_{2^{n-1}}]}{\delta\phi_{p-2^{n-1}}}\right) \end{aligned} \tag{3.6}$$

I_n is invariant under the parity transformation P_n, with:

$$\begin{aligned} t_n &\to -t_n \\ \phi_{1+2^{n-1}} &\to -\phi_{1+2^{n-1}} - \frac{\delta I_{n-1}}{\delta\phi_1} \\ \phi_p &\to \phi_p && \text{for } p < 1+2^{n-1} \\ \phi_p &\to \phi_p + \partial_n\phi_{p-2^{n-1}} && \text{for } p > 1+2^{n-1}. \end{aligned} \tag{3.7}$$

Under the symmetry P_n, the Lagrangian density varies by a pure derivative, $\mathcal{L}_n \to \mathcal{L}_n - \partial_n(\mathcal{L}_{n-1})$.

As for the rest of the parity transformations P_p of the fields in the bulk, with $1 \leq p \leq n-1$, their existence can be proven by induction.

Assume that the full bulk parity symmetry exists at level $n-1$, that is, field transformations exist that leave invariant $I_{n-1}[\phi_1, \ldots, \phi_{2^{n-1}}]$ for all transformations $t_p \to -t_p$, $1 \leq p \leq n-1$. Then, the triangular nature of the Jacobian of the transformation $(\phi_1, \ldots, \phi_{2^{n-1}}) \to P_p(\phi_1, \ldots, \phi_{2^{n-1}})$ at level $n-1$ implies that one can extend this transformation law for the new fields that occur at level n, $(\phi_1, \ldots, \phi_{2^n}) \to P_p(\phi_1, \ldots, \phi_{2^n})$ and that I_n, as given in equation (3.6), is invariant under P_p, in a way that generalizes equation (3.5).

Conversely, the parity symmetry and power counting imply that I_n must be of the form (3.6). This shows that the number of parameters of the theory is the same in bulk quantization, with any given choice of the the bulk dimension n, as in the standard formulation. These parameters are just those of an action that is renormalizable by power counting in 4 dimensions.

We can now write the action in the following form:

$$\begin{aligned} \int \mathrm{d}^4x\mathrm{d}t^1\mathrm{d}t^2\ldots\mathrm{d}t^n\; s_n\left(\bar{\Psi}_{1+2^{n-1}}\left(\partial_n\phi_1 + \phi_{1+2^{n-1}}\frac{\delta I_{n-1}[\phi_1 \ldots, \phi_{2^{n-1}}]}{\delta\phi_1}\right)\right. \\ \left. + \sum_{p=2}^{2^{n-1}} \bar{\Psi}_p \frac{\delta I_{n-1}[\phi_1 \ldots, \phi_{2^{n-1}}]}{\delta\phi_p}\right). \end{aligned} \tag{3.8}$$

A propagation occurs in the new direction t^n, while the equation of motions of the formulation at degree $n-1$ are enforced in a BRST invariant way. Because the action is s_n exact, and $\phi_{1+2^{n-1}}$ is the momentum with respect to t_n of ϕ_1, the correlation functions that one can compute in the $(4+n)$-dimensional theory, at an equal bulk component t_n, are identical to those computed in the theory defined by I_{n-1}. The proof is just as in the case of a one-dimensional bulk, and uses the BRST invariance and the translation and parity symmetries in the bulk. Finally, the correlation functions computed in the $(4+n)$-dimensional theory, where all argument only involve a single point T in the bulk, are identical to those computed from the basic four dimensional I_0, that is,

$$G_N^{I_0}\Big(x_1,\dots,x_N\Big) = G_N^{I_n}\Big((x_1,t^{p_1})\dots,(x_N,t^{p_N})\Big)|_{t^{p_1}=\dots=t^{p_N}=T^p}. \quad (3.9)$$

Due to translation invariance, such a correlation function is independent on the choice of T. An another interesting expression of the action at level n is:

$$\int \mathrm{d}^4x\mathrm{d}t^1\mathrm{d}t^2\dots\mathrm{d}t^n \;\big(s_n(\bar\Psi_{1+2^{n-1}}\partial_n\phi_1) \\ + s_n\bar s_n(\bar\Psi_{1+2^{n-1}}\Psi_1 + I_{n-1}[\phi_1,\dots,\phi_{2^{n-1}}])\big). \quad (3.10)$$

It shows that the Hamiltonian at level n is a supersymmetric term, $H = \frac{1}{2}\{Q_n,\bar Q_n\}$, which involves in the very simple way the action at level $n-1$.

This is a pleasure to thank Daniel Zwanziger for discussions related to this work. The research of Laurent Baulieu was supported in part by DOE grant DE-FG02-96ER40959.

References

[1] L. Baulieu and D. Zwanziger, *Nucl. Phys. B* **581** (2000) 604 [hep-th/9909006]; *JHEP* **0108** (2001) 015 [hep-th/0107074].
[2] L. Baulieu, P. Antonio Grassi and D. Zwanziger, *Nucl. Phys. B* **597** (2001) 583 [hep-th/0006036].
[3] G. Parisi and Y.S. Wu. *Sci. Sin.* **24** (1981) 484.
[4] P.H. Daamgard and H. Huffel, *Phys. Rep.* **152** (1983) 227.
[5] L. Baulieu, *A Curious Relation Between Gravity and Yang-yMills Theories*, Talk given in the memory of E.S. Fradkin [hep-th/0007027].
[6] L. Baulieu and M. Bellon, *Phys. Lett. B* **507** (2001) 265 [hep-th/0102075].

1

SEMINAR 2

SPECIAL HOLONOMY SPACES AND M-THEORY

M. CVETIČ

Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, PA 19104, U.S.A.

[1]Based on les Houches 2001 lectures given by M. Cvetič.

Contents

SPECIAL HOLONOMY SPACES AND M-THEORY

M. Cvetič[1,5], G.W. Gibbons[2], H. Lü[3] and C.N. Pope[4,5]

Abstract

We review the construction of regular p-brane solutions of M-theory and string theory with less than maximal supersymmetry whose transverse spaces have metrics with special holonomy, and where additional fluxes allow for brane resolutions *via* transgression terms. We summarize properties of resolved M2-branes and fractional D2-branes, whose transverse spaces are Ricci flat eight-dimensional and seven-dimensional spaces of special holonomy. Recent developments in the construction of new G_2 holonomy spaces are also reviewed.

1 Introduction

Regular supergravity solutions with less than maximal supersymmetry may provide viable gravity duals to strongly coupled field theories with less than maximal supersymmetry. In particular, the regularity of such solutions at small distances sheds light on confinement and chiral symmetry breaking in the infrared regime of the dual strongly coupled field theory [1].

We shall briefly review the construction of such regular supergravity solutions with emphasis on resolved M2-branes of 11-dimensional supergravity and fractional D2-branes of type IIA supergravity, which provide viable gravity duals of strongly coupled three-dimensional theories with $\mathcal{N} = 2$ and $\mathcal{N} = 1$ supersymmetry.

---

[1]Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, PA 19104, U.S.A.

[5]Isaac Newton Institute for Mathematical Sciences, 20 Clarkson Road, University of Cambridge, Cambridge CB3 0EH, UK

[2]DAMTP, Centre for Mathematical Sciences, Cambridge University Wilberforce Road, Cambridge CB3 OWA, UK

[3]Michigan Center for Theoretical Physics, University of Michigan Ann Arbor, MI 48109, U.S.A.

[4]Center for Theoretical Physics, Texas A&M University, College Station, TX 77843, U.S.A.

This construction has been referred to as a “resolution *via* transgression” [2]. It involves the replacement of the standard flat transverse space by a smooth space of special holonomy, *i.e.* a Ricci-flat space with fewer covariantly constant spinors. Furthermore, additional field strength contributions are involved, which are provided by harmonic forms in the space of special holonomy. Transgression–Chern-Simons terms modify the equation of motion and/or Bianchi identity for the original p-brane field strength. In Section 2 the construction will be reviewed in general and then applied to resolved M2-branes and D2-branes.

The explicit construction of such solutions has led to mathematical developments, for example obtaining harmonic forms for a large class of special holonomy metrics. As a prototype example we shall review the construction of the metric and the middle-dimensional forms for the Stenzel manifolds in $D = 2n$ (with $n \geq 2$ integer) [3, 4]. We shall also briefly mention examples of known G_2 holonomy spaces and their associated harmonic forms. We also discuss the old as well as the new two-parameter metric with $Spin(7)$ holonomy [5, 6] and the associated harmonic forms. We shall then briefly summarize the properties of resolved M2-branes [2, 4] and fractional D2-branes [2, 7] as well as fractional M2-brane whose transverse space is that of the new $Spin(7)$ holonomy metrics [5]. All these developments will be reviewed in in Section 3.

In the subsequent Section 4 we review the most recent progress on explicit constructions of G_2 holonomy spaces. In particular, we highlight the construction of general G_2 holonomy spaces whose principal orbits are S^3 bundles over S^3 and the intriguing connection of those spaces to a unified description of deformed and resolved conifolds in six-dimensions.

In Section 5 we also outline directions of current and future research, in particular the study of singular G_2 holonomy spaces and their implications for four-dimensional non-Abelian chiral theories that can arise from a compactification of M-theory on such classes of special holonomy spaces.

The work presented in these lectures was initiated in [2] and further pursued in a series of papers that provide both new technical mathematical results and physics implications for resolved p-brane configurations [3–7]. Recent progress in the construction of new special holonomy spaces was initiated in [5, 6], resulting in the first example of asymptotically locally conical metric with $Spin(7)$ holonomy. A subsequent series of papers, which appeared after the lectures had been given, developed these techniques and provided explicit analyses of classes of cohomogeneity-one G_2 holonomy metrics, with the primary focus on those whose principal orbits are S^3 bundles over S^3 [8–23].

2 Resolution *via* transgression

2.1 Motivation

The AdS_{D+1}/CFT_D correspondence [24–26] provides a quantitative insight into strongly coupled superconformal gauge theories in D dimensions, by studying the dual supergravity solutions. The prototype supergravity dual is the D3-brane of type IIB theory, with the classical solution

$$\begin{aligned} \mathrm{d}s_{10}^2 &= H^{-1/2}\,\mathrm{d}x\cdot\mathrm{d}x + H^{1/2}\,(\mathrm{d}r^2 + r^2\,\mathrm{d}\Omega_5^2)\,, \\ F_{(5)} &= \mathrm{d}^4x\wedge\mathrm{d}H^{-1} + \hat{*}(\mathrm{d}^4x\wedge\mathrm{d}H^{-1})\,, \\ H &= 1+\frac{R^4}{r^4}\cdot \end{aligned} \tag{2.1}$$

In the decoupling limit $H = 1+\frac{R^4}{r^4} \rightarrow \frac{R^4}{r^4}$ this reduces to $AdS_5\times S^5$, which provides a gravitational dual of the strongly coupled $\mathcal{N}=4$ super-Yang-Mills (SYM) theory.

Of course, the ultimate goal of this program is to elucidate strongly coupled YM theory, such as QCD, that has no supersymmetry. But for the time being important steps have been taken to obtain viable (regular) gravitational duals of strongly coupled field theories with less than maximal supersymmetry. In particular, within this framework we shall shed light on gravity duals of field theories in $D=\{2,3,4\}$ with $\mathcal{N}=\{1,2\}$ supersymmetry.

As a side comment, within $D=5$ $\mathcal{N}=2$ gauged supergravity progress has been made (see [27–29] and references therein) in constructing domain wall solutions, both with vector-multiplets *and* hyper-multiplets, which lead to smooth solutions that provide viable gravity duals of $D=4$ $\mathcal{N}=1$ conformal field theories. Note however, that often the higher dimensional interpretation of this approach, and thus a direct connection to string and M-theory, is not clear. The aim in these lectures is to discuss the string and M-theory embeddings of configurations with less than maximal supersymmetry, and the field theory interpretation of such gravity duals.

A procedure for obtaining a supergravity solution with lesser supersymmetry is to replace the flat transverse 6-dimensional space $\mathrm{d}s_6^2 = \mathrm{d}r^2 + r^2\,\mathrm{d}\Omega_5^2$ of the D3-brane in (2.1) with a smooth non-compact Ricci-flat space with fewer Killing spinors. In this case the metric function H still satisfies $\Box H = 0$, but now $\Box$ is the Laplacian in the new Ricci-flat transverse space. This procedure ensures one has a solution with reduced supersymmetry; however the solution for H can be singular at the inner boundary of the transverse space, signifying the appearance of the (distributed) D3-brane source there.

A resolution of the singularity (and the removal of the additional source) can take place if one turns on additional fluxes ("fractional" branes). Within the D3-brane context, the Chern-Simons term of type IIB supergravity modifies the equations of motion:

$$\begin{aligned} \mathrm{d}F_{(5)} &= \mathrm{d}{*}F_{(5)} = F^{\mathrm{NS}}_{(3)} \wedge F^{\mathrm{RR}}_{(3)} = \tfrac{1}{2i} F_{(3)} \wedge \bar{F}_{(3)}\,, \\ F_{(3)} &\equiv F^{\mathrm{RR}}_{(3)} + \mathrm{i}\, F^{\mathrm{NS}}_{(3)} = m L_{(3)}\,, \end{aligned} \tag{2.2}$$

where $L_{(3)}$ is a complex harmonic self-dual 3-form on the 6-dimensional Ricci-flat space. Depending on the properties of L_3, this mechanism may allow for a smooth and thus viable supergravity solution. This is precisely the mechanism employed by Klebanov and Strassler, which in the case of the deformed conifold yields a supergravity dual of $D = 4$ $\mathcal{N} = 1$ SYM theory. (For related and follow up work see, for example, [30–38]. For earlier work see, for example, [39–42].)

In a general context, the resolution *via* transgression [2] is a consequence of the Chern-Simons-type (transgression) terms that are ubiquitous in supergravity theories. Such terms modify the Bianchi identities and/or equations of motion when additional field strengths are turned on. p-brane configurations with $(n+1)$-transverse dimensions, *i.e.* with "magnetic" field strength $F_{(n)}$, can have additional field strengths $F_{(p,q)}$ which, *via* transgression terms, modify the equations for $F_{(n)}$:

$$\mathrm{d}F_{(n)} = F_{(p)} \wedge F_{(q)}; \quad (p + q = n + 1)\,. \tag{2.3}$$

If the $(n+1)$-dimensional transverse Ricci-flat space admits a harmonic p-form $L_{(p)}$ then the equations of motion are satisfied if one sets $F_{(p)} = mL_{(p)}$, and by duality $F_{(q)} \sim \mu * L_{(p)}$. Depending on the L^2 normalizability properties of $L_{(p)}$, one may be able to obtain resolved (non-singular) solutions.

2.2 Resolved M2-brane

The transgression term in the 4-form field equation in 11-dimensional supergravity is given by

$$\mathrm{d}{*}F_{(4)} = \tfrac{1}{2} F_{(4)} \wedge F_{(4)}\,, \tag{2.4}$$

and the modified M2-brane Ansatz takes the form

$$\begin{aligned} \mathrm{d}\hat{s}^2_{11} &= H^{-2/3}\, \mathrm{d}x^\mu\, \mathrm{d}x^\nu\, \eta_{\mu\nu} + H^{1/3}\, \mathrm{d}s^2_8\,, \\ F_{(4)} &= \mathrm{d}^3x \wedge \mathrm{d}H^{-1} + m\, L_{(4)}\,, \end{aligned} \tag{2.5}$$

where $L_{(4)}$ is a harmonic self-dual 4-form in the 8-dimensional Ricci-flat transverse space. The equation for H is then given by

$$\Box H = -\tfrac{1}{48} m^2\, L^2_{(4)}\,. \tag{2.6}$$

For related work see, for example, [34, 44–48].

2.3 Resolved D2-brane

The transgression modification in the 4-form field equation in type IIA supergravity is

$$\mathrm{d}\left(e^{\frac{1}{2}\phi}\,\hat{*}F_4\right) = F_{(4)} \wedge F_{(3)}\,, \tag{2.7}$$

and the modified D2-brane Ansatz takes the form:

$$\begin{aligned} \mathrm{d}s_{10}^2 &= H^{-5/8}\,\mathrm{d}x^\mu\,\mathrm{d}x^\nu\,\eta_{\mu\nu} + H^{3/8}\,\mathrm{d}s_7^2\,, \\ F_{(4)} &= \mathrm{d}^3x \wedge \mathrm{d}H^{-1} + m\,L_{(4)}\,, \quad F_{(3)} = m\,L_{(3)}\,, \quad \phi = \tfrac{1}{4}\log H\,, \end{aligned} \tag{2.8}$$

where $G_{(3)}$ is a harmonic 3-form in the Ricci-flat 7-metric $\mathrm{d}s_7^2$, and $L_{(4)} = *L_{(3)}$, with $*$ the Hodge dual with respect to the metric $\mathrm{d}s_7^2$. The function H satisfies

$$\Box H = -\tfrac{1}{6}m^2 L_{(3)}^2\,, \tag{2.9}$$

where $\Box$ denotes the scalar Laplacian with respect to the transverse 7-metric $\mathrm{d}s_7^2$. Thus the deformed D2-brane solution is completely determined by the choice of Ricci-flat 7-manifold, and the harmonic 3-form supported by it.

2.4 Other examples

In general the transgression terms modify field equations or Bianchi identities as given in (2.3), thus allowing resolved branes with $(n+1)$ transverse dimensions for the following additional examples in M-theory and string theory:

- (i) D0-brane: $\mathrm{d}{*F_{(2)}} = *F_{(4)} \wedge F_{(3)}$;
- (ii) D1-brane: $\mathrm{d}{*F_{(3)}^{\mathrm{RR}}} = F_{(5)} \wedge F_{(3)}^{\mathrm{NS}}$;
- (iii) D4-brane: $\mathrm{d}F_{(4)} = F_{(3)} \wedge F_{(2)}$;
- (iv) IIA string: $\mathrm{d}{*F_{(3)}} = F_{(4)} \wedge F_{(4)}$;
- (v) IIB string: $\mathrm{d}{*F_{(3)}^{\mathrm{NS}}} = F_{(5)} \wedge F_{(3)}^{\mathrm{RR}}$;
- (vi) heterotic 5-brane: $\mathrm{d}F_{(3)} = F_{(2)}^i \wedge F_{(2)}^i$.

In what follows, we shall focus on resolved M2-branes and briefly mention fractional D2-branes. For details of other examples and their properties, see *e.g.*, [2, 3, 49, 50].

3 Special holonomy spaces, harmonic forms and resolved branes

The construction of resolved supergravity solutions necessarily involves the explicit form of the metric on the Ricci-flat special holonomy spaces. These spaces fall into the following classes:

- Kähler spaces in $D = 2n$ dimensions (n-integer) with $SU(n)$ holonomy, and two covariantly constant spinors. There are many examples, with the Stenzel metric on T^*S^n providing a prototype. They are typically asymptotically conical (AC);

- Hyper-Kähler spaces in $D = 4n$ with $Sp(n)$ holonomy, and $n+1$ covariantly constant spinors. Subject to certain technical assumptions, Calabi's metric on the co-tangent bundle of $\mathbf{CP}^n$ is the only complete irreducible cohomogeneity one example [51];

- In $D = 7$ there are exceptional G_2 holonomy spaces with one covariantly constant spinor. Until recently only three AC examples were known [52, 53], but new metrics have been recently constructed in [9, 10, 12–23] and will be discussed in a separate Section 4;

- In $D = 8$ there are exceptional $Spin(7)$ holonomy spaces with one covariantly constant spinor; until recently only one AC example was known [52, 53]. New metrics were recently constructed in [5, 6, 12, 14, 15].

Here the focus is on a construction of cohomogeneity one spaces that are typically asymptotic to cones over Einstein spaces. Recent mathematical developments evolved in two directions: (i) construction of harmonic forms on known Ricci-flat spaces (see in particular [3, 7]), (ii) construction of new exceptional holonomy spaces [5, 6, 9, 10, 12–23]. In the following subsections we illustrate these developments by

- Summarizing results on the construction of harmonic forms on the Stenzel metric [3];

- Briefly mentioning the results for the old G_2 holonomy metrics [2, 7];

- Presenting results for the new $Spin(7)$ two-parameter metrics [5, 6], and

- Summarizing the implications of these prototype special holonomy spaces for resolved M2-branes and D2-branes.

In Section 4 we shall then summarize the recent new constructions of G_2 holonomy spaces and the implications for M-theory dynamics on such spaces.

3.1 Harmonic forms for the Stenzel metric

The Stenzel [54] construction provides a class of complete non-compact Ricci-flat Kähler manifolds, one for each even dimension, on the co-tangent bundle of the $(n{+}1)$-sphere, $T^\star S^{n+1}$. These are asymptotically conical, with principal orbits that are described by the coset space $SO(n{+}2)/SO(n)$, and they have real dimension $d = 2n + 2$.

3.1.1 Stenzel metric

In the following we summarize the relevant results for the construction of the Stenzel metric. (For more details see [3].) This construction [3, 54] of the Stenzel metric starts with L_{AB}, which are left-invariant 1-forms on the group manifold $SO(n + 2)$. By splitting the index as $A = (1, 2, i)$, we have that L_{ij} are the left-invariant 1-forms for the $SO(n)$ subgroup, and so the 1-forms in the coset $SO(n + 2)/SO(n)$ will be

$$\sigma_i \equiv L_{1i}\,, \qquad \tilde\sigma_i \equiv L_{2i}\,, \qquad \nu \equiv L_{12}\,. \tag{3.1}$$

The metric Ansatz takes the form:

$$\mathrm{d}s^2 = \mathrm{d}t^2 + a^2\sigma_i^2 + b^2\,\tilde\sigma_i^2 + c^2\,\nu^2\,, \tag{3.2}$$

where a, b and c are functions of the radial coordinate t. One defines Vielbeine

$$e^0 = \mathrm{d}t\,, \qquad e^i = a\,\sigma_i\,, \qquad e^{\tilde i} = b\,\tilde\sigma_i\,, \qquad e^{\tilde 0} = c\,\nu\,, \tag{3.3}$$

for which one can introduce a holomorphic tangent-space basis of complex 1-forms ϵ^α:

$$\epsilon^0 \equiv -e^0 + \mathrm{i}\,e^{\tilde 0}\,, \qquad \epsilon^i = e^i + \mathrm{i}\,e^{\tilde i}\,. \tag{3.4}$$

Defining $a = e^\alpha$, $b = e^\beta$, $c = e^\gamma$, and introducing the new coordinate η by $a^n\,b^n\,c\,\mathrm{d}\eta = \mathrm{d}t$, one finds [3] that the Ricci-flat equations can be obtained from a Lagrangian $L = T - V$ which can be written as a "supersymmetric Lagrangian": $L = \frac12 g_{ij}\,(\mathrm{d}\alpha^i/\mathrm{d}\eta)\,(\mathrm{d}\alpha^j/\mathrm{d}\eta) - \frac12 g^{ij}\,\frac{\partial W}{\partial\alpha^i}\,\frac{\partial W}{\partial\alpha^j}$. The solution of the first-order equations yields the explicit solution:

$$a^2 = R^{\frac{1}{n+1}}\,\coth r\,; b^2 = R^{\frac{1}{n+1}}\,\tanh r\,; h^2 = c^2 = \tfrac{1}{n+1} R^{-\frac{n}{n+1}}\sinh^n(2r)\,, \tag{3.5}$$

where $R(r) \equiv \int_0^r (\sinh 2u)^n\,\mathrm{d}u$, and the radial coordinate r is introduced as $\mathrm{d}t = h\,\mathrm{d}r$.

For each n the result is expressible in relatively simple terms. For example,

$$R = \sinh^2 r\,; R = \tfrac{1}{8}(\sinh 4r - 4r)\,; R = \tfrac{2}{3}(2 + \cosh 2r)\sinh^4 r\,, \qquad (3.6)$$

for $n = 1, 2, 3$, respectively. The case $n = 1$ is the Eguchi-Hanson metric [43], and $n = 2$ it is the deformed conifold [55].

As r approaches zero, the metric takes the form

$$\mathrm{d}s^2 \sim \mathrm{d}r^2 + r^2 \tilde{\sigma}_i^2 + \sigma_i^2 + \nu^2\,, \qquad (3.7)$$

which has the structure locally of the product $\mathbf{R}^{n+1} \times S^{n+1}$, with S^{n+1} being a "bolt". As r tends to infinity, the metric becomes

$$\mathrm{d}s^2 \sim \mathrm{d}\rho^2 + \rho^2 \left\{ \frac{n^2}{(n+1)^2}\nu^2 + \frac{n}{2(n+1)^2}\left(\sigma_i^2 + \tilde{\sigma}_i^2\right) \right\}, \qquad (3.8)$$

representing a cone over the Einstein space $SO(n+2)/SO(n)$.

3.1.2 Harmonic middle-dimension (p, q) forms

An Ansatz compatible with the symmetries of the Stenzel metric is of the form:

$$\begin{aligned} L_{(p,q)} \;=\;& f_1\, \epsilon_{i_1\cdots i_{q-1} j_1 \cdots j_p}\, \bar{\epsilon}^0 \wedge \bar{\epsilon}^{i_1} \wedge \cdots \wedge \bar{\epsilon}^{i_{q-1}} \wedge \epsilon^{j_1} \wedge \cdots \wedge \epsilon^{j_p} \\ &+ f_2\, \epsilon_{i_1\cdots i_{p-1} j_1 \cdots j_q}\, \epsilon^0 \wedge \epsilon^{i_1} \wedge \cdots \wedge \epsilon^{i_{p-1}} \wedge \bar{\epsilon}^{j_1} \wedge \cdots \wedge \bar{\epsilon}^{j_q}\,, \end{aligned} \qquad (3.9)$$

with f_1, f_2 being functions of r, only. The harmonicity condition becomes $\mathrm{d}L_{(p,q)} = 0$, since $*L_{(p,q)} = \mathrm{i}^{\,p-q}\, L_{(p,q)}$. The functions f_1, f_2 are solutions of coupled first-order homogeneous differential equations, yielding a solution that is finite as $r \to 0$:

$$f_1 = \; q\, {}_2F_1\left[\tfrac{1}{2}p, \tfrac{1}{2}(q+1), \tfrac{1}{2}(p+q)+1; -(\sinh 2r)^2\right], \qquad (3.10)$$

$$f_2 = -\, p\, {}_2F_1\left[\tfrac{1}{2}q, \tfrac{1}{2}(p+1), \tfrac{1}{2}(p+q)+1; -(\sinh 2r)^2\right]. \qquad (3.11)$$

For any specific integers (p, q), these are elementary functions of r.

For the two special cases of greatest interest, they have the following properties:

- (p, p)-forms in $4p$-dimensions: $f_1 = -f_2 = \frac{p}{(\cosh r)^{2p}}$ with $|L_{(p,p)}|^2 = \frac{\text{const.}}{(\cosh r)^{4p}}$ falls-off fast enough as $r \to \infty$. This turns out to be the only L^2 normalizable form;

- $(p+1,p)$-forms in $(4p+2)$-dimensions. As $r \to \infty$: $|L_{(p+1,p)}|^2 \sim \frac{1}{[\sinh(2r)]^{2p}}$ which is marginally L^2-non-normalizable.

From the viewpoint of physics, the case in $2(n+1) = 4$ dimensions with an L^2-normalizable $L_{(1,1)}$-form is precisely the example of the resolved self-dual string discussed in [2].

In $2(n+1) = 6$ dimensions, the $L_{(2,1)}$-form was constructed in [1], and provides a resolution of the D3-brane. Since $L_{(2,1)}$ is only *marginally non-normalizable* as $r \to \infty$, the decoupling limit of the space-time does not give an AdS_5, but instead there is a logarithmic modification. In particular, this modification accounts for a renormalization group running of the difference of the inverse-squares of the two gauge group couplings in the dual $SU(N) \times SU(N+M)$ SYM [41].

On the other hand in $2(n+1) = 8$ dimensions the L^2 normalizable $L_{(2,2)}$-form supports additional fluxes that resolve the original M2-brane, whose details are given in [3].

It turns out that one can construct regular supersymmetric resolved M2-branes for many other examples of 8-dimensional special holonomy transverse spaces, such as the original $Spin(7)$ holonomy transverse space [2], a number of new Kähler spaces [2,7], and hyper-Kähler spaces [4].

3.2 Old G_2 holonomy metrics and their harmonic forms

3.2.1 Resolved cones over $S^2 \times S^4$ and $S^2 \times \mathbf{CP}^2$

The first complete Ricci-flat 7-dimensional metrics of G_2 holonomy were obtained in [52, 53]. There were two types. The first type comprises two examples of R^3 bundles over four-dimensional quaternionic-Kähler Einstein base manifolds M. These spaces are of cohomogeneity one, with principal orbits that are S^2 bundles over M (sometimes referred to as the twistor space over M). For the two examples that arise, M is S^4 or $\mathbf{CP}^2$. The two G_2 manifolds have principal orbits that are $\mathbf{CP}^3$ (S^2 bundle over S^4), or the flag manifold $SU(3)/(U(1) \times U(1))$ (S^2 bundle over $\mathbf{CP}^2$), respectively. These two manifolds are the bundles of self-dual 2-forms over S^4 or $\mathbf{CP}^2$ respectively. They approach $\mathbf{R}^3 \times S^4$ or $\mathbf{R}^3 \times \mathbf{CP}^2$ locally near the origin.

The derivations for the two cases, with the principal orbits being S^2 bundles either over S^4 or over $\mathbf{CP}^2$, proceed essentially identically. In the notation of [53], the G_2 metrics are given by

$$d\hat{s}_7^2 = h^2\, dr^2 + a^2\, (D\mu^i)^2 + b^2\, ds_4^2\,, \tag{3.12}$$

where μ^i are coordinates on $\mathbf{R}^3$ subject to $\mu^i\,\mu^i = 1$, and ds_4^2 is the metric on S^4 or $\mathbf{CP}^2$ scaled to have $R_{\alpha\beta} = 3g_{\alpha\beta}$. The 1-forms A^i are $su(2)$

Yang-Mills instanton potentials, and

$$D\mu^i \equiv \mathrm{d}\,\mu^i + \epsilon_{ijk}\, A^j\, \mu^k\,. \tag{3.13}$$

The field strengths $J^i \equiv \mathrm{d}A^i + \frac{1}{2}\epsilon_{ijk}\, A^j \wedge A^k$ satisfy the algebra of the unit quaternions, $J^i_{\alpha\gamma}\, J^j_{\gamma\beta} = -\delta_{ij}\,\delta_{\alpha\beta} + \epsilon_{ijk}\, J^k_{\alpha\beta}$. The harmonic 3-form (other than the covariant constant one), was constructed in [4]: it is smooth and L^2-normalizable.

3.2.2 Resolved cone over $S^3 \times S^3$

The second type of complete 7-dimensional manifold of G_2 holonomy obtained in [52, 53] is again of cohomogeneity one, with principal orbits that are topologically $S^3 \times S^3$. The manifold is the spin bundle of S^3; near the origin it approaches locally $\mathbf{R}^4 \times S^3$.

The Ricci-flat metric on the spin bundle of S^3 is given by [52, 53].

$$\mathrm{d}s_7^2 = \alpha^2\,\mathrm{d}r^2 + \beta^2\left(\sigma_i - \frac{1}{2}\Sigma_i\right)^2 + \gamma^2\,\Sigma_i^2\,, \tag{3.14}$$

where the functions α, β and γ are given by

$$\alpha^2 = \left(1 - \frac{1}{r^3}\right)^{-1}, \qquad \beta^2 = \tfrac{1}{9}r^2\left(1 - \frac{1}{r^3}\right), \qquad \gamma^2 = \tfrac{1}{12}r^2\,. \tag{3.15}$$

Here Σ_i and σ_i are the two sets of left-invariant 1-forms on two independent $SU(2)$ group manifolds. The principal orbits $r =$ constant are therefore S^3 bundles over S^3. Since the bundle is trivial, they are topologically $S^3 \times S^3$, although not with the standard product metric. The radial coordinate runs from $r = a$ to $r = \infty$.

This metric admits a regular harmonic 3-form, explicitly constructed in [2]: it is square-integrable at short distance, but gives a linearly divergent integral at large distance. The short-distance square-integrability is enough to give a regular deformed D2-brane solution, even though $L_{(3)}$ is not L^2-normalizable.

3.3 New Spin(7) holonomy metrics and their harmonic forms

3.3.1 The old metric and harmonic 4-forms

Until recently only one explicit example of a complete non-compact metric on a $Spin(7)$ holonomy space was known [52, 53]. The principal orbits are S^7, viewed as an S^3 bundle over S^4. The solution (3.16) is asymptotic to a cone over the "squashed" Einstein 7-sphere, and it approaches $\mathbf{R}^4 \times S^4$ locally at short distance (*i.e.* $r \approx \ell$). The metric is of the form:

$$\mathrm{d}s_8^2 = \left(1 - \frac{\ell^{10/3}}{r^{10/3}}\right)^{-1} \mathrm{d}r^2 + \tfrac{9}{100}\, r^2 \left(1 - \frac{\ell^{10/3}}{r^{10/3}}\right) h_i^2 + \tfrac{9}{20} r^2\, \mathrm{d}\Omega_4^2\,, \tag{3.16}$$

where $h_i \equiv \sigma_i - A^i_{(1)}$, and the σ_i are left-invariant 1-forms on $SU(2)$, $\mathrm{d}\Omega_4^2$ is the metric on the unit 4-sphere, and $A^i_{(1)}$ is the $SU(2)$ Yang-Mills instanton on S^4. The σ_i can be written in terms of Euler angles as $\sigma_1 = \cos\psi\, \mathrm{d}\theta + \sin\psi\, \sin\theta\, \mathrm{d}\varphi$, $\sigma_2 = -\sin\psi\, \mathrm{d}\theta + \cos\psi\, \sin\theta\, \mathrm{d}\varphi$, $\sigma_3 = \mathrm{d}\psi + \cos\theta\, \mathrm{d}\varphi$. A regular L^2 normalizable harmonic 4-form in this metric was obtained in [2].

3.3.2 The new $Spin(7)$ holonomy metric

The generalization that we shall consider involves allowing the S^3 fibers of the previous construction themselves to be "squashed". Namely, the S^3 bundle is itself written as a $U(1)$ bundle over S^2 leading to the following "twice squashed" Ansatz:

$$\mathrm{d}\hat{s}_8^2 = \mathrm{d}t^2 + a^2\, (D\mu^i)^2 + b^2\, \sigma^2 + c^2\, \mathrm{d}\Omega_4^2\,, \tag{3.17}$$

where a, b and c are functions of the radial variable t. (The previous $Spin(7)$ example has $a = b$.) Here

$$\mu_1 = \sin\theta\, \sin\psi\,, \qquad \mu_2 = \sin\theta\, \cos\psi\,, \qquad \mu_3 = \cos\theta\,, \tag{3.18}$$

are the S^2 coordinates, subject to the constraint $\mu_i \mu_i = 1$, and

$$D\mu^i \equiv \mathrm{d}\mu^i + \epsilon_{ijk}\, A^j_{(1)}\, \mu^k\,, \quad \sigma \equiv \mathrm{d}\varphi + \mathcal{A}_{(1)}\,, \quad \mathcal{A}_{(1)} \equiv \cos\theta\, \mathrm{d}\psi - \mu^i\, A^i_{(1)}\,, \tag{3.19}$$

where the field strength $\mathcal{F}_{(2)}$ of the $U(1)$ potential $\mathcal{A}_{(1)}$ turns out to be given by: $\mathcal{F}_{(2)} = \frac{1}{2}\epsilon_{ijk}\, \mu^k\, D\mu^i \wedge D\mu^j - \mu^i\, F^i_{(2)}$.

The Ricci-flatness conditions can be satisfied by solving the first-order equations coming from a supersymmetric Lagrangian, yielding the following special solution (for details see [5,6]):

$$\begin{aligned} \mathrm{d}s_8^2 \;=\;& \frac{(r-\ell)^2\, \mathrm{d}r^2}{(r-3\ell)(r+\ell)} + \frac{\ell^2\, (r-3\ell)(r+\ell)}{(r-\ell)^2}\, \sigma^2 \\ &+ \tfrac{1}{4}(r^2 - 3\ell)(r+\ell)\, (D\mu^i)^2 + \tfrac{1}{2}(r^2 - \ell^2)\, \mathrm{d}\Omega_4^2. \end{aligned} \tag{3.20}$$

The quantity $\frac{1}{4}[\sigma^2 + (D\mu^i)^2]$ is the metric on the unit 3-sphere, and so in this case we find that the metric smoothly approaches $\mathbf{R}^4 \times S^4$ locally, at small distance ($r \to 3\ell$), in the same way that it does in the previously-known example. Therefore it has the same topology as the old $Spin(7)$ holonomy space. On the other hand, it locally approaches $\mathcal{M}_7 \times S^1$ at large distance. Here $\mathcal{M}_7$ denotes the 7-manifold of G_2 holonomy on the $\mathbf{R}^3$ bundle over S^4 [52,53]. Asymptotically the new metric behaves like a circle bundle over an asymptotically conical manifold in which the length of the $U(1)$ fibers tends to a constant; in other words, it is ALC.

If one takes r to be negative, or instead analytically continues the solution so that $\ell \to -\ell$ (keeping r positive), one gets a different complete manifold. Thus instead of (3.20), the quantity $\frac{1}{4}(\sigma^2 + (D\mu^i)^2 + \mathrm{d}\Omega_4^2)$ is precisely the metric on the unit 7-sphere, and so as r approaches ℓ the metric $\mathrm{d}s_8^2$ smoothly approaches $\mathbf{R}^8$. At large r the function b, which is the radius in the $U(1)$ direction σ, approaches a constant, and so the metric tends to an S^1 bundle over a 7-metric of the form of a cone over $\mathbf{CP}^3$; it has the same asymptotic form as (3.20). The manifold in this case is topologically $\mathbf{R}^8$.

In [5,6] the general solution to the first-order system of equations is obtained, leading to additional families of regular metrics of $Spin(7)$ holonomy, which are complete on manifolds $\mathbf{B}_8^{\pm}$ that are similar to $\mathbf{B}_8$. These additional metrics have a non-trivial integration constant which parameterizes inequivalent solutions. (For details see [6] and Appendix A of [5].)

L^2 normalizable harmonic 4-forms for the new $Spin(7)$ 8-manifolds were obtained in [5].

3.4 Applications: Resolved M2-branes and D2-branes

The explicit construction of harmonic 4-forms on 8-dimensional Ricci flat spaces led to analytic expressions for resolved M2-brane solutions, while the 3-forms (and dual 4-forms) of G_2 holonomy spaces led to analytic expressions for fractional D2-branes. Their properties, such as supersymmetry conditions, flux integrals and aspects of the dual field theories, were discussed in [2–4,7,48].

Resolved M2-branes on a suitable eight-dimensional space can be supported by L^2-normalizable harmonic forms and thus they are regular at short distance and have decoupling limits at large distance that yield AdS_4. They have no conserved additional (fractional) charges. The dual 3-dimensional field theory is superconformal (with $\mathcal{N} = 1$ or $\mathcal{N} = 2$ supersymmetry), and is in turn perturbed by marginal operators associated with pseudo-scalar fields [48]. On the other hand the fractional D2-branes have conserved fractional charges corresponding to D4-branes wrapping the 2-cycles dual to S^4 or $\mathbf{CP}^2$ in $\mathcal{M}_7$, or to NS-NS 5-branes wrapping the 3-cycle dual to S^3 in $\mathcal{M}_7$.

An interesting application of these new $Spin(7)$ holonomy spaces is the construction of fractional M2-branes. After reduction on S^1 these give D2-branes with additional fractional magnetic charge associated with D4-branes wrapping 2-cycles *and* D6-branes wrapping 4-cycles. The fact that the resolved M2-brane on the new $Spin(7)$ holonomy space has non-zero fractional charge is a consequence of the asymptotically locally conical structure of the new $Spin(7)$ holonomy space.

4 New G_2 holonomy metrics

Subsequent to these lectures there have been major developments the construction of new holonomy spaces and the study of their implications for M-theory dynamics. In part motivated by the construction of the new two-parameter $Spin(7)$ holonomy metrics with ALC structure [5, 6] (described in the previous Sect. 3), new constructions of G_2 (as well as $Spin(7)$) holonomy spaces [9, 10, 12–23] have been given. The implications from M-theory on such spaces for the dynamics of the resulting $\mathcal{N} = 1$, $D = 4$ field theory [56–58] are attracting considerable attention. Specifically, it has been proposed that M-theory compactified on a certain singular seven-dimensional space with G_2 holonomy might be related to an $\mathcal{N} = 1$, $D = 4$ gauge theory [26, 56–58, 60] that has no conformal symmetry. The quantum aspects of M-theory dynamics on spaces of G_2 holonomy can provide insights into non-perturbative aspects of four-dimensional $\mathcal{N} = 1$ field theories, such as the preservation of global symmetries and phase transitions. For example, reference [56] provides an elegant exposition and study of these phenomena using the three original manifolds of G_2 holonomy that were obtained in [52, 53].

One related development in this direction is the discovery of M3-brane configurations [8, 9]. These have a flat 4-dimensional world-volume and a transverse space that is a deformation of the G_2 manifold, and with the 4-form field strength is turned on. They turn out to have zero charge and ADM mass, leading to naked singularities at small distances.

4.1 Classification of G_2 holonomy spaces with $S^3 \times S^3$ orbits

In another recent development, G_2 metrics have been obtained that make contact with the six-dimensional resolved and deformed conifolds. This work is described in detail in [20] (see also [18, 19]). We consider a generalization of the original Ansatz [52, 53] (3.14) for metrics of cohomogeneity one with $S^3 \times S^3$ principal orbits. The more general Ansatz is given by

$$\begin{aligned} \mathrm{d}s_7^2 = {} & \mathrm{d}t^2 + c^2\,(\Sigma_3 - \sigma_3)^2 + f^2\,(\Sigma_3 + g_3\,\sigma_3)^2 \\ & + a^2\,[(\Sigma_1 + g\,\sigma_1)^2 + (\Sigma_2 + g\,\sigma_2)^2] + b^2\,[(\Sigma_1 - g\,\sigma_1)^2 + (\Sigma_2 - g\,\sigma_2)^2], \end{aligned} \tag{4.1}$$

where Σ_i and σ_i are again two sets of left-invariant 1-forms of $SU(2)$, and the six coefficients a, b, c, f, g and g_3 depend only on t. In the orthonormal basis

$$\begin{aligned} &e^0 = \mathrm{d}t\,, \quad e^1 = a\,(\Sigma_1 + g\,\sigma_1)\,, \quad e^2 = a\,(\Sigma_2 + g\,\sigma_2)\,, \quad e^3 = c\,(\Sigma_3 - \sigma_3)\,,\\ &e^4 = b\,(\Sigma_1 - g\,\sigma_1)\,, \quad e^5 = b\,(\Sigma_2 - g\,\sigma_2)\,, \quad e^6 = f\,(\Sigma_3 + g_3\,\sigma_3)\,, \end{aligned} \tag{4.2}$$

there is a natural candidate for an invariant associative 3-form, namely

$$\begin{aligned} \Phi &= e^0 \wedge (e^1 \wedge e^4 + e^2 \wedge e^5 + e^3 \wedge e^6) - (e^1 \wedge e^2 - e^4 \wedge e^5) \wedge e^3\\ &\quad +(e^1 \wedge e^5 - e^2 \wedge e^4) \wedge e^6\,. \end{aligned} \tag{4.3}$$

Requiring the closure and co-closure of this 3-form gives a set of first-order equations for G_2 holonomy [20],

$$\begin{aligned} \dot a &= \frac{c^2\,(a^2-b^2) + [4a^2\,(a^2-b^2) - c^2\,(5a^2-b^2) - 4a\,b\,c\,f]\,g^2}{16a^2\,b\,c\,g^2}\,,\\ \dot b &= -\frac{c^2\,(a^2-b^2) + [4b^2\,(a^2-b^2) + c^2\,(5b^2-a^2) - 4a\,b\,c\,f]\,g^2}{16a\,b^2\,c\,g^2}\,,\\ \dot c &= \frac{c^2 + (c^2 - 2a^2 - 2b^2)\,g^2}{4a\,b\,g^2}\,,\\ \dot f &= -\frac{(a^2-b^2)\,[4a\,b\,f^2\,g^2 - c\,(4a\,b\,c + a^2\,f - b^2\,f)\,(1-g^2)]}{16a^3\,b^3\,g^2}\,,\\ \dot g &= -\frac{c\,(1-g^2)}{4a\,b\,g}\,, \end{aligned} \tag{4.4}$$

together with an algebraic equation for g_3:

$$g_3 = g^2 - \frac{c\,(a^2-b^2)(1-g^2)}{2a\,b\,f}\,. \tag{4.5}$$

There are two combinations of the equation (4.4) that can be integrated explicitly, giving two invariants built out of the metric functions. These two constants are nothing but the coefficients in front of the volume forms for the respective three-spheres in the associated three-form: $\Phi = m\,\sigma_1 \wedge \sigma_2 \wedge \sigma_3 + n\,\Sigma_1 \wedge \Sigma_2 \wedge \Sigma_3 + \cdots$, which may be seen to be constant by imposing closure of Φ (see [19]). Ultimately, the system (4.4) can be reduced to a single non-linear second-order differential equation.

The general solution of the first-order equation (4.4) is not known. Of course the asymptotically conical G_2 metric (3.14) is a solution. An explicit, singular, solution was found in [18,19]. Another exact solution, found earlier in [10], is

$$\begin{aligned} ds_7^2 &= \frac{(r^2-\ell^2)}{(r^2-9\ell^2)}\,dr^2\,\tfrac{1}{12}(r-\ell)(r+3\ell)[(\Sigma_1-\sigma_1)^2+(\Sigma_2-\sigma_2)^2] \\ &\quad +\tfrac{1}{12}(r+\ell)(r-3\ell)[(\Sigma_1+\sigma_1)^2+(\Sigma_2+\sigma_2)^2] \\ &\quad +\tfrac{1}{9}r^2\,(\Sigma_3-\sigma_3)^2+\tfrac{4}{9}\ell^2\,\frac{r^2-9\ell^2}{r^2-\ell^2}\,(\Sigma_3+\sigma_3)^2\,. \end{aligned} \tag{4.6}$$

The radial coordinate runs from an S^3 bolt at $r=3\ell$ to an asymptotic region as r approaches infinity. The metric is asymptotically locally conical, with the radius of the circle with coordinate $(\psi+\widetilde{\psi})$ stabilising at infinity. The metric is closely analogous to an ALC $Spin(7)$ metric on the $\mathbf{R}^4$ bundle over S^4 that was found previously [5, 6].

Although explicit solutions to the first-order system (4.4) are not in general known, it is nevertheless possible to study the system by a combination of approximation and numerical methods. Specifically, one can perform a Taylor expansion around the bolt at a minimum radius where the $S^3\times S^3$ orbits degenerate, and use this to set initial data just outside the bolt for a numerical integration towards large radius. The criterion for a complete non-singular metric is that the metric functions should be well-behaved at large distance, either growing linearly with distance as in an AC metric, or else with one or more metric coefficients stabilising to fixed values asymptotically, as in an ALC metric such as (4.6). This method is discussed in detail in [13, 18, 20], and it is established there that there exist three families of non-singular ALC metrics, each with a non-trivial parameter λ that gives the size of a stabilising circle at infinity relative to the size of the bolt at short distance. The metrics, denoted by $\mathbf{B}_7$, $\mathbf{D}_7$ and $\widetilde{\mathbf{C}}_7$, have bolts that are a round S^3, a squashed S^3 and $T^{p,q}=S^3\times S^3/U(1)_{(p,q)}$ respectively, where $p/q=\sqrt{m/n}$ and m,n are the two explicit integration constants of the first-order system (4.4) that we discussed previously. The radius of the stabilising circle ranges from zero at $\lambda=0$ to infinity at $\lambda=\infty$. As one takes the limit $\lambda\to 0$, the ALC G_2 metric approaches the direct product of a six-dimensional Ricci-flat Kähler metric and a vanishing circle. This limit is known mathematically as the Gromov-Hausdorff limit.

The cases of most immediate interest are $\mathbf{B}_7$ and $\mathbf{D}_7$. Their Gromov-Hausdorff limits are a vanishing circle times the deformed conifold, and a vanishing circle times the resolved conifold, respectively [18, 20]. On the other hand, as λ goes to infinity, they both approach the original AC metric of [52, 53]. If, therefore, we begin with a solution $(\text{Minkowski})_4\times Y_7$ in M-theory, with Y_7 being a $\mathbf{B}_7$ or $\mathbf{D}_7$ metric, then we can dimensionally reduce it on the circle that stabilises at infinity, thereby obtaining a solution of the type IIA string. The radius of the M-theory circle, R, is related to the string coupling constant $g_{\rm str}$ by $g_{\rm str}=R^{3/2}$. This means that taking the Gromov-Hausdorff limit in $\mathbf{B}_7$ or $\mathbf{D}_7$ corresponds to the weak-coupling limit in the type IIA string, and the ten-dimensional solution becomes the product of (Minkowski)$_4$ with the deformed or resolved conifold. In the strong-coupling domain, where λ goes to infinity, these two ten-dimensional solutions become unified *via* the $\mathbf{B}_7$ and $\mathbf{D}_7$ solutions in M-theory.

A yet more general system of cohomogeneity one G_2 metrics with $S^3\times S^3$ principal orbits was obtained recently in [23]. The construction was based on an approach developed recently by Hitchin [61], in which one starts from an Ansatz for an associative 3-form, and derives first-order equations *via* a system of Hamiltonian flow equations. These first-order equations can be shown to imply that a certain metric derived from the 3-form has G_2 holonomy. By applying this procedure to the case of $S^3 \times S^3$ principal orbits, it was shown in [23] that the metric

$$\begin{aligned} ds_7^2 = dt^2 & \\ &+\frac{1}{y_1}[(n\,x_1+x_2\,x_3)\Sigma_1^2+(m\,n+x_1^2-x_2^2-x_3^2)\Sigma_1\,\sigma_1+(m\,x_1+x_2\,x_3)\sigma_1^2] \\ &+\frac{1}{y_2}[(n\,x_2+x_3\,x_1)\Sigma_2^2+(m\,n+x_2^2-x_3^2-x_1^2)\,\Sigma_2\sigma_2+(m\,x_2+x_3\,x_1)\sigma_2^2] \\ &+\frac{1}{y_3}[(n\,x_3+x_1\,x_2)\Sigma_3^2+(m\,n+x_3^2-x_1^2-x_2^2)\,\Sigma_3\sigma_3+(m\,x_3+x_1\,x_2)\sigma_3^2] \end{aligned} \tag{4.7}$$

has G_2 holonomy if the functions x_i and y_i, which depend only on t, satisfy the first-order Hamiltonian system of equations

$$\dot{x}_1 = \sqrt{\frac{y_2\,y_3}{y_1}}\,, \qquad \dot{y}_1 = \frac{m\,n\,x_1 + (m+n)\,x_2\,x_3 + x_1\,(x_2^2+x_3^2-x_1^2)}{\sqrt{y_1\,y_2\,y_3}}\,, \tag{4.8}$$

and cyclically for the 2 and 3 directions. In addition, the conserved Hamiltonian must vanish, which implies that

$$\begin{aligned} &4y_1\,y_2\,y_3 + m^2\,n^2 - 2m\,n\,(x_1^2+x_2^2+x_3^2) - 4(m+n)\,x_1\,x_2\,x_3 \\ &+x_1^4+x_2^4+x_3^4-2x_1^2\,x_2^2-2x_2^2\,x_3^2-2x_3^2\,x_1^2 = 0\,. \end{aligned} \tag{4.9}$$

The above first-order system encompasses all the previous cases as specialisations. In particular, the first-order system for the metrics (4.1) is obtained by making the specialisation $x_1 = x_2$, $y_1 = y_2$. If, instead, one sets $m = n = 1$, the system reduces to one studied in [9, 10].

In addition to these $SU(2) \times SU(2)$ invariant metrics with principal orbits $S^3 \times S^3$, one may take various Inönü-Wigner contractions to give metrics with principal orbits $T^3 \times S^3$, or other orbit types constructed from the possible contractions of $SU(2)$ [23]. In the particular case of $T^3 \times S^3$ orbits, the resulting first-order system is that of [22].

We find that the general set of equation (4.8) does not seem to yield new classes of regular solutions, other than those already classified [20] for the first-order system (4.4).

5 Conclusions and open avenues

In these lectures we have presented a summary of some recent developments in the construction of regular p-brane configurations with less than maximal supersymmetry. In particular, the method involves the introduction of complete non-compact special holonomy metrics and additional fluxes, supported by harmonic-forms in special holonomy spaces, which modify the original p-brane solutions *via* Chern-Simons (*transgression*) terms.

The work led to a number of important *mathematical developments* which we have also summarized. Firstly, the construction of harmonic forms for special holonomy spaces in diverse dimensions was reviewed, and the explicit construction of harmonic forms for Stenzel metrics was summarized. Secondly, a construction of new two-parameter $Spin(7)$ holonomy spaces was discussed. These have the property that they interpolate asymptotically between a local $S^1 \times \mathcal{M}_7$, where the length of the circle is finite and $\mathcal{M}_7$ is the G_2 holonomy space with the topology of the S^2 bundle over S^4, while at small distance they approach the "old" $Spin(7)$ holonomy space with the topology of the chiral spin bundle over S^4.

These mathematical developments also led to a number of important *physics implications*, relevant for the properties of the resolved p-brane solutions. In particular, the focus was on the properties of resolved M2-branes with 8-dimensional special holonomy transverse spaces, for example Stenzel, hyper-Kähler and $Spin(7)$ holonomy spaces, and the results for the fractional D2-branes with three 7-dimensional G_2 holonomy transverse spaces.

After the lectures were given, there was major progress in constructing new G_2 holonomy spaces and studying the M-theory dynamics on such spaces. We have summarized this progress in Section 4, and in particular highlighted the classification of general G_2 holonomy spaces with $S^3 \times S^3$ principal orbits.

Until recently, the emphasis has been on finding new G_2 manifolds that are complete and non-singular. However, M-theory compactified on such spaces necessarily gives only Abelian and non-chiral $\mathcal{N} = 1$ theories in four dimensions. To obtain non-Abelian chiral theories from M-theory, one needs to consider compactifications on singular G_2 manifolds. One explicit realisation of such an M-theory compactification has an interpretation as an S^1 lift of type IIA theory, compactified on an orientifold, with intersecting D6-branes and O6 orientifold planes [62]. Non-Abelian gauge fields arise at the locations of coincident branes, and chiral matter arises at the intersections of D6-branes. Interestingly, these constructions provide [63] the first three-family supersymmetric standard-like models with intersecting D6-branes. The S^1 lift of these configurations results in singular G_2 holonomy metrics in M-theory. Co-dimension four ADE-type singularities are associated with the location of the coincident D6-branes, and co-dimension seven singularities are associated with the location of the intersection of two D6-branes in type IIA theory [56, 62, 64–66].

Further analyses of co-dimension seven singularities of the G_2 holonomy spaces, leading to chiral matter, were given in [64–66] and subsequent work [67–73]. It is expected that there exists a wide class of singular 7-manifolds with G_2 holonomy that yield non-Abelian $\mathcal{N} = 1$ supersymmetric four-dimensional theories with chiral matter. The explicit construction of such metrics would provide a starting point for further studies of chiral M-theory dynamics.

A recent study of an explicit class of singular G_2 holonomy spaces was given in [74]. These are cohomogeneity two metrics foliated by twistor spaces, that is S^2 bundles over self-dual Einstein four-dimensional manifolds M_4. The 4-manifold is chosen to be a self-dual Einstein space with orbifold singularities. An investigation of this construction was carried out in [68]. In [74] the most general self-dual Einstein metrics of triaxial Bianchi IX type, which have an $SU(2)$ isometry acting transitively on 3-dimensional orbits that are (locally) S^3, were considered.

Specialisation to biaxial solutions with positive cosmological constant yields manifolds that are compact, but in general with singularities. The radial coordinate ranges over an interval that terminates at endpoints where the $SU(2)$ principal orbits degenerate; to a point (a NUT) at one end, and to a two-dimensional surface (a bolt) that is (locally) S^2 at the other. Only for very special values of the NUT parameter is the metric regular at both ends. In general, however, one encounters singularities at both endpoints of the radial coordinate. In the generic case, a specific choice of the period for the azimuthal angle allows the singularity at the S^2 bolt to be removed, but then the NUT has a co-dimension four orbifold singularity. Alternatively, choosing the periodicity appropriate for regularity at the NUT, there will be a co-dimension two singularity on the S^2 bolt. The associated seven-dimensional G_2 holonomy spaces therefore have singularities of the same co-dimensions. The co-dimension four NUT singularities may admit an M-theory interpretation associated with the appearance of non-Abelian gauge symmetries, and the circle reduction of M-theory on these G_2 holonomy spaces may have a type IIA interpretation in terms of coincident D6-branes [68]. On the other hand, the co-dimension two singularities at the bolts do not seem to have a straightforward interpretation in M-theory.

Research is supported in part by DOE grant DE-FG02-95ER40893, NSF grant No. PHY99-07949, Class of 1965 Endowed Term Chair and NATO Collaborative research grant 976951 (M.C.), in full by DOE grant DE-FG02-95ER40899 (H.L.) and in part by DOE grant DE-FG03-95ER40917 (C.P.).

References

[1] I.R. Klebanov and M.J. Strassler, *JHEP* **0008** (2000) 052 [hep-th/0007191].

[2] M. Cvetič, H. Lü and C.N. Pope, *Nucl. Phys. B* **600** 103 (2001) [hep-th/0011023].

[3] M. Cvetič, G.W. Gibbons, H. Lü and C.N. Pope, Ricci-flat metrics, harmonic forms and brane resolutions [hep-th/0012011], to appear in *Comm. of Math. Phys.*

[4] M. Cvetič, G.W. Gibbons, H. Lü and C.N. Pope, *Nucl. Phys. B* **617** (2001) 151 [hep-th/0102185].

[5] M. Cvetič, G.W. Gibbons, H. Lü and C.N. Pope, *Nucl. Phys. B* **620** (2002) 29 [hep-th/0103155].

[6] M. Cvetič, G.W. Gibbons, H. Lü and C.N. Pope, *New cohomogeneity one metrics with Spin(7) holonomy* [math.DG/ 0105119].

[7] M. Cvetič, G.W. Gibbons, H. Lü and C.N. Pope, *Nucl. Phys. B* **606** (2001) 18 [hep-th/0101096].

[8] M. Cvetič, H. Lü and C.N. Pope, *Nucl. Phys. B* **613** (2001) 167 [hep-th/0105096].

[9] M. Cvetič, G.W. Gibbons, H. Lü and C.N. Pope, *Nucl. Phys. B* **620** (2002) 3 [hep-th/0106026].

[10] A. Brandhuber, J. Gomis, S.S. Gubser and S. Gukov, *Nucl. Phys. B* **611** (2001) 179 [hep-th/0106034].

[11] M. Cvetič, G.W. Gibbons, James T. Liu, H. Lü and C.N. Pope, *A New Fractional D2-brane, G_2 Holonomy and T-duality* [hep-th/0106162].

[12] H. Kanno and Y. Yasui, *On Spin(7) holonomy metric based on $SU(3)/U(1)$* [hep-th/0108226].

[13] M. Cvetič, G.W. Gibbons, H. Lü and C.N. Pope, *Cohomogeneity one manifolds of Spin(7) and $G(2)$ holonomy* [hep-th/0108245].

[14] S. Gukov and J. Sparks, *Nucl. Phys. B* **625** (2002) 3 [hep-th/0109025].

[15] M. Cvetič, G.W. Gibbons, H. Lü and C.N. Pope, *Orientifolds and slumps in G2 and Spin(7) metrics* [hep-th/0111096].

[16] G. Curio, B. Kors and D. Lust, *Fluxes and branes in type II vacua and M-theory geometry with $G(2)$ and Spin(7) holonomy* [hep-th/0111165].

[17] H. Kanno and Y. Yasui, *On Spin(7) holonomy metric based on $SU(3)/U(1)$. II* [hep-th/0111198].

[18] M. Cvetič, G.W. Gibbons, H. Lü and C.N. Pope, *Phys. Rev. Lett.* **8** (2002) 121602 [hep-th/0112098].

[19] A. Brandhuber, *Nucl. Phys. B* **629** (2002) 393 [hep-th/0112113].

[20] M. Cvetič, G.W. Gibbons, H. Lü and C.N. Pope, *Phys. Lett. B* **534** (2002) 172 [hep-th/0112138].

[21] Cvetič, G.W. Gibbons, H. Lü and C.N. Pope, *Almost special holonomy in type IIA and M theory* [hep-th/0203060].

[22] S. Gukov, S.T. Yau and E. Zaslow, *Duality and fibrations on $G(2)$ manifolds* [hep-th/0203217].

[23] Z.W. Chong, M. Cvetič, G.W. Gibbons, H. Lu, C.N. Pope and P. Wagner, *General metrics of $G(2)$ holonomy and contraction limits* [hep-th/0204064].

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

[25] S.S. Gubser, I.R. Klebanov and A.M. Polyakov, *Phys. Lett. B* **428** (1998) 105 [hep-th/9802109].

[26] E. Witten, *Adv. Theor. Math. Phys.* **2** (1998) 253 [hep-th/980215].

[27] K. Behrndt and M. Cvetič, *Nucl. Phys. B* **609** (2001) 183 [hep-th/0101007].

[28] A. Ceresole, G. Dall'Agata, R. Kallosh and A. Van Proeyen, *Phys. Rev. D* **64** (2001) 104006 [hep-th/0104056].

[29] K. Behrndt and G. Dall'Agata, *Nucl. Phys. B* **627** (2002) 357 [hep-th/0112136].

[30] M. Graña and J. Polchinski, *Phys. Rev. D* **63** (2001) 026001 [hep-th/0009211].

[31] J. Maldacena and C. Nuñez, *Int. J. Mod. Phys. A* **16** (2001) 822 [hep-th/0007018].

[32] S.S. Gubser, *Supersymmetry and F-theory realization of the deformed conifold with three-form flux* [hep-th/0010010].

[33] L.A. Pando Zayas and A.A. Tseytlin, *JHEP* **0011** (2000) 028 [hep-th/0010088].

[34] K. Becker and M. Becker, *JHEP* **0011** (2000) 029 [hep-th/0010282].

[35] M. Bertolini, P. Di Vecchia, M. Frau, A. Lerda, R. Marotta and I. Pesando, *JHEP* **0102** (2001) 014 [hep-th/0011077].

[36] O. Aharony, *JHEP* **0103** (2001) 012 [hep-th/0101013].

[37] E. Caceres and R. Hernandez, *Phys. Lett. B* **504** (2001) 64 [hep-th/0011204].

[38] J.P. Gauntlett, N. Kim and D. Waldram, *Phys. Rev. D* **63** (2001) 126001 [hep-th/0012195].

[39] I.R. Klebanov and E. Witten, *Nucl. Phys. B* **536** (1998) 199 [hep-th/9807080].

[40] S.S. Gubser and I.R. Klebanov, *Phys. Rev. D* **58**, (1998) 125025 [hep-th/9808075].

[41] I.R. Klebanov and N. Nekrasov, *Nucl. Phys. B* **574** (2000) 263 [hep-th/9911096].

[42] I.R. Klebanov and A.A. Tseytlin, *Nucl. Phys. B* **578** (2000) 123 [hep-th/0002159].

[43] T. Eguchi and A.J. Hanson, *Phys. Lett. B* **74** (1978) 249.

[44] K. Becker and M. Becker, *Nucl. Phys. B* **477** (1996) 155 [hep-th/9605053].

[45] M.J. Duff, J.M. Evans, R.R. Khuri, J.X. Lu and R. Minasian, *Phys. Lett. B* **412** (1997) 281 [hep-th/9706124].

[46] S.W. Hawking and M.M. Taylor-Robinson, *Phys. Rev. D* **58** (1998) 025006 [hep-th/9711042].

[47] K. Becker, *JHEP* **0105** (2001) 003 [hep-th/0011114].

[48] C.P. Herzog and I.R. Klebanov, *Phys. Rev. D* **63** (2001) 126005 [hep-th/0101020].

[49] C.P. Herzog and P. Ouyang, *Nucl. Phys. B* **610** (2001) 97 [hep-th/0104069].

[50] P. Herzog, I.R. Klebanov and P. Ouyang, *D-branes on the conifold and $N = 1$ gauge/gravity dualities* [hep-th/0205100], published in Les Houches 2001 proceedings.

[51] A.S. Dancer and A. Swann, *J. Geom. Phys.* **21** (1997) 218.

[52] R.L. Bryant and S. Salamon, *Duke Math. J.* **58** (1989) 829.

[53] G.W. Gibbons, D.N. Page and C.N. Pope, *Commun. Math. Phys.* **127** (1990) 529.

[54] M.B. Stenzel, *Manuscr. Math.* **80** (1993) 151.

[55] P. Candelas and X.C. de la Ossa, *Nucl. Phys. B* **342** (1990) 246.

[56] M. Atiyah and E. Witten, *M-theory dynamics on a manifold of $G(2)$ holonomy* [hep-th/0107177].

[57] B.S. Acharya, *On realising $N = 1$ super Yang-Mills in M theory* [hep-th/0011089].

[58] M. Atiyah, J. Maldacena and C. Vafa, *J. Math. Phys.* **42** (2001) 3209 [hep-th/0011256].

[59] J.D. Edelstein and C. Nuñez, *JHEP* **0104** (2001) 028 [hep-th/0103167].

[60] M. Aganagic and C. Vafa, *Mirror symmetry and a* $G(2)$ *flop* [`hep-th/0105225`].

[61] N. Hitchin, *Stable forms and special metrics* [`math.DG/0107101`].

[62] M. Cvetič, G. Shiu and A.M. Uranga, *Nucl. Phys. B* **615** (2001) 3 [`hep-th/0107166`].

[63] M. Cvetič, G. Shiu and A.M. Uranga, *Three-family supersymmetric standard like models from intersecting brane worlds* [`hep-th/0107143`]; *Phys. Rev. Lett.* **87** (2001) 201801 [`hep-th/0107143`].

[64] E. Witten, *Anomaly cancellation on* G_2 *manifold* [`hep-th/0108165`].

[65] B. Acharya and E. Witten, *Chiral fermions from manifolds of* G_2 *holonomy* [`hep-th/0109152`].

[66] M. Cvetič, G. Shiu and A.M. Uranga, *Chiral type II orientifold constructions as M-theory on* G_2 *holonomy spaces* [`hep-th/0111179`].

[67] R. Roiban, C. Romelsberger and J. Walcher, *Discrete torsion in singular* G_2*-manifolds and real* LG [`hep-th/0203272`].

[68] K. Behrndt, *Singular 7-manifolds with* G_2 *holonomy and intersecting 6-branes* [`hep-th/0204061`].

[69] A.M. Uranga, *Localized instabilities at conifolds* [`arXiv:hep-th/0204079`].

[70] L. Anguelova and C.I. Lazaroiu, *M-theory compactifications on certain "toric" cones of* G_2 *holonomy* [`hep-th/0204249`].

[71] L. Anguelova and C.I. Lazaroiu, *M-theory on "toric"* G_2 *cones and its type II reduction* [`hep-th/0205070`].

[72] P. Berglund and A. Brandhuber, *Matter from* G_2 *manifolds* [`hep-th/0205184`].

[73] R. Blumenhagen, V. Braun, B. Kors and D. Lust, *Orientifolds of K3 and Calabi-Yau Manifolds with Intersecting D-branes* [`hep-th/0206038`].

[74] M. Cvetič, G. Gibbons, H. Lü and C. Pope, *Bianchi IX self-dual metrics and* G_2 *manifolds*, UPR-1002-T, to appear.

SEMINAR 3

FOUR DIMENSIONAL NON-CRITICAL STRINGS

F. FERRARI

Joseph Henry Laboratories,
Princeton University,
Princeton NJ 08544, U.S.A.
and
Institut de Physique,
Université de Neuchâtel,
Rue A.-L. Bréguet 1, 2000 Neuchâtel,
Switzerland

Contents

FOUR DIMENSIONAL NON-CRITICAL STRINGS

F. Ferrari

Abstract

This is a set of lectures on the gauge/string duality and non-critical strings, with a particular emphasis on the discretized, or matrix model, approach. After a general discussion of various points of view, I describe the recent generalization to four dimensional non-critical (or five dimensional critical) string theories of the matrix model approach. This yields a fully non-perturbative and explicit definition of string theories with eight (or more) supercharges that are related to four dimensional CFTs and their relevant deformations. The space-time as well as world-sheet dimensions of the supersymmetry preserving world-sheet couplings are obtained. Exact formulas for the central charge of the space-time supersymmetry algebra as a function of these couplings are calculated. They include infinite series of string perturbative contributions as well as all the non-perturbative effects. An important insight on the gauge theory side is that instantons yield a non-trivial $1/N$ expansion at strong coupling, and generate open string contributions, in addition to the familiar closed strings from Feynman diagrams. We indicate various open problems and future directions of research.

1 Introduction

String theory is an unequalled subject for the extensive techniques that it uses and the scope of ideas on which it relies. This great variety certainly is at the origin of part of the excitement in the field. Yet, it pertains to the weakest point of the theory: the lack of unifying principles, and of a consistent non-perturbative definition on which to rely. It is thus essential to stand back and to strive to find a synthesis, if only a partial one. It

This work was supported in part by a Robert H. Dicke fellowship and by the Swiss National Science Foundation. The author is on leave of absence from Centre National de la Recherche Scientifique, École Normale Supérieure, Paris, France.

is with this motivation that we will discuss below the gauge theory/string theory duality. We will propose a point of view [1] that connects different approaches developed over the years. From this perspective, we are able to obtain several new results, including explicit non-perturbative definitions of many string theories and non-trivial exact formulas.

2 Many paths to the gauge/string duality

We start by discussing succinctly some of the many different facts that suggest a gauge theory/string theory duality.

2.1 Confinement

A compelling evidence for the relationship between ordinary four dimensional gauge theories like QCD and string theory is the similarity of their particle spectra. In both cases one expects to find an infinite set of resonances with masses on a Regge trajectory,

$$M_J^2 \sim \frac{J}{\alpha'}\,, \tag{2.1}$$

where M_J is the mass of the resonance, J its spin, and α' sets the length scales. In the string picture, $1/\alpha'$ is identified with the string tension. In the gauge theory picture, the dimensionless coupling constant $g^2_{\rm YM}$ is replaced by a scale Λ after dimensional transmutation, and $1/\alpha' \sim \Lambda^2$. Ordinary four dimensional Yang-Mills is believed to confine, and the spectrum (2.1) is characteristic of a confining gauge theory. Confinement is the consequence of the collimation of the chromoelectric flux lines, which generalize the ordinary Faraday flux lines of electrodynamics. The collimation can be demonstrated in the strong coupling approximation on the lattice [2], where it appears to be a consequence of the compactness of the gauge group. The string theory dual is simply the theory describing the dynamics of the tubes that the collimated flux lines form. The collimation of flux lines is interpreted in terms of a dual superconductor picture [3]. The idea is that the relevant degrees of freedom in the strongly coupled Yang-Mills theory are magnetically charged. One then assumes that those magnetic charges condense, which implies that the chromoelectric flux is squeezed into vortices, in the same way as the condensation of electric charges squeezes the magnetic field into Abrikosov vortices in an ordinary superconductor.

The chromoelectric flux vortices, or tubes, have a definite thickness of order $1/\Lambda$. The above description in terms of strings thus seems to be at best phenomenological. The well-studied fundamental strings [4] indeed have zero thickness. Equivalently, one would need a fundamental description of relativistic theories based on light electric and magnetic charges, and none

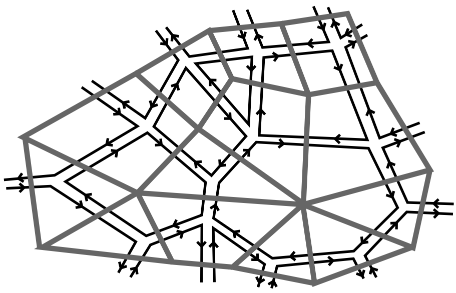


Fig. 1. Part of a typical Feynman diagram in an $SU(N)$ gauge theory, depicted in the double line representation. A dual representation (gray lines) is obtained by associating a p-gon to each vertex of order p. The p-gons generate an oriented discretized Riemann surface. The power of $1/N^2$ counts the genus of these surfaces.

were known until the proposal in [1]. Those difficulties explain why the excitement initiated in the late sixties eventually faded, and the subject remained dormant for several decades.

2.2 Large N

A seemingly very different argument in favor of the gauge theory/string theory duality is due to 't Hooft [5]. In an $SU(N)$ gauge theory, with fields transforming in the adjoint representation, the Feynman graphs can be depicted using a double-line representation corresponding to the double-index notation $A^m_{\bar{n}}$ for the fields. This representation makes the relationship with discretized Riemann surfaces obvious (see Fig. 1). Moreover, by taking $g^2_{\mathrm{YM}}N$ (or equivalently after renormalization the mass scale Λ) to be an N-independent constant, surfaces of genus g comes with a factor $1/N^{2g}$. The large N expansion of $SU(N)$ Yang-Mills theory is thus a reordering of the Feynman diagrams with respect to their topology. This shows that a perturbative discretized closed oriented string theory of coupling constant $\kappa \sim 1/N$ is equivalent to the reordered perturbative Yang-Mills theory. By perturbative, we mean that the correspondence works *a priori* for the contributions to the Yang-Mills path integral that can be represented in terms of Feynman diagrams. Note that the world-sheet of the string theory is embedded in the four dimensional Minkowski space, since the real-space Feynman rules imply that each vertex comes with a space-time label on which we integrate.

The weakness of the above argument is obvious: the discretized world sheets are only vaguely reminiscent of the continuous world-sheets of

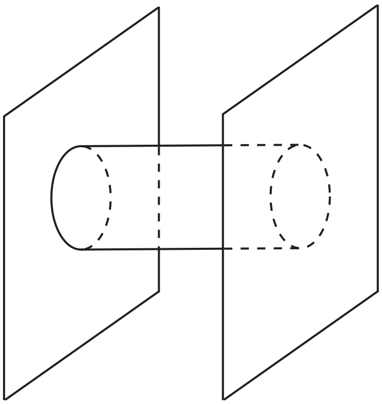


Fig. 2. Interaction between D-branes. One can interpret the picture as either a one-loop diagram for an open string whose ends are stuck on the branes (this is described by a Yang-Mills theory), or as a tree diagram for a closed string propagating in the bulk (this is described by a quantum gravity theory).

ordinary string theory. Only very large Feynman graphs, which have a very large number of faces, may give a good approximation to continuous world sheets. One could then argue that in the IR, which is the relevant regime for confinement, the gauge coupling is large and thus the large Feynman graphs indeed dominate. This picture is only heuristic and might at best yield an effective string theory description, similar to the one discussed in Section 2.1. We will see however that it is extremely fruitful to pursue this idea further. In a slightly different context, and with an important additional physical input, we will be able to use the 't Hooft representation in a controlled way.

2.3 D-branes

Yet another argument in favor of the gauge/string correspondence has been through the use of D-branes [6]. A Dp-brane can be described perturbatively as a $p+1$ dimensional defect in space-time on which open strings can end [7]. When N D3-branes are put on top of each other, the low energy dynamics of the open strings is described by the $\mathcal{N}=4$ supersymmetric Yang-Mills theory with gauge group $SU(N)$. On the other hand, at large N, the whole of the N D3-branes is an heavy object that can be described by a classical solitonic solution in type IIB supergravity [8]. These complementary descriptions of the D-branes (see Fig. 2), involving either open strings or a soliton in a closed string theory, is equivalent to the gauge/string duality.

This picture can be made precise, and we refer the reader to other lectures at this school [9] in which many details about this correspondence can be found.

The above construction fits very well with the discussion of Section 2.2. The closed string coupling constant indeed turns out to be proportional to $1/N$, and $g^2_{\rm YM}N$ is independent of this coupling. However, it seems at first sight impossible to make it consistent with the discussion of Section 2.1. The $\mathcal{N}=4$ theory is indeed a conformal field theory, and thus does not confine. There is then no length scale to set the string tension. Moreover, the correspondence involves type IIB strings, which are of zero thickness and live in ten dimensions! Remarkably, it is the very fact that more than four dimensions are involved that makes the use of fundamental strings and the description of conformal gauge theories possible. Of the six additional dimensions, five play a rather technical and model-dependent role. They are best viewed as additional degrees of freedom on the world sheet that are necessary to account for the many "matter" fields of the $\mathcal{N}=4$ super Yang-Mills theory. Those fields transform under an $SO(6)$ R-symmetry, which explains why the five dimensions actually form a 5-sphere. The other, "fifth" dimension ϕ, is much more interesting. The form of the non-trivial five dimensional metric is actually fixed by conformal invariance to be the AdS_5 metric,

$$\mathrm{d}s_5^2 = \mathrm{d}\phi^2 + \mathrm{e}^{2\phi/R}\,\mathrm{d}s_4^2\,, \tag{2.2}$$

where $\mathrm{d}s_4^2$ is the four dimensional Minkowski metric. The radius R turns out to be proportional to $(g^2_{\rm YM}N)^{1/4}$. From (2.2) we see explicitly that ϕ can be interpreted as a renormalization group flow parameter, since a shift in ϕ can be absorbed in a rescaling of the Minkowski metric. The string tension varies in the fifth dimension and is set by the RG scale! The original open strings that generate the Yang-Mills dynamics are naturally attached to the brane in the far UV, $\phi\rightarrow\infty$, where all the information about the gauge theory is encoded. Confining theories can be obtained by adding some relevant operators to the $\mathcal{N}=4$ Yang-Mills action. The form of the metric is then in general

$$\mathrm{d}s_5^2 = \mathrm{d}\phi^2 + a^2(\phi)\mathrm{d}s_4^2\,, \tag{2.3}$$

where the function $a(\phi)$ is well approximated by $\exp(\phi/R)$ in the UV region $\phi\rightarrow\infty$, but remains to be determined in the IR. At the confining scale $\phi=\phi_*\sim\ln\Lambda$, $a(\phi)$ will typically have a minimum. It is then energetically favoured for the fundamental strings to sit in this region, and this implies that they acquire an effective thickness $1/\Lambda$ from the four dimensional point of view [10].

The three aspects discussed in Sections 2.1–2.3 are thus fully consistent, in a subtle and interesting way. We will better understand the origin of the fifth dimension in the next subsection, but the fact that conformal field theories play a prominent role remains a rather weird and not very well understood feature. The discussion in Section 3 will shed some light on this part of the story.

2.4 Non-critical strings

The most direct approach to construct a string theory dual to a four dimensional field theory is to try to quantize directly the string with a target space of dimension $D = 4$. Because of the quantum anomaly in the Weyl symmetry, the world sheet metric g_{ab}, which classically is a mere auxiliary field, does not decouple. Using world sheet diffeomorphisms, we can always put the metric in the form $g = \hat{g}\mathrm{e}^{\phi}$, where $\hat{g}$ depends on a finite number of moduli. The conformal factor ϕ, which is called the Liouville field, then acquires a non-trivial world sheet dynamics. When $D \leq 1$, or more precisely for a world-sheet theory of central charge $c \leq 1$, it is possible, under rather well-controlled assumptions, to work out this dynamics explicitly [11]. For example, in the case $D = 1$, the classical world sheet action is

$$S = \frac{1}{4\pi\alpha'} \int \mathrm{d}^2\sigma \sqrt{g}\, (\partial_a X \partial^a X + \lambda_0 + \alpha' \Phi R) \ . \tag{2.4}$$

The world sheet scalar X corresponds to the embedding coordinate and the string coupling constant is $g_{\mathrm{s}0} = \mathrm{e}^{\Phi}$. A world sheet cosmological constant λ_0 must also be included because the Weyl symmetry is broken quantum mechanically. Integrating out g then yields an action of the form

$$S = \frac{1}{4\pi\alpha'} \int \mathrm{d}^2\sigma \sqrt{\hat{g}} \left(\partial_a X \partial^a X + \partial_a \phi \partial^a \phi + T(\phi) + \alpha' \Phi(\phi) \hat{R} \right) , \tag{2.5}$$

which shows that the Liouville field ϕ plays the role of a new dimension. The physics is not uniform in this new dimension because of the non-trivial "tachyon" T and dilaton Φ backgrounds, which are determined by requiring world sheet conformal invariance. More details may be found in the review [12]. Unfortunately, the case $D > 1$ is much more difficult and a direct analysis has never been performed. However, it can be noted [13] that the most general ansätz compatible with the symmetries of the problem, in particular with D-dimensional Lorentz invariance, is

$$S = \frac{1}{4\pi\alpha'} \int \mathrm{d}^2\sigma \sqrt{\hat{g}}\, \left(a^2(\phi) \partial_a X_\mu \partial^a X^\mu + \partial_a \phi \partial^a \phi + \cdots \right) , \tag{2.6}$$

where the dots represent various possible background fields. The main difference with the $D \leq 1$ case is that the $D + 1$ dimensional metric is not flat. When $D = 4$, the Liouville field is but the fifth dimension discussed in Section 2.3, and we recover in particular equation (2.3). For the string theory (2.6) to describe a gauge theory, additional consistency conditions must be satisfied. In particular, the open strings generating the Yang-Mills dynamics must live either at the horizon $a = 0$ or at the point of infinite string tension $a = \infty$. The latter choice, that corresponds to a "fundamental" brane living in the far UV, seems more natural and is consistent with the discussion of Section 2.3. In either case, the open string theory will have only vector states and the zig-zag symmetry of the Wilson loops will be satisfied [13].

String theories based on actions like (2.6) are extremely difficult to solve. Even in the simplest case $D \leq 1$ described by (2.5), only a partial analysis can be given. The reason is twofold: first the world sheet theories are complicated interacting two dimensional field theories, which makes the analysis of string perturbation theory particularly involved; second the string coupling can grow due to the non-trivial dilaton background, which can altogether invalidate the use of the string perturbative framework. In spite of these daunting difficulties, the case $D \leq 1$ has actually been solved to all orders of string perturbation theory in a series of remarkable papers [14–17]. The basic idea [15] is to consider a discretized version of the D dimensional string theory, which turns out quite surprisingly to be easier to study than the original continuum model. The continuous world sheets are approximated by discretized surfaces made up of flat polygons of area ℓ^2. The curvature is concentrated on a discrete lattice L on which the vertices of the polygons lie. The discrete world sheet fields are defined at the center of the polygons, or equivalently on the dual lattice $\tilde{L}$. For example, the discrete Polyakov path integral $Z_{\mathrm{P}}(\ell)$ defining string perturbation theory based on the action (2.4) with the cut-off $1/\ell$ can be written

$$\begin{aligned} Z_{\mathrm{P}}(a) &= \sum_{h\geq 0} \int_{\substack{\text{genus } h \\ \text{world sheets}}} [\mathrm{d}g_{ab}(\sigma)\mathrm{d}X(\sigma)]\, \mathrm{e}^{-S} \Big|_{\text{cut-off } 1/\ell} \\ &= \sum_{h\geq 0} g_{\mathrm{s}0}^{2h-2} \sum_{\substack{\text{genus } h \\ \text{lattices } L}} \mathrm{e}^{-\frac{\lambda_0 \ell^2 |\tilde{L}|}{4\pi\alpha'}} \int \prod_{i\in\tilde{L}} \mathrm{d}X_i \prod_{\substack{\text{links } \langle kl\rangle \\ \text{of } \tilde{L}}} \Delta\left(|X_k - X_l|/\sqrt{\alpha'}\right). \end{aligned} \tag{2.7}$$

In the above formula $|\tilde{L}|$ represents the number of vertices of the lattice $\tilde{L}$, or equivalently the number of polygons in the discretization of the world sheet. The function Δ is a Gaussian and is derived from the kinetic term for X in (2.4). The cosmological constant λ_0 can *a priori* be renormalized, since the sum over lattices of fixed size and the integration over the X_is can generate a counterterm proportional to $|\tilde{L}|$.

We can now use 't Hooft's idea described in Section 2.2. The very involved combinatorial problem corresponding to the discrete lattice sum (2.7) is conveniently encoded in the Feynman graph expansion of a matrix theory. Originally, 't Hooft considered four dimensional Yang-Mills theories, but the argument is straightforwardly extended to any matrix field theory. If the matrix field theory lives in D dimension(s), then the discretized world sheets are embedded in the D dimensional Minkowski space. It is easy to see that the matrix theory corresponding to (2.7) is a quantum mechanics based on a single $N \times N$ hermitian matrix M, and that we have

$$\mathrm{e}^{Z_{\mathrm{P}}(\ell)} = \int \left[\mathrm{d}^{N^2} M(\tau)\right] \, \exp\left[-N \int \mathrm{d}\tau \; \mathrm{tr}\Big(\frac{1}{2}\dot{M}^2 + \frac{1}{2\alpha'} M^2 + U(M)\Big)\right] . \tag{2.8}$$

The euclidean time τ corresponds to the embedding coordinate X in the string theory. Terms in M^p in the potential $U(M)$ generate p-gons in the discretization of the world sheets, see Figure 1. The metric on the world sheet is determined by giving an area ℓ^2 to the p-gons. We can actually restrict ourselves to the simple potential

$$U(M) = \frac{g}{3!} M^3 . \tag{2.9}$$

More complicated potentials yield the same theory in the continuum limit. In terms of the matrix theory variables, we have

$$Z_{\mathrm{P}}(\ell) = \sum_{h \geq 0} N^{2-2h} \sum_{\substack{\text{genus } h \\ \text{lattices } \tilde{L}}} (-g)^{|\tilde{L}|} \int \prod_{i \in \tilde{L}} \mathrm{d}\tau_i \prod_{\substack{\text{links } \langle kl \rangle \\ \text{of } \tilde{L}}} \Delta_{\mathrm{E}}\left(|\tau_k - \tau_l|/\sqrt{\alpha'}\right) . \tag{2.10}$$

The power of N comes from the standard 't Hooft's analysis, and yields a string coupling $g_{\mathrm{s}0} = 1/N$. The power $|\tilde{L}|$ of g corresponds to the number of vertices in the ordinary representation of the Feynman diagrams, or equivalently to the number of polygons in the dual representation, see Figure 1. The sum is over the lattices $\tilde{L}$, which is of course equivalent to the sum over the dual lattices L as in (2.7).

The matrix quantum mechanics (2.8) can be mapped onto a problem of free fermions [14], and is thus exactly solvable. In particular, the discrete Polyakov partition function, which is proportional to the radius r of the embedding dimension due to translational invariance, can be calculated since it is related to the ground state energy $E(g)$,

$$E(g) = - \lim_{r \to \infty} \frac{Z_{\mathrm{P}}(\ell)}{r} \cdot \tag{2.11}$$

Any other observable of the discrete string theory could be related to quantum mechanical amplitudes in a similar way. However, and as was stressed in Section 2.2, we are still far from our goal of solving a continuum string theory. Indeed, we must find a way to implement consistently the continuum limit $\ell \to 0$. At finite ℓ, the identification of (2.7) and (2.10) must be taken with a grain of salt, because the link factor Δ_{E} is the one dimensional euclidean propagator, which is a simple exponential, unlike the Gaussian Δ that corresponds to the Polyakov action. On the other hand, it is very plausible that any link factor for which a consistent continuum limit can be defined will yield the unique consistent continuum string theory.

The fact that a continuum limit can be defined for (2.10) relies on the non-trivial property that the average number of polygons in the relevant Feynman graphs for (2.8) diverges for a negative critical value of the coupling g [14, 15]. Indeed, the ground state energy admits an expansion

$$E(g) = \sum_{h\geq 0} N^{2-2h} E_h(g) = \sum_{h\geq 0} g_{\mathrm{s}0}^{2h-2} E_h(g)\,, \tag{2.12}$$

where

$$E_h(g) = \sum_{k=0}^{\infty} E_{h,k}(-g)^k\,. \tag{2.13}$$

The numerical coefficient $E_{h,k} < 0$ gives the contribution from genus h surfaces with a fixed number k of polygons. It is given by the terms in (2.10) with $|\tilde{L}| = k$. The series (2.13) has a finite radius of convergence which is independent of h, and the critical coupling corresponds to the point where it diverges [14]. This means that the large N expansion of the matrix quantum mechanics breaks down at $g = g_{\mathrm{c}}$. When $g \to g_{\mathrm{c}}$, the terms with a high power of k, or equivalently the surfaces with a large number of polygons, dominate the sum (2.13). *The limit $g \to g_{\mathrm{c}}$ is thus a continuum limit in which 't Hooft's heuristic picture of Section 2.2 becomes precise.* At large k, $E_{h,k}$ picks up a term proportional to $(-g_{\mathrm{c}})^{-k}$, and thus we expect the contribution of surfaces of size $|\tilde{L}|$ to be proportional to $(g/g_{\mathrm{c}})^{|\tilde{L}|} \simeq \exp\bigl(-|\tilde{L}|(g_{\mathrm{c}} - g)\bigr)$. There is actually a logarithmic correction to that formula, and $g_{\mathrm{c}} - g$ is replaced by ϵ defined by

$$g_{\mathrm{c}} - g = -\epsilon \ln \epsilon\,. \tag{2.14}$$

Comparing with (2.7), we get the renormalized world sheet cosmological constant

$$\lambda = 4\pi\alpha' \frac{\epsilon}{\ell^2}\,\cdot \tag{2.15}$$

This equation gives the precise relation between the coupling g of the matrix quantum mechanics and the world sheet cut-off $1/\ell$. In the continuum limit, equation (2.12) naïvely suggests that $E(g)$ diverges, since the fixed genus contributions $E_h(g)$ themselves diverge. However, the limit can still be made consistent because the divergences are very specific, and can be compensated for by a simple multiplicative renormalization of the string coupling g_{s0} [15–17]. Indeed, we have

$$E_h(g) \underset{g \to g_c}{\propto} \frac{1}{\epsilon^{2h-2}}, \tag{2.16}$$

which shows that in the double scaling limit [16, 17]

$$N \to \infty, \quad \epsilon \to 0, \quad N\epsilon = \text{constant} = 1/\kappa, \tag{2.17}$$

the ground state energy (2.12) has a finite limit E_{scaled}, with an expansion of the form[1]

$$E_{\text{scaled}} = \sum_{h \geq 0} E_h\, \kappa^{2h-2}. \tag{2.18}$$

The double scaling limit (2.17) shows that the renormalized string coupling is

$$g_s = \frac{1}{N\ell^2} = \frac{g_{s0}}{\ell^2}, \tag{2.19}$$

and thus have a non-trivial world sheet dimension two. The dimensionless coupling κ is a combination of the genuine string coupling and of the cosmological constant,

$$\kappa = \frac{4\pi\alpha' g_s}{\lambda}. \tag{2.20}$$

The non-critical strings based on a world sheet theory of central charge $c < 1$ can be treated in a similar way. The matrix quantum mechanics (2.8) is replaced by a simple matrix integral

$$\int \mathrm{d}^{N^2} M\, \mathrm{e}^{-N\,\mathrm{tr}\, U(M; g_j)} \tag{2.21}$$

with a general potential

$$U(M; g_j) = \frac{1}{2} M^2 + \sum_{j=3}^{p} \frac{g_j}{j!} M^j. \tag{2.22}$$

[1]The sphere and torus contributions are actually logarithmically divergent. This can be understood from the point of view of the continuum theory, but is beyond the scope of the present review. Details can be found in the excellent lecture notes [18].

By adjusting k of the couplings to special values, we can go to a kth-order critical point, called the kth Kazakov critical point, for which $c = 1 - 3(2k - 3)^2/(2k-1)$. The double scaling limit (2.17) takes the general form

$$N \to \infty\,, \quad \epsilon \to 0\,, \quad N\epsilon^{1-\gamma_{\rm str}/2} = \text{constant}, \tag{2.23}$$

with some exponent $\gamma_{\rm str} = -1/(k+1)$, but qualitatively the results are very similar to the $c = 1$ case. All the (p, q) minimal CFTs can be obtained in this way, by considering a slight generalization of (2.21) based on a two-matrix model. All the correlators can be studied, and the RG flows between the various theories are described by a nice mathematical structure based on the KdV hierarchy. Details can be found in [19].

Let us conclude this introduction with an important remark. The simple potential (2.9) does not have a ground state. This means that $E(g)$ can only be defined as a sum over Feynman graphs. We could try to use another potential, for example $U(M) = gM^4/4!$, but the positivity of the partition function (2.7) implies that the coupling g in (2.10) must be negative. As explained in details in Section 7 of the review [19], it is actually impossible to obtain a non-perturbative definition of the unitary string theories using the matrix models. The "solutions" of the $D \le 1$ string theories thus yield the observables to all orders of perturbation theory, but not beyond. Of course, the Polyakov formulation of string theory, that was our starting point, is perturbative in nature. Equations like (2.11) are thus perfectly consistent and must be understood as a statement about the asymptotic perturbative series. However, it was originally expected that the powerful new formulation in terms of a matrix theory in the double scaling limit could give insights into a non-perturbative definition of string theory. The fact that this is not the case is a major drawback of the classic matrix model approach. We will have much more to say in Section 3 about this very important point of principle.

3 Four dimensional non-critical strings

The discretized approach advocated in [15] is indisputably the most fruitful to study the $D \le 1$ string theories (2.5). It is then most natural to try to extend the same ideas to the $D > 1$ theories (2.6). The first concrete results in this direction were obtained only recently by the present author [1]. The results of [1], together with the insights gained in a series of preparatory [20, 21] and subsequent [22, 23] works, strongly suggest that the discretized approach in $D > 1$ (and in particular in the case $D = 4$ on which we will focus) has considerable power. In particular, it provides an explicit *non-perturbative* definition of the $D = 4$ non-critical strings, in sharp contrast with the $D \le 1$ cases.

Table 1. The magic square of the non-critical string theories. The continuum approach side of the $c = 1$ barrier was finally cleared in 1997, but the discretized approach side remained unviolated until recently. We will try to shed some light on the material hidden behind the question mark in the following.

Continuous world sheets with $c \leq 1$ $$Z = \int [\mathrm{d}g_{ab}(\sigma)\mathrm{d}X(\sigma)]\,\mathrm{e}^{-S}$$ $c+1$ dimensional flat target space $$\mathrm{d}s^2 = \mathrm{d}\phi^2 + \mathrm{d}X^2$$	Discretized world sheets with $c \leq 1$ $$\mathrm{e}^{Z(\ell)} = \int \mathrm{d}^{N^2}M\,\mathrm{e}^{-N\,\mathrm{tr}\,U(M;g_j)}$$ Matrix integral with Kazakov critical points in the double scaling limit
Continuous world sheets with $c > 1$ $$Z = \int [\mathrm{d}g_{ab}(\sigma)\mathrm{d}\vec{X}(\sigma)]\,\mathrm{e}^{-S}$$ $c+1$ dimensional curved target space $$\mathrm{d}s^2 = \mathrm{d}\phi^2 + a^2(\phi)\,\mathrm{d}\vec{X}^2$$	Discretized world sheets with $c > 1$?

The first step to go from $D \leq 1$ to $D = 4$ is straightforward: the $D = 0$ matrix integrals (2.21) or $D = 1$ matrix path integral (2.8) are replaced by $D = 4$ matrix path integrals

$$\int \left[\mathrm{d}M(x)\right]\,\exp\left[-N\int \mathrm{d}^4x\,\mathcal{L}(M,\partial M)\right], \tag{3.1}$$

where M represents in general a collection of $N \times N$ hermitian matrices and $\mathcal{L}$ is a lagrangian density. In the original 't Hooft's example [5], we have four matrices corresponding to the four components of the vector potential A_μ, and $\mathcal{L}$ is the $SU(N)$ Yang-Mills lagrangian with the 't Hooft coupling $g_{\mathrm{YM}}^2 N$ or equivalently after renormalization the dynamically generated scale Λ chosen to be independent of N. The path integral (3.1) defines a discretized string theory with a four dimensional target space[2]. Of course, and as stressed in Section 2.2, the real challenge is to succeed in taking the continuum limit. This was done in Section 2.4 for the simple $D \leq 1$ integrals (2.8) or (2.21) by adjusting the parameters in the interaction potential to special Kazakov critical points. However, the pure Yang-Mills path integral does not have any free parameter, and thus this procedure cannot be straightforwardly extended.

The way out of this problem was first proposed in [20]. The idea is to consider gauge theories with Higgs fields in the adjoint representation of the gauge group. The adjoint Higgs fields correspond to additional hermitian matrices M on which we integrate in (3.1). The Higgs theory depends on couplings parametrizing the Higgs potential. Varying those couplings amounts to varying the masses of the gauge bosons, or equivalently the effective gauge coupling. It was then argued in [20] that for some special values of the Higgs couplings, that correspond to W bosons of masses of order Λ, critical points may exist. It could then be possible to define a consistent continuum limit, in strict parallel to the $D \leq 1$ cases. The consistency of those ideas were checked on a simple toy model [20]. Supersymmetry plays no role in the discussion, and in particular the model studied in [20] was purely bosonic. Unfortunately, and even though lattice calculations seem encouraging [24], the present-day analytical tools do not make the search for critical points in purely bosonic Yang-Mills/adjoint Higgs theories possible.

This is where supersymmetry enters the game: we do have some control on supersymmetric Yang-Mills/Higgs theories, and we can try to apply our ideas in this context. A typical example is pure $\mathcal{N} = 2$ supersymmetric $SU(N)$ gauge theory. The adjoint Higgs field is automatically included in this theory because it is a supersymmetric partner of the gauge bosons. Strictly speaking, this theory is parameter-free, because the Higgs coupling is related to the gauge coupling by supersymmetry, and the latter is replaced by a mass scale in the quantum theory. However, there is a freedom in the choice of the Higgs expectation values, because the Higgs potential has flat directions that are protected by a non-renormalization theorem. The path integral (3.1) is then parametrized by a set of boundary conditions for the Higgs field at infinity. Such parameters are called moduli of the field theory, and one can show that there are $N-1$ moduli for the pure $SU(N)$ theory. For our purposes, the moduli will play the role of the couplings g_j in the

[2]Considering $SU(N)$ instead of $U(N)$ amounts to integrating over traceless hermitian matrices in (3.1). This does not change the discretized string interpretation in the large N limit.

potential (2.22) of the $D \leq 1$ theories. In more general $\mathcal{N} = 2$ gauge theories, one may have quark mass parameters in addition to the moduli. Mass parameters and moduli are on an equal footing in our discussion, and we will denote them collectively by $\mathcal{M} = \{m_j\}$.

There is a subtle but fundamental difference between the space of couplings g_j for the simple matrix integrals of Section 2 and the space $\mathcal{M}$ for the gauge theories: the Yang-Mills path integrals are non-perturbatively defined for *all* values of the moduli or parameters, unlike the matrix integrals in $D \leq 1$. As discussed at the end of Section 2.4, this has some important consequences, because the Kazakov critical points in $D \leq 1$ always lie at the "wrong" couplings, for example $g < 0$ for the integral (2.8) with a quartic potential. As a consequence, the $D \leq 1$ continuum string theories can only be defined in perturbation theory. We see that in four dimensions, we cannot run into this problem: we can only get *non-perturbative* definitions of string theories, from the *non-perturbative* gauge theory path integrals. The hard part is of course to find Kazakov critical points on $\mathcal{M}$.

3.1 Four dimensional CFTs as Kazakov critical points

It is well-known that there are so-called singularities on the moduli/parameter space $\mathcal{M}$ of $\mathcal{N} = 2$ gauge theories where a non-trivial low energy physics develops [25]. For example, in the pure $SU(N)$ case, one can adjust $p \geq 2$ moduli to special values and get an interacting CFT in the infrared [26]. The nature of such CFTs has remained rather mysterious, because the light degrees of freedom include both electric and magnetic charges, and thus a conventional local field theoretic description does not exist. Such theories are nevertheless of primary interest, because effective theories of light electric and magnetic charges are believed to play an important role in real world QCD, as explained at the end of Section 2.1.

The main result of our investigations is then the following:
The non-trivial CFTs on the moduli space of supersymmetric gauge theories, that can be obtained by adjusting a finite, N-independent, number of moduli, are the four dimensional generalizations of the Kazakov critical points. They can be used to define double scaling limits that yield string theories dual to the corresponding CFT with the possible relevant deformations.

A direct consequence of this claim is that the theories of light electric and magnetic charges can be described at a fundamental level by string theories, a result in perfect harmony with the discussion in Section 2. Moreover, many CFTs, including gauge CFTs, can be constructed in this way. For example, $\mathcal{N} = 4$ super Yang-Mills with gauge group $SU(p+1)$ can be obtained on the moduli space of a parent $\mathcal{N} = 4$ theory with gauge group $SU(N)$, by adjusting p moduli. The very possibility of defining a double scaling limit automatically demonstrates that there is a string dual, whose continuous world sheets are constructed from the Feynman diagrams of the *parent* gauge theory.

A basic property of a Kazakov critical point is that the large N expansion of the matrix theory breaks down in its vicinity. This is due to divergences at each order in $1/N$. For example, the divergence of the coefficient E_h defined in (2.12) is given by (2.16). Those divergences are crucial, since the whole idea of the double scaling limits rely on the possibility of compensating the divergences by taking $N \rightarrow \infty$ and approaching the critical point in a correlated way. This procedure picks up the most divergent, universal, terms that correspond to the continuum string theory. A basic consequence of our claim is thus that the large N expansion of $\mathcal{N} = 2$ supersymmetric gauge theories should break down at singularities on the moduli space! More precisely, one can distinguish two classes of singularities: those that are obtained by adjusting a large number of order N of moduli, and others that are obtained by adjusting a finite, N-independent number of moduli. Our claim implies that divergences in the large N expansion must be found in the second case. This is by itself a rather strong and new statement about the behaviour of the large N expansion of certain gauge theories. If the $\mathcal{N} = 2$ parent gauge theory admits a string dual (not to be confused with the string theories produced in the double scaling limits!), it implies that the string theory does not admit a well-defined perturbation theory at the critical points.

The breakdown of the $1/N$ expansion can be given a simple interpretation. Commonly, trying to find a good approximation scheme to describe a non-trivial critical point is difficult. A typical example is ϕ^4 theory in dimension D. The theory has two parameters, the bare mass m (or "temperature") and the bare coupling constant g. By adjusting the temperature, we can go to a point where we have massless degrees of freedom, and a non-trivial Ising CFT in the IR. The difficulty is that the renormalized fixed point coupling g_* is large, and thus ordinary perturbation theory in g fails. It is meaningless to try to calculate universal quantities like critical exponents as power series in g, since those are g-independent. Either the tree-level, g-independent contributions are exact and the corrections vanish (this occurs above the critical dimension, which is $D_{\mathrm{c}} = 4$ for the Ising model, and we have a trivial fixed point $g_* = 0$ well described by mean field theory), or the expansion parameter corresponds to a relevant operator and corrections to mean field theory are plagued by untamable IR divergencies. The critical points on the moduli space of $\mathcal{N} = 2$ super Yang-Mills are very similar to the Ising critical point below the critical dimension, and $1/N$ is very similar to the ϕ^4 coupling g. They are characterized by a set of critical exponents [26] that are independent of the number of colours N of the parent gauge theory in which the CFT is embedded. These critical exponents cannot consistently be calculated in a $1/N$ expansion. Even the simple monopole critical points that are known to be trivial are not described consistently by the $N \to \infty$ limit of the original gauge theory, because electric-magnetic duality is not implemented naturally in this approximation scheme. Note that the argument does not apply to CFTs obtained by adjusting a large number $\sim N$ of moduli, because those are N-dependent, and the large N expansion can then certainly be consistent.

The fact that the divergences have an IR origin implies that the string theories obtained in the double scaling limits are dual to the relevant deformations of the CFT at the critical point. Indeed, the scaling limits pick up the most IR divergent contributions, which are due to the light degrees of freedom only. This can be checked explicitly [1, 22], as we will see in Section 3.4.

Our arguments so far have been qualitative, and the reader may feel rather uncomfortable. Indeed, it is at first sight hard to imagine how to test our ideas by explicit calculations [27]. For example, one would like to compute explicitly some observables in the $1/N$ expansion, and check explicitly that the large N expansion does break down at the critical point, and that the divergences are specific enough for a consistent double scaling limit to be defined. We explain below how such calculations can actually be done. This involves a rather surprising result about the large N behaviour of instanton contributions at strong coupling.

3.2 Instantons and large N

Unlike the $D \leq 1$ matrix models, which are exactly solvable, only a small number of observables can be calculated in $\mathcal{N} = 2$ gauge theories, following Seiberg and Witten [25, 28]. Those observables are physically very important, since they correspond to the leading term in a derivative expansion of the low energy effective action. In particular, the central charge of the supersymmetry algebra, and thus the mass of the BPS states, can be calculated exactly by using Gauss' law. However, those observables are very special mathematically, because they pick up only a one-loop term from ordinary perturbation theory. The non-trivial physics comes entirely from an infinite series of instanton contributions.

Relevant contributions in the large N limit are usually assumed to come from the sum of the Feynman graphs at each order in $1/N^2$ [5], as reviewed in Section 2. On the other hand, instantons are usually disregarded [29]. This is due to the fact that the instanton action is proportional to N in the 't Hooft's scaling $g^2_{\mathrm{YM}} \propto 1/N$. The effects of instantons of topological charge k and size $1/v$ are thus proportional, in the one-loop approximation which is exact for $\mathcal{N}=2$ super Yang-Mills, to

$$\mathrm{e}^{-8\pi^2 k/g_{\mathrm{YM}}^2} = \left(\frac{\Lambda}{v}\right)^{kN\beta_0}, \tag{3.2}$$

where β_0 is a coefficient of order 1 given by the one-loop β function. This formula suggests that the only smooth limit of instanton contributions when $N\rightarrow\infty$ is zero, with exponentially suppressed corrections [29]. Large instantons (small v), if relevant, would produce catastrophic exponentially large contributions, and if one is willing to assume that the large N limit makes sense the only physically sensible conclusion seems to be that instantons are irrelevant variables. This argument is independent of supersymmetry. In real-world QCD, the ratio Λ/v in (3.2) would simply be replaced by a more complicated function. In QCD, instantons of all sizes (all v) can potentially contribute, and this led Witten to argue that the instanton gas must vanish [29]. In Higgs theories, like $\mathcal{N}=2$ super Yang-Mills, the Higgs vevs introduce a natural cutoff v on the size of instantons, and for v large enough ("weak coupling") the instanton gas can exist but is just negligible at large N. At small v ("strong coupling"), we run into the same difficulties as in QCD.

The above argument would seem to imply, at least superficially, that the Seiberg-Witten observables are not suited for testing our ideas on the nontrivial behaviour of the large N expansion. However, it was realized in [21] that there is a major loophole in the analysis based on (3.2). The point is that (3.2) gives the contribution at fixed instanton number k, whereas the full instanton contribution is given by an infinite series of the form

$$S(v/\Lambda) = \sum_{k=1}^{\infty} c_k \left(\frac{\Lambda}{v}\right)^{kN\beta_0}. \tag{3.3}$$

As long as $|v|\gg|\Lambda|$, the series converges, and a term-by-term analysis makes sense. In particular, we do expect the sum S to vanish exponentially at large N. However, at strong coupling $|v|\ll|\Lambda|$, things are entirely different because the large instantons make the sum diverges, and it is meaningless to isolate a particular term in the sum. This corresponds to the intuitive idea [29] that the instanton gas disappears. However, this does not imply that $S(v/\Lambda)$ itself is ill-defined in the strong coupling regime, because the series (3.2) can have a smooth analytic continuation. In some sense, instanton will transmute into something new through the process of analytic continuation. What this might be can be discovered by working out explicitly the large N limit of the analytic continuation of (3.3) [21], and the result turns out to be extremely interesting.

3.3 A toy model example

Since we do not want to assume that the reader is familiar with the technology of the exact results in $\mathcal{N}=2$ gauge theories, we will discuss a simple toy example instead. Our toy example actually plays a role in the full calculation in the gauge theory [1, 21], but it is much simpler and nonetheless illustrates the main points discussed in Sections 3.1 and 3.2. The model is based on the equation

$$P(\sigma) = \sigma \prod_{i=1}^{N-1} (\sigma + v_i) = \Lambda^N. \tag{3.4}$$

The observables are the roots of this equation. The moduli space $\mathcal{M}$ is parametrized by the v_is. Critical points C_p are obtained when $p \geq 2$ roots coincide. The vicinity of C_p is described by the equation

$$\sigma^p + \sum_{k=2}^{p} t_{p-k}\sigma^{p-k} = 0\,. \tag{3.5}$$

The t_ks depend on the v_is and correspond to the $p-1$ independent relevant deformations of C_p. They are associated with a set of critical exponents

$$\delta_k^{(p)} = \frac{1}{p-k}\,\cdot \tag{3.6}$$

For example, if the most relevant operator t_0 is turned on, then the separation of the roots is of order $t_0^{\delta_0^{(p)}}$.

Let us focus on the root σ_N that goes to zero in the classical, or weak coupling, limit $\Lambda/v_i \to 0$. In this limit, σ_N is given by an instanton series of the type (3.3),

$$\sigma_N = \sum_{k=1}^{\infty} \tilde{c}_k\, \Lambda^{kN}, \tag{3.7}$$

where the coefficients $\tilde{c}_k$ can be expressed in terms of the polynomial P,

$$\tilde{c}_k = \frac{1}{k!}\left(\frac{1}{P'}\frac{\mathrm{d}}{\mathrm{d}\sigma}\right)^{k-1} \cdot \frac{1}{P'}\,\Big|_{\sigma=0}\cdot \tag{3.8}$$

The radius of convergence of (3.7) as a function of the distribution of the v_is can be calculated at large N with the methods of [21]. For our purposes, it is enough to consider the simple case $v_1 = \cdots = v_{N-1} = v$. In terms of the dimensionless ratio

$$r = v/\Lambda\,, \tag{3.9}$$

the expansion (3.7) takes the form

$$\sigma_N(r) = v \sum_{k=1}^{\infty} c_k \, r^{-kN} \tag{3.10}$$

with $c_1 = 1$, $c_2 = 1 - N$, etc. The radius of convergence of the series (3.10) is finite because $\sigma_N(r)$ has branch cuts. The branching points occur when the root σ_N coincides with another root of the equation (3.4), and thus correspond to critical points of the type C_2. It is elementary to show that there are N critical values r_* of r given by

$$r_*^N = -\frac{N^N}{(N-1)^{N-1}} \tag{3.11}$$

and for which

$$\sigma_N(r_*) = \sigma_* = -v/N. \tag{3.12}$$

We see that $\sigma_N(r_*)$ is of order $1/N$. There are N different analytic continuations of $\sigma_N(r)$ for $|r| < |r_*|$, corresponding to the N branching points (3.11), and which yield any of the N roots of (3.4). To obtain the large N expansion of $\sigma_N(r)$ for $|r| < |r_*|$, we write (3.4) as

$$\ln(1+x) + \frac{1}{N} \ln \frac{x}{1+x} = \ln \frac{1}{\rho_\alpha}\,, \tag{3.13}$$

where $x = \sigma/v$ and $\rho_\alpha = \mathrm{e}^{i\alpha} r = \mathrm{e}^{2i\pi k/N} r$, $0 \le k \le N-1$. The integer k labels the different analytic continuations. In the $1/N$ expansion, $\alpha = 2\pi k/N \in [0, 2\pi[$ is actually a continuous variable. At leading order, the term proportional to $1/N$ in (3.13) is negligible, and we get $\sigma_N(r)/v = -1 + 1/\rho_\alpha + o(1)$. Corrections are obtained by substituting $x \to -1 + 1/\rho_\alpha + \Delta$ in (3.13) and solving for Δ. The first few terms are

$$\frac{\sigma_N(r)}{v} = -1 + \frac{1}{\rho_\alpha} - \frac{\ln(1-\rho_\alpha)}{N\rho_\alpha} + \frac{1}{2\rho_\alpha N^2}\Big[\big(\ln(1-\rho_\alpha)\big)^2 + \frac{2\rho_\alpha \ln(1-\rho_\alpha)}{1-\rho_\alpha}\Big] + \mathcal{O}(1/N^3). \tag{3.14}$$

The equation (3.14) displays all the main features of the large N expansion of the analytic continuations of instanton series. The same features are found in the full gauge theory calculations [21]. First of all, it is an asymptotic series *with expansion parameter* $1/N$. This is very different from the series in $1/N^2$ that is generated by the 't Hooft expansion in terms of Feynman diagrams[3]. It means that open strings must be

[3]Let us emphasize that the $SU(N)$ gauge theories we are considering have fields in the adjoint representation only.

present in any string theoretic description of the $\mathcal{N} = 2$ gauge theory. The physics underlying this fact is that the pure closed string background is singular at strong coupling. The singularity is resolved by inflating branes [30]. The open strings must be the strings that are attached to these branes [21]. A second general qualitative feature of (3.14) is that the contribution at each order in $1/N$ is given by a series in $1/\rho_\alpha$, by writing $\ln(1-\rho_\alpha) = \ln(-\rho_\alpha) + \ln(1-1/\rho_\alpha) = \ln(-\rho_\alpha) - \sum_{j=1}^{\infty} 1/(j\rho_\alpha^j)$. Since an instanton contributes $1/\rho_\alpha^N$, we may interpret those terms as coming from fractional instantons. The physical picture is then that instantons have disintegrated through the process of analytic continuation in the strong coupling region[4]. The third qualitative feature is that the large N expansion breaks down at $\rho_\alpha = 1$, which corresponds to the critical point where two roots coincide.

This last feature is of course crucial for our purposes. The origin of the divergences is exactly as discussed in Section 3.1. Introducing

$$\epsilon = \rho_\alpha - 1\,, \tag{3.15}$$

the separation of the two colliding roots goes like $\delta\sigma \propto \epsilon^{1/2}$ when $\epsilon \to 0$, the critical exponent $\delta_0^{(2)} = 1/2$ being given by (3.6). However, it is straightforward to see that in the leading large N approximation $\delta_0^{(2)} = 1$. This erroneous result is obtained by noting that the formula (3.14) gives the large N expansion of the $N-1$ roots $\sigma_1, \ldots, \sigma_{N-1}$ for $\rho_\alpha > 1$, a regime in which $\sigma_N = 0$ to all orders in $1/N$ due to the exponential decrease of instantons. The divergences in the corrections to the leading approximation in (3.14) when $\epsilon \to 0$ signal the failure of the $1/N$ expansion to yield the correct critical exponent, as discussed in Section 3.1.

The next step is to check whether the divergences are specific enough for a double scaling limit to be defined. For this purpose, let us consider the rescaled observable

$$s = N\sigma_N/v \tag{3.16}$$

and the double scaling limit

$$N \to \infty\,, \quad \epsilon \to 0\,, \quad N\epsilon - \ln N = \text{constant} = 1/\kappa + \ln\kappa\,. \tag{3.17}$$

By plugging in (3.17) into (3.14), and discarding terms of order κ^2 and higher, we get

$$\begin{aligned} s = &\big(-1/\kappa - \ln\kappa - \ln N\big) + \big(-\ln(-1) + \ln\kappa + \ln N - \kappa\ln\kappa - \kappa\ln N\big) \\ &+ \big(-\kappa\ln(-1) + \kappa\ln N + \kappa\ln\kappa\big) + \cdots \end{aligned} \tag{3.18}$$

[4]The fractional instanton picture is only heuristic, because we have not found the corresponding field configurations (that must be singular in the original field variables), and also because at large N the fractional topological charge is vanishingly small.

We have grouped together in (3.18) the terms that come from a given order in $1/N$ in (3.14). We see that subtle cancellations between the different terms make the result finite to the order we consider,

$$s = -1/\kappa - \ln(-1) - \kappa \ln(-1) + \mathcal{O}(\kappa^2)\,. \tag{3.19}$$

It is actually very easy to see that the cancellations will work to all orders and beyond. By using the scalings (3.16) and (3.17) in the exact equation (3.13), we indeed obtain a non-perturbative equation determining s,

$$s\mathrm{e}^s = \frac{1}{\kappa}\,\mathrm{e}^{-1/\kappa}\,. \tag{3.20}$$

This is an *exact* result for the double scaled theory, from which we can in particular derive the full asymptotic series in κ and thus recover (3.19). The analogous string theoretic results that we obtain by using a gauge theory instead of the simple toy model (3.4) are described below.

3.4 Exact results in 4D string theory

For concreteness, let us write down explicitly some of the exact results obtained in [1]. The critical points we consider are Argyres-Douglas critical points AD_p which have $p-1$ relevant deformations t_k, $0 \le k \le p-2$. The space-time dimensions of the t_ks can be deduced from the exact solution for the parent $\mathcal{N} = 2$ supersymmetric gauge theory, and read

$$[t_k]_{\mathrm{space-time}} = \frac{2(p-k)}{p+2}\,\cdot \tag{3.21}$$

The double scaling limit, similar to (2.23), is

$$N \to \infty\,, \quad t_k \to 0\,, \quad \tau_k = N^{1-k/n}\,t_k = \text{constant}\,. \tag{3.22}$$

By identifying the most relevant operator τ_0 with the world sheet cosmological constant, and by using the reasoning in Section 2.4, equation (2.15) and below, we deduce that the world sheet cut-off $1/\ell$ scales as

$$\ell^2 \sim 1/N \tag{3.23}$$

and that the world sheet dimensions are

$$[\tau_k]_{\text{world sheet}} = \frac{2(p-k)}{p}\,\cdot \tag{3.24}$$

The string theoretic central charge z of the supersymmetry algebra as a function of the world sheet couplings τ_k reads

$$z(\gamma) = \frac{1}{4\tau_0^{1/p}} \oint_\gamma \frac{uT'(u)}{\sqrt{1-\mathrm{e}^{-T(u)}}}\,\mathrm{d}u\,, \tag{3.25}$$

with

$$T(u) = \sum_{k=0}^{p-2} \tau_k u^k + u^p \,. \tag{3.26}$$

The contour γ encircles any two roots of the polynomial T, and corresponds to a choice of electric and magnetic charges. The formula (3.25) includes an infinite series of string perturbative corrections as well as all the non-perturbative contributions. It can be viewed as the "realistic" generalization of the toy model equation (3.20). For example, in the special case where only the most relevant operator is turned on, it is natural to introduce the coupling $\kappa = 1/\tau_0$, to rescale $u \to \kappa^{-1/p} u$, and to consider the contours γ_{jk} encircling the roots $u_j = \exp(i\pi(1+2j)/p)$ and u_k. A straightforward calculation then yields

$$z(\gamma_{jk}) = \frac{p}{4\kappa} \oint_{\alpha_{jk}} \frac{u^p \, \mathrm{d}u}{\sqrt{1 - \mathrm{e}^{-(1+u^p)/\kappa}}} = \mathrm{e}^{i\pi(j+k+1)/p} \sin\bigl(\pi(j-k)/p\bigr) \, \mathcal{I}_p(\kappa) \tag{3.27}$$

where

$$\mathcal{I}_p(\kappa) = \frac{p}{(p+1)\kappa} + \int_0^{1/\kappa} \Bigl(\frac{1}{\sqrt{1-\mathrm{e}^{-x}}} - 1 \Bigr) (1-\kappa x)^{1/p} \mathrm{d}x. \tag{3.28}$$

The asymptotic expansion of $\mathcal{I}_p(\kappa)$ can be obtained by noting that when $x \sim 1/\kappa$, $1/\sqrt{1-\exp(-x)} - 1$ is exponentially small. We thus have

$$\mathcal{I}_p(\kappa) = \frac{p}{(p+1)\kappa} + \sum_{k=0}^{K} \frac{\Gamma(k-1/p)}{\Gamma(-1/p)\Gamma(k+1)} I_k \, \kappa^k + \mathcal{O}(\kappa^{K+1}) \tag{3.29}$$

with

$$I_k = \int_0^\infty \Bigl(\frac{1}{\sqrt{1-\mathrm{e}^{-x}}} - 1 \Bigr) x^k \, \mathrm{d}x = \frac{(-1)^{k+1}}{k+1} \int_0^1 \frac{\bigl(\ln(1-t)\bigr)^{k+1}}{2t^{3/2}} \, \mathrm{d}t. \tag{3.30}$$

The first integrals I_k can be calculated by expanding the logarithm in powers of t,

$$I_0 = 2\ln 2 \,, \quad I_1 = \frac{\pi^2}{6} - 2(\ln 2)^2 \,, \quad I_2 = \frac{8(\ln 2)^2}{3} - \frac{2\pi^2 \ln 2}{3} + 4\zeta(3) \,, \tag{3.31}$$

which gives the first string loops corrections.

An important point is that the central charge Z, and thus the BPS masses, of the parent gauge theory scales as [1]

$$Z \sim N^{-1/p} z \sim \ell^{2/p}\,, \tag{3.32}$$

and thus the continuum UV limit on the world sheet $\ell \to 0$ corresponds to a low energy limit of the parent gauge theory $Z \to 0$. This shows explicitly that in the double scaling limit we are left with the low energy degrees of freedom only, as was already deduced from general arguments at the end of Section 3.1.

3.5 Further insights

3.5.1 Full proofs

The consideration of the protected observables on the gauge theory side, that correspond to the central charge or equivalently to the low energy effective action, is not enough to give a full proof of the existence of double scaling limits. One should study in principle *all* the observables, including those with a non-trivial perturbative expansion, or equivalently the full path integral. Moreover, the heuristic picture for the appearance of a continuum string theory in the limit relies on the observation that very large Feynman graphs dominate near the critical points, as reviewed in Section 2.4. This can in principle be checked on generic amplitudes but obviously not on the BPS observables for which perturbation theory is trivial. One should also give a proof that the continuum limit does correspond to a genuine continuum string theory, a fact that is extremely difficult to check directly even in the $D \leq 1$ cases.

Unfortunately, generic amplitudes cannot be calculated in Yang-Mills theory. For that reason, it is interesting to consider simplified models that can be exactly solved. A particularly interesting one was studied in [22]. The model is a two dimensional non-linear σ model which is a very close relative to $\mathcal{N} = 2$ super Yang-Mills in four dimensions. It has an exactly calculable central charge with the same non-renormalization theorems as in four dimensions and the same BPS mass formula. The analysis sketched in Sections 3.1–3.4 can thus be reproduced, with qualitatively the same results (appearance of "fractional instantons", breakdown of the large N expansion at critical points, possibility to define double scaling limits for which exactly known BPS amplitudes have a finite limit). Moreover, and this is the main point, the two-dimensional model is exactly solvable in the large N limit: the large N Feynman graphs can be explicitly summed up. It is then possible to give rigorous proofs of the existence of the double scaling limits, and to fully characterize the double scaled theories. In particular, we do obtain the expected continuum limits. Those results are extremely encouraging and strongly suggest that the scaling limits are consistent in the case of the gauge theories as well. We invite the reader to consult [22] for more details.

3.5.2 Non-perturbative non-Borel summable partition functions

As we have already emphasized, a very important point of principle is that the four dimensional double scaling limits are non-perturbative. This is to be contrasted with the classic $D \leq 1$ cases, for which the non-Borel summable observables are not defined beyond perturbation theory. We have discussed this aspect in some details in [23], where interesting mechanisms that allow for a non-perturbative definition of non-Borel summable partition functions are explicitly worked out in a class of simple theories akin to the model studied in [22].

4 Open problems

The results obtained in [1] show that a generalization of the discretized approach to four dimensional non-critical strings is possible. However, we have only scratched the surface of the subject, and many points would deserve further investigations. For example, the critical points used in [1] only correspond to one family amongst many others that can be found on the moduli space of supersymmetric gauge theories. A classification of these critical points exists [26], with an ADE pattern, and the work in [1] could certainly be generalized to the most general cases. Another interesting line of research is the study of the renormalization group flow equations between the different critical points that can be derived by using the explicit formulas of [1]. In the classic low dimensional cases, a very elegant mathematical structure emerges (generalized KdV hierarchy) when one studies the flows (see for example Sects. 4 and 8 of [19]). It would be extremely interesting to discover a similar structure in the four dimensional theories. Yet another intriguing point is that the approach of [1] suggests a relationship between the analytic continuation of the sum over gauge theory instantons in the double scaling limit and sums over Riemann surfaces with boundaries which define the perturbative series of the resulting string theories. It would be worth studying if this correspondence can be proven directly by looking at the moduli space of instantons at large N and large instanton number k. Finally, it is highly desirable to construct directly in the continuum the string theories obtained in the double scaling limits. As in the classic $D \leq 1$ case, this would provide an important consistency check of the discretized approach. Moreover, we believe that the ADE "exactly solvable" four dimensional non-critical strings could play a role similar to the ADE minimal two dimensional CFTs, and shed considerable light into the structure of non-perturbative string theories.

I would like to thank the Les Houches summer school for providing a very nice learning and thinking environment. I am grateful to Rafael Hernández, Igor Klebanov, Ivan Kostov and particularly Alexandre Polyakov for comments on the work presented in these lectures. The research was supported by a Robert H. Dicke fellowship from Princeton University, and more recently by the Swiss National Science Foundation and the University of Neuchâtel.

References

[1] F. Ferrari, *Nucl. Phys. B* **617** (2001) 348.

[2] K. Wilson, *Phys. Rev. D* **10** (1974) 2445.

[3] S. Mandelstam, *Phys. Rep.* **236** (1976) 245; G. 't Hooft, in *High Energy Physics*, Proc. European Phys. Soc. Int. Conf., edited by A. Zichichi (Bologna 1976).

[4] J. Polchinski, *String Theory*, two volumes (Cambridge University Press, 1998).

[5] G. 't Hooft, *Nucl. Phys. B* **72** (1974) 461.

[6] J. Maldacena, *Adv. Theor. Math. Phys.* **2** (1998) 231; S. Gubser, I.R. Klebanov and A.M. Polyakov, *Phys. Lett. B* **428** (1998) 105; E. Witten, *Adv. Theor. Math. Phys.* **2** (1998) 253.

[7] J. Polchinski, *Phys. Rev. Lett.* **75** (1995) 4724.

[8] G. Horowitz and A. Strominger, *Nucl. Phys. B* **360** (1991) 197.

[9] J.M. Maldacena, *Large N field theories and gravity*, Les Houches lecture in this volume; I.R. Klebanov, *D-Branes on the Conifold and $\mathcal{N} = 1$ Gauge/Gravity Dualities*, Les Houches lecture in this volume [hep-th/0205100].

[10] J. Polchinski and L. Susskind, *String Theory and the Size of Hadrons*, NSF-ITP-01-185, SU-ITP 01/56 [hep-th/0112204].

[11] A.M. Polyakov, *Phys. Lett. B* **103** (1981) 207; T. Curtright and C. Thorn, *Phys. Rev. Lett.* **48** (1982) 1309; J.-L. Gervais and A. Neuveu, *Nucl. Phys. B* **199** (1982) 59.

[12] P. Ginsparg and G.W. Moore, *Lectures on 2D Gravity and 2D String Theory,* TASI lectures 1992 [hep-th/9304011].

[13] A.M. Polyakov, *Nucl. Phys. Proc. Suppl.* **68** (1998) 1; A.M. Polyakov, *Int. J. Mod. Phys. A* **14** (1999) 645.

[14] É. Brézin, C. Itzykson, G. Parisi and J.-B. Zuber, *Comm. Math. Phys.* **59** (1978) 35.

[15] F. David, *Nucl. Phys. B* **257** (1985) 45; V.A. Kazakov, *Phys. Lett. B* **150** (1985) 282; J. Ambjörn, B. Durhuus and J. Fröhlich, *Nucl. Phys. B* **257** (1985) 433.

[16] É. Brézin and V.A. Kazakov, *Phys. Lett. B* **236** (1990) 144; M.R. Douglas and S. Shenker, *Nucl. Phys. B* **355** (1990) 635; D.J. Gross and A.A. Migdal, *Phys. Rev. Lett.* **64** (1990) 127.

[17] É. Brézin, V.A. Kazakov and Al.B. Zamolodchikov, *Nucl. Phys. B* **338** (1990) 673; D. Gross and M. Milkovic, *Phys. Lett. B* **238** (1990) 217; G. Parisi, *Phys. Lett. B* **238** (1990) 209; P. Ginsparg and J. Zinn-Justin, *Phys. Lett. B* **240** (1990) 333.

[18] I.R. Klebanov, *String theory in two dimensions,* lectures at the ICTP Spring School on String Theory and Quantum Gravity (Trieste, 1991) [hep-th/9108019].

[19] P. Di Francesco, P. Ginsparg and J. Zinn-Justin, *Phys. Rep.* **254** (1995) 1.

[20] F. Ferrari, *Phys. Lett. B* **496** (2000) 212; F. Ferrari, *J. High Energy Phys.* **6** (2001) 57.

[21] F. Ferrari, *Nucl. Phys. B* **612** (2001) 151.

[22] F. Ferrari, *Large N and double scaling limits in two dimensions,* NEIP-01-008, PUPT-1997, LPTENS-01/11 [hep-th/0202002].

[23] F. Ferrari, *Non-perturbative double scaling limits,* PUPT-1998, NEIP-01-009, LPTENS-02/11 [hep-th/0202205].

[24] E.H. Fradkin and S.H. Shenker, *Phys. Rev. D* **19** (1979) 3682.

[25] N. Seiberg and E. Witten, *Nucl. Phys. B* **426** (1994) 19; Erratum B **430** (1994) 485.

[26] P.C. Argyres and M.R. Douglas, *Nucl. Phys. B* **448** (1995) 93; T. Eguchi, K. Hori, K. Ito and S.-K. Yang, *Nucl. Phys. B* **471** (1996) 430.

[27] A.M. Polyakov, private communication (2001).

[28] P.C. Argyres and A.E. Faraggi, *Phys. Rev. Lett.* **74** (1995) 3931; A. Klemm, W. Lerche, S. Yankielowicz and S. Theisen, *Phys. Lett. B* **344** (1995) 169.

[29] E. Witten, *Nucl. Phys. B* **149** (1979) 285.

[30] C.V. Johnson, A.W. Peet and J. Polchinski, *Phys. Rev. D* **61** (2000) 86001.

SEMINAR 4

U-OPPORTUNITIES: WHY IS TEN EQUAL TO TEN?

B. JULIA

Laboratoire de Physique Théorique,
ENS, 24 rue Lhomond,
75231 Paris Cedex 05, France

Contents

U-OPPORTUNITIES: WHY IS TEN EQUAL TO TEN?

B. JULIA

Abstract

It seems to me at this time that two recent developments may permit fast progress on our way to understand the symmetry structure of toroidally (compactified and) reduced M-theory. The first indication of a (possibly) thin spot in the wall that prevents us from deriving *a priori* the U-duality symmetries of these models is to be found in the analysis of the hyperbolic billiards that control the chaotic time evolution of (quasi)homogeneous anisotropic String, Supergravity or Einstein cosmologies near a spacelike singularity. What happens is that U-duality symmetry controls chaos *via* negative constant curvature. On the other hand it was noticed in 1982 that (symmetrizable) "hyperbolic" Kac-Moody algebras have maximal rank ten, exactly like superstring models and that two of these four rank ten algebras matched physical theories. My second reason for optimism actually predates also the previous breakthrough, it was the discovery in 1998 of surprising superalgebras extending U-dualities to all $(p+1)$-forms (associated to p-branes). They have a super-nonlinear sigma model structure similar to the symmetric space structure associated to 0-forms and they obey a universal self-duality field equation. As the set of forms is doubled to implement electric-magnetic duality, they obey a set of first order equations. More remains to be discovered but the beauty and subtlety of the structure cannot be a random emergence from chaos. In fact we shall explain how a third maximal rank hyperbolic algebra BE_{10} controls heterotic cosmological chaos and how as predicted Einstein's General Relativity fits into the general picture.

1 Classifications

It is well known to conformal field theorists but a much more general and venerable fact that positive definite symmetric matrices with integer entries tend to appear in many classification problems. More precisely the ADE philosophy is to list such occurrences and try to relate them to each other.

Let us be a bit more specific, this set of problems corresponds to the emergence in various contexts of symmetric matrices with diagonal entries equal to 2 and negative integral off-diagonal entries. The prototype of a successful correspondence is the work of Brieskorn (with some help from Grothendieck) realising the simple complex Kleinian singularities related to discrete subgroups of $SU(2)$ as rational singularities in the set of unipotent elements of the corresponding complex Lie group of type ADE. The singularity is at the conjugacy class of subregular elements in other words the elements whose centraliser has 2 more dimensions than regular ones namely $r+2$ instead of r, the group rank. The resolution of the singularity introduces exceptional divisors whose intersection matrix is the opposite of the group Cartan matrix, there the ADE matrices arise in two different disguises but there is a relation to the Lie group in both cases.

Halfway (in time) between the Brieskorn results and the classification of modular invariant partition functions by Cappelli *et al.* $N=8$ supergravity was constructed in 4 dimensions as well as its toroidal decompactification family up to 11 spacetime dimensions. It was quickly remarked to me by Y. Manin that the E_r internal symmetry groups appearing upon compactification on a r-torus suggested a role for (regular) del Pezzo surfaces, these are variants of CP_2 or $CP_1 \times CP_1$ which admit a canonical projective embedding . However the other pure supergravities in 4 dimensions with fewer supersymmetries also belonged to families which together formed a magic triangle of theories. They led also to ADE groups yet not all in split form, this meant that real geometry was to be tackled rather than the simpler complex algebraic one.

It also rapidly became clear if not rigorously established that the E_r family included the infinite dimensional $E_9 \equiv E_8^{(1)}$ in 2 dimensions *i.e.* after compactification on a 9-torus. Our partial understanding of these duality groups includes their $A_{r-1} \times \mathbb{R}$ subgroups related to the r ignorable coordinates. The infinite dimensional hidden symmetry is related to the existence of a Lax pair in 2 residual dimensions leading to a quasi integrable situation yet it was known that chaos remained prevalent in some particular flows.

In 1982 I noticed that a naive extrapolation to 1 dimension would suggest a role for E_{10} as one would expect its subalgebra A_9 to appear there, however the implementation was problematic. Nevertheless a similar analysis of so-called type I supergravity in other words pure (without matter multiplet) supergravity in 10 dimensions led to the suggestion of a corresponding role for the other hyperbolic Kac-Moody algebra DE_{10} the "overextended" D_8 in split form (namely $SO(8,8)$ affinized with one more $SL(2,\mathbb{R})$ generating subgroup right next to the affinizing $SL(2,\mathbb{R})$ in the Dynkin diagram as implied by the A_{r-1} argument). Now according to Bourbaki (actually Chein) E_{10} and DE_{10} are exactly the two simply laced hyperbolic Kac-Moody algebras of maximal rank. There are only two more (non simply laced) hyperbolic Kac-Moody algebras of maximal rank: BE_{10} (see Sect. 3) and CE_{10} (to be seen).

Let us recall that hyperbolicity here means that the Dynkin diagram becomes a product of finite or affine Dynkin diagrams after removal of any node (this gives nice arithmetic properties as well). The rank ten here is directly related to the rank eight of E_8, the largest exceptional simple Lie group which itself comes about also from some subtle classification analysis. On the other hand ten is the critical dimension of superstring models and as such it is related to quantum conformal invariance. How could the two derivations be related? They must be as the rank of the Lie group is closely related to the dimension of the compactification torus!

Still a puzzling fact emerged from the subsequent construction of heterotic strings namely the possibility of duality groups of rank higher than 8 for instance 16 in dimension 3 ($SO(8, 24)$) corresponding putatively to rank 18 in one dimension. This paradoxical apparent violation of the bound ten on the rank will be clarified in Section 3. Let us also remark that we are talking of superstrings and (bosonic) Kac-Moody algebras, hence fermionic structures should emerge in an *a priori* bosonic context. The upper limit 26 does not arise so simply yet, although it can be argued to be the sum of ten and sixteen the latter being the rank of the two even euclidean unimodular lattices dictated again by quantum anomaly considerations.

Let us add some examples of ADE objects. The basic one is the set of integral positive definite matrices occuring as Cartan matrices of simple simply laced complex Lie groups. The positive definiteness, resp positive semi-definiteness, resp hyperbolicity guarantees a relatively simple classification, for instance in the positive definite case (that of finite dimensional simply laced Lie algebras) there is a finite number of objects for each rank; on the other hand the three cases of E_6, E_7, E_8 look exceptional in this context. In the semi-definite case also called affine Kac-Moody situation very much the same is true; in the hyperbolic case however as we have seen the rank is bounded and there are finitely many instances of a given rank except when it is equal to two.

The finite dimensional (irreducible) Coxeter groups generated by reflections are closely related objects, their list encompasses that of the Weyl groups of the simple Lie groups but one gets 3 extra Coxeter diagrams with rotation angle $2\pi/5$ and an infinite family of dihedral groups of rotation angles $2\pi/k$ for non cristallographic k's integers at least 7. These are all nonsimply laced cases. It is important to note that non simply laced Lie groups have a symmetrisable Cartan matrix: a basic assumption for most of

Kac-Moody theory and Borcherds algebras. The reflections preserve a symmetric form that is a symmetrisation of the non symmetric Cartan matrix.

By definition Coxeter groups admit a finite presentation by involutions S_i , $i = 1, \ldots, r$ satisfying

$$(S_i S_j)^{m_{ij}} = I\,,\; i \neq j. \tag{1.1}$$

In the simply laced case we consider only exponents $m_{ij} = 2$ for commuting involutions or 3 for dihedral subgroups of order 6. The matrix is encoded by a Dynkin diagram with r vertices and simple bonds for exponent 3; it turns out that no loop is allowed, that at most three legs occur and finally that the sum $1/m+1/n+1/p$ of the inverses of the number of vertices (including the potentially trivalent vertex) on each leg must be strictly larger than one. One recognises two infinite families A_k, D_k and three exceptions E_6, E_7, E_8 with numbers of vertices respectively $(m, n, p) = (2, 3, 3), (2, 3, 4), (2, 3, 5)$, let us notice that $E_5 \equiv D_5$ corresponds to $(2, 3, 2)$ and $E_4 \equiv A_4$ to $(2, 3, 1)$. $E_3 \equiv A_2 \times A_1$ which is semi-simple but not simple is the next group in the E_r family.

The list of affine Kac-Moody algebras is very closely related to the list of finite dimensional simple Lie algebras. It permits one loop for the $A_k^{(1)}$ Dynkin diagrams, sums of inverse numbers of vertices at one three-valent vertex equal to one, two three valent vertices for $D_k^{(1)}$ $k \geq 5$ resp. one 4-valent vertex for $D_4^{(1)}$. The list of hyperbolic diagrams can be found in [1,2]. Their defining property given above guarantees the Lorentzian signature of the invariant bilinear form on the Cartan subalgebra. The hyperbolic algebras one meets in supergravity theories are overextensions of finite Lie algebras, the construction was discussed in [3,4]. Not all overextensions are hyperbolic but all hyperbolic algebras of rank at least 7 are overextensions (called superaffine in [1]). The derivation of the signature is straightforward, if A is a kxk affine Cartan matrix it has null determinant and signature $(+^{k-1}, 0)$, the $(k + 1)$st line and column contain a 2 at their intersection and a -1 at the affine column or row, the corresponding quadratic form after completing the square has manifest Lorentzian signature. Another characteristic property we shall use in Section 3 is the fact that the Weyl chamber on the unit hyperboloid has finite volume (it may actually be non compact, for instance when the rank is strictly larger than 5 [5,6]).

Let us now discuss in more detail the case of overextended A_{k-1}. The affine $A_{k-1}^{(1)}$ has a circular Dynkin diagram, its (over)extension has just one line and one extra vertex attached to it, now the criterion of hyperbolicity prevents overextended A_8 to be hyperbolic, overextended A_7 is the last hyperbolic HA_9 in the family, and this can be viewed again as a consequence of the fact that there are only three Lie groups of type E. So one may say that the exceptionality of the E family is related to the bound on the rank of the hyperbolic HA_{k+1}. We shall see in Section 3 the dramatic difference between pure gravity dynamics at a singularity in ten or eleven dimensions as a consequence of this algebraic fact.

Let us recall that E_{10} and HD_{10} were identified in [3], the Weyl chamber of HB_{10} (in other words overextended B_8) was suggestively recognised by [7] as controlling the classical chaos of heterotic string theory yet following Narain and Sen one expects the U-duality group to have rank 16 in 3 dimensions and not 8. The answer lies in the simple observation that the classical action is a real functional and the real equations are invariant under a real Lie group, the precise real form of which is critically important. But as an expert (M. Reid) puts it, real algebraic geometry is 2^N times more difficult than complex algebraic geometry with N large.

2 Real forms of Lie algebras

We are familiar with the classification of complex simple Lie algebras as a monument of group theory. Its relative simplicity is permitted by the algebraic closure of the field of complex numbers, indeed the main tool is the simultaneous diagonalisation of commuting Cartan generators (observables) and the analysis of the root spectrum (quantum numbers). Up to central elements (non simple-connectedness) and up to isomorphism there is a unique compact form of the associated Lie group. The theory of non compact forms and their representation theory is much richer and even the split (also called maximally non-compact forms) have complicated representations. The existence of the split form follows from the observation that the structure constants of a complex Lie algebra can be taken to be integers in the appropriate basis. The restriction of the field of coefficients from the complex to the real numbers or even the natural integers is possible, the choice of real "angles of rotations" in the Cartan-Chevalley basis defines the split form. For $SL(2,C)$ the split form is $SL(2,\mathbb{R})$ whereas the compact form is $SU(2)$ and the arithmetic form $SL(2,\mathbb{Z})$.

Over the complex numbers or in the compact case the Cartan subalgebras are all conjugate to each other, not so for other real forms G even the split one, still there is a finite collection of inequivalent ones. The classification of real forms of simple Lie groups has been given by E. Cartan together with the classification of maximally symmetric spaces. Since then two strategies have been applied. Either one selects a maximally non-compact Cartan torus in G and diagonalises as many observables as possible over the reals, there appears a maximal split subalgebra S [16] p. 116 inside our noncompact real form G and the roots project onto roots or twice the roots of S, this is the Tits-Satake theory with bicoloured diagrams as developped for instance in [8]. One key result is that the roots restricted to the noncompact Cartan generators form a not necessarily reduced root system. In particular this implies that after choosing the compact real form that contains the Cartan (compact) generators resp. their multiples by the imaginary unit i (for the non-compact ones), all "imaginary roots" *i.e.* those that vanish on the (maximal) set of non-compact Cartan generators must be associated to compact eigengenerators.

The other strategy leaves more freedom, given a non-compact real form even after restricting oneself to the case of a maximally compact Cartan subalgebra. Vogan introduced other bicoloured diagrams for that situation. For instance the case of fully compact tori arises sometimes for non-compact Lie algebras and is very interesting. Now the arbitrariness with the Vogan strategy stems from the fact that the choice of simple roots is not unique there. The bicoloured diagrams may be chosen to have all or all but one compact vertices. Clearly this choice permits an easier identification of compact subgroups whereas the Tits-Satake strategy is best for split subgroups. It is natural to expect a dual result to the previous one namely that now no root can vanish on the (maximally) compact part of the Cartan torus, in other words one departs maximally from the complex root analysis. We refer to [9] for an introduction to the Borel-de Siebenthal-Murakami-Vogan theory.

The real rank l of a simple real Lie algebra of full rank r is the maximal dimension of an Abelian subalgebra of diagonalisable (called semi-simple) generators of non-compact type (whose Killing norm is positive). $0 \leq l \leq r$, with $r = l$ in the split case and $l = 0$ in the compact case. If one starts from the split form the compact form is obtained by multiplying some generators by $\sqrt{-1}$. For instance given the standard basis e, f, h of $SL(2,\mathbb{R})$ the compact $SU(2)$ is generated by $(e - f)$ which is already compact as well as $i(e + f)$ and ih. We see that the diagonalisability of ih has been lost over the real numbers.

Turning now to the applications we may read off the tables of [8, 10] that the restricted root system associated to the non-split $E_7(-5)$ of real rank 4 (the U-duality symmetry of $N = 6$ 4d supergravity reduced to three dimensions) is that of its maximal split subalgebra F_4 (one speaks of Freudenthal-Tits geometry of type F_4) and this is a common feature of all Maxwell-Einstein $N = 2$, $d = 5$ supergravities constructed by Günaydin *et al.* in 1983 after reduction to three dimensions. We refer to [10] p. 534 for the Tits Satake diagram of type EVI symmetric space or real form of E_7. On this diagram the white dots denote non compact $SL(2,\mathbb{R})$ subalgebras, three of them building up a compactification $SL(4,\mathbb{R})$ symmetry. The black dots refer to compact $SU(2)$'s.

In the tables one finds also the multiplicity of the restricted roots; for instance for the Lorentz group $SO(1,3)$ of real rank 1 the Tits-Satake diagram is composed of two disconnected white dots with one double arrow between them. This can be understood as follows, the root analysis over the real numbers goes through for the noncompact (white) Cartan generator(s). Upon projection of the full root system on its restriction to linear forms over the noncompact part of the Cartan subalgebra this may imply multiple occurrence of the same restricted roots, when this happens for two simple roots one joins the corresponding (white) dots by a double arrow. The real Lie algebra admits a complex structure precisely when all restricted roots have multiplicity two. On the other hand black dots project to zero. When the restricted root system is not reduced, ie for some geometries of type B_k one must also give the multiplicity of the doubles of the restricted roots.

Our present interest in the Tits-Satake analysis stems from the observation that for the orthogonal groups of the form $SO(8, 8+p)$ with p between 1 and 16 the real rank is 8 and the geometry is of type $SO(8,9)$. Now such groups arise as U-duality groups in three dimensions and the sigma model Lagrangians that control chaos in the quasi homogeneous situation must lead to the overextension BE_{10} if one is to recover the experimental discovery of [7].

3 Chaos controlled by symmetry

In a remarkable analysis [7] it was found that near a cosmological singularity the chaotic behaviour of essentially one dimensional (homogeneous) string theories was well approximated by an Anosov flow in the hyperbolic billiard defined by the Weyl cell of hyperbolic Kac-Moody algebras. For M-theory (alias in this approximation 11d SUGRA or type II String theory) E_{10} emerged, and for type I SUGRA DE_{10} replaces it. As we mentioned above both Lie algebras had surfaced as tantalising candidates for hidden symmetries in exactly these situations 20 years ago. Symmetry controls chaos which is not so surprising when negative curvature non-compact symmetric spaces are the arena. What remains more puzzling is the precise sense in which the 2 dimensional reductions of these models can be called integrable and still allow in their midst chaotic islands: they are well known to admit Lax pairs...

The chaotic solutions are even older and this tension between order and ergodicity is one of the most striking features of these Lagrangian theories. Another surprise of this work was the emergence of BE_{10}'s Weyl cell as the billiard relevant for heterotic and type I cosmological singularities. The relevant U-duality group in three dimensions is $SO(8,24)$ and its overextension is not hyperbolic. How should the overextension of $SO(8,9)$ take its place. On the other hand it was also remarked in [3] that BE_{10} and CE_{10} were nonsimply laced analogues of the previous two maximal rank algebras and ought to appear somewhere, half this prediction has been now fulfilled. We shall explain momentarily the reconciliation of $SO(8,24)$ and $SO(8,9)$.

In [3] it was also predicted that the overextension HA_3 of $SL(2,\mathbb{R})$, the Ehlers group of stationary General Relativity, was the more conservative candidate for a hidden symmetry but now in a well tested theory and again in the homogeneous situation. It was a very powerful experience to meet Alex Feingold and Igor Frenkel in Chicago that summer who independently and for mathematical reasons were working on [4] and had developed the theory of superaffine algebras (better called hyperaffine maybe) as a handle for hyperbolic algebras while I had been concentrating on the A_{k-1} concept for physical reasons. We learned a lot from each other then and it is a good place to express my thanks to the organisers I. Singer, P. Sally, G. Zuckerman, H. Garland and M. Flato for their invitation. Now this experience immediately suggested that the corresponding billiard should appear in the celebrated Belinsky-Khalatnikov-Lifschitz chaos. This time it appeared only in the quasi homogeneous situation *i.e.* not quite in the 1-dimensional setting, yet again the three dimensional U-duality controlled chaos by the same mechanism [11]. Furthermore the earlier observation by the belgian team of the absence of BKL chaos for pure gravity beyond 10 dimensions exactly matched the above remark about HA_9 as the largest hyperbolic overextension of a group of A type. It was known that A_k is precisely the U-duality group of the three dimensional reduction of pure gravity in $k+3$ dimensions [12].

Let us now see how the dimensionally reduced action of such a gravity theory (on a torus by homogeneity) implies the dynamical mechanism that is approximated by a hard walled billiard for time evolution near a singularity. We shall begin with the case of split U-dualities ie the better studied examples of [12]. As is well known in three dimensions all propagating fields can be dualised to scalars and form a noncompact symmetric space. There is a choice between two descriptions here. Either as it is the case when one actually discovers the system upon dimensional reduction one works in a fixed (or partially fixed) gauge and one uses coordinates on the symmetric space; for instance by using the Iwasawa decomposition one may use affine coordinates on the Borel subgroup AN of the noncompact group $G = KAN$ where K is the maximal compact subgroup of G. Or else one may restore a local K gauge invariance and use G valued scalar fields, this is the better way to discover and restore symmetries but the formalism developed in [12] and references therein is best suited for the fixed gauge approach.

The general action distinguishes the dilaton fields ie the coordinates along the (fully) non-compact Cartan subalgebra Lie(A) from the other scalar fields that correspond to positive root generators. It is simply a sum of quadratic kinetic terms for the latter weighted by appropriate exponentials of their respective roots *i.e.* linear forms in the dilatons plus free kinetic terms for the dilatons themselves. These exponential factors are responsible for the walls of the effective potential of the piecewise Kasner metric evolving in time [7, 11] after suitable overextension, in particular one must include the so-called symmetry wall.

Let us now recall the Iwasawa decomposition in the general case. It can be studied in [10] for instance. We are now considering a real Lie algebra with a maximally non-compact Cartan subalgebra whose noncompact part we denote by a. The maximal compact subgroup has a higher dimension than in the split case and the coset space has lower dimension. The decomposition reads now

$$\mathrm{Lie}(G) = \mathrm{Lie}(K) + a + n \tag{3.1}$$

where n is the nilpotent subalgebra of positive restricted root vectors.

Clearly in the non-split case the Cartan subalgebra is to be replaced by its non compact part and the nilpotent subalgebra of positive root vectors by that of positive restricted root vectors. Again the restricted roots form a possibly nonreduced root system. This brings two possible complications: firstly the multiplicities but they do not change the walls of the billiards and hence are irrelevant but also the nonreduced roots for some $SO(2k+1)$ restricted root systems and geometries.

Applications will appear soon, firstly in [13] we shall examine the replacement of $SO(8,24)$ by $SO(8,9)$; this paper contains also anomaly free string realisations of the DE_{10} theory as well as that of a theory with U-duality $SO(8,9)$ (this work was started before [7] and independently). Other instances of non split forms of U-dualities occuring in pure supergravities in 4d the so-called magic triangle will be analysed i [14] in order to check the precise control of chaos by symmetry in this more general situation. In many cases the simple rule of overextension of 3d U-duality is sufficient to analyse the chaotic or non chaotic behaviour of the flow. It is important to reach a more rigorous level of characterisation of the chaos and work is being done in this direction whereas theorems are available on the non-chaotic side [15].

I am grateful to J. McKay and M. Henneaux for references and discussions.

References

[1] A.P. Ogg, *Can. J. Math.* **36** (1984) 800.

[2] C. Saçioglu, *J. Phys. A* **22** (1989) 3753.

[3] B. Julia, *Lect. Appl. Math.* **21** (1985) 355.

[4] A. Feingold and I. Frenkel, *Math. Ann.* **263** (1983) 87.

[5] Bourbaki, *Groupes et algèbres de Lie 4, 5, 6* (Hermann, 1968).

[6] E.B. Vinberg, *Geom. II Encycl. Math. Sci.* **29** (1993).

[7] T. Damour and M. Henneaux, *Phys. Rev. Lett.* **86** (2001) 4749.

[8] S. Araki, *J. Math. Osaka City Un.* **13** (1962) 1.

[9] A.W. Knapp, *Lie groups beyond an introduction* (Birkhäuser, 1996).

[10] S. Helgason, *Differential geometry, Lie groups and symmetric spaces* (Academic Press, 1978).

[11] T. Damour, M. Henneaux, B. Julia and H. Nicolai, *Phys. Lett. B* **509** (2001) 323.

[12] E. Cremmer, B. Julia, H. Lu and C. Pope [`hep-th 9909099`].

[13] A. Hanany, B. Julia and A. Keurentjes, LPTENS 02-24 [`hep-th/0210xxx`].

[14] M. Henneaux and B. Julia, *Hyperbolic billiards of pure $d = 4$ SUGRAS*, in preparation [`hep-th/0210xxy`].

[15] L. Andersson and A.D. Rendall [`gr-qc/0001047`], see also [`gr-qc/0202069`].

[16] A. Borel and J. Tits, *Pub. IHES* **27** (1965) 55.

SEMINAR 5

EXACT ANSWERS TO APPROXIMATE QUESTIONS – NONCOMMUTATIVE DIPOLES, OPEN WILSON LINES AND UV-IR DUALITY

SOO-JONG REY

School of Physics & Center for Theoretical Physics, Seoul National University, Seoul 151-747, Korea

Contents

EXACT ANSWERS TO APPROXIMATE QUESTIONS – NONCOMMUTATIVE DIPOLES, OPEN WILSON LINES AND UV-IR DUALITY

Soo-Jong Rey

Abstract

In this lecture, I put forward conjectures asserting that, in all noncommutative field theories, (1) open Wilson lines and their descendants constitute a complete set of interpolating operators of "noncommutative dipoles", obeying dipole relation, (2) infrared dynamics of the noncommutative dipoles is dual to ultraviolet dynamics of the elementary noncommutative fields, and (3) open string field theory is a sort of noncommutative field theory, whose open Wilson lines are interpolating operators for closed strings. I substantiate these conjectures by various intuitive arguments and explicit computations of one- and two-loop Feynman diagrammatics.

1 Introduction and conjectures

The most salient feature of noncommutative field theories, as opposed to the conventional commutative field theories, is that physical excitations are described by a "noncommutative dipole" – weakly interacting, nonlocal object. One can visualize them as follows. Denote their center-of-mass momentum and dipole moment as $\mathbf{k}$ and ℓ, respectively. According to the "dipole" picture, originally developed in [1] and more recently reiterated in [2], the two quantities are related each other:

$$\ell^a = \theta^{ab}\mathbf{k}_b. \tag{1.1}$$

Here, θ^{ab} denotes the noncommutativity parameter[1]:

$$\left\{\mathbf{x}^a, \mathbf{x}^b\right\}_\star = i\theta^{ab}. \tag{1.2}$$

[1]Here, $\{\cdot\}$ refers to the Moyal commutator, defined in terms of the $\star$-product:

$$\{A(x_1)B(x_2)\}_\star := \exp\left(\frac{i}{2}\partial_1\wedge\partial_2\right)A(x_1)B(x_2) \qquad \text{where} \qquad \partial_1\wedge\partial_2 := \theta^{ab}\partial_1^a\partial_2^b.$$

Evidently, in the commutative limit, $\theta^{ab} \to 0$, the dipole shrinks in size and reduces to pointlike excitations.

Operationally, a noncommutative field theory is defined by an action functional of putative *elementary* fields, collectively denoted as Φ. Elementary quanta are created/annihilated by none other than the elementary fields, Φ. Apart from radiative corrections at higher-orders, in weak-coupling perturbation theory, they constitute a complete set of *pointlike* excitation spectrum. On the other hand, the above argument implies that, in a generic noncommutative field theory, in addition to point-like excitations, there ought to be dipole-like excitations as well. An immediate and interesting question is *"In what precise manner, does the noncommutative field theory produce dipole-like excitations?"*. A related questions are *"If the dipole-like excitations are present, does that imply that the elementary field Φ is non-unitary? If so, should one introduce a new field for each dipole-like excitation?"*. I claim that all these provoking questions can be answered, in noncommutative field theories, by introducing a set of operators, nicknamed as <u>open Wilson lines</u> (OWLs) [3, 4].

I found it illuminating to draw an analogy with the situation in QED. In QED, the spectrum includes, in addition to electron and proton quanta created/annihilated by respective fundamental fields, bound-states such as positronium or hydrogen atom, whose characteristic scale in wave functions or in "parton distribution functions", is set by the Bohr radius. This does not mean, however, that perturbative unitarity is violated, or a new field creating/annihilating the positronium or the hydrogen atom needs to be introduced. Rather, the bound-states are properly understood as *poles* of *two-particle*-irreducible Green functions, for which the only technical difficulty would be non-analyticity of the fixed-order perturbation theory near the two-particle threshold. The perturbation theory remains valid outside the threshold region, and hence is better than the approach relying only on dispersion relation, for which the discontinuity across the branch cut ought to be known for *all* energies. Morally speaking, I view the noncommutative dipole as counterpart of the QED bound-state, created/annihilated by the open Wilson line operators.

In answering the question posed above, I put forward the following conjectures:

• 1. <u>open Wilson lines</u>: in a generic noncommutative field theory, there *always* exist a special class of composite operators, open Wilson lines, $W_k[\Phi]$, and their descendants, $(\Phi W)_k\,[\Phi]$.

• 2. <u>noncommutative dipoles</u>: the open Wilson lines and their descendants constitute a complete set of interpolating operators for creating/annihilating the "noncommutative dipoles", obeying the dipole relation, equation (1.1).

• 3. UV-IR duality: infrared dynamics of the noncommutative dipoles, and hence the open Wilson lines, $W_k[C]$, is dual to ultraviolet dynamics of the elementary fields, Φ's.
• 4. Closed Strings from Open Strings: extended to string field theories, open Wilson lines made out of open string field are interpolating operators for closed strings, both on and off-shell.
I would like to motivate these conjectures from the following considerations. Once one recalls the Weyl-Moyal correspondence, the conjecture 1 is rather transparent: the open Wilson lines in Moyal formulation correspond, in Weyl formulation, to the familiar Wilson loops, *viz.* "master fields" in the large $N \sim \mathrm{Pf}\theta$ limit. The well-known fact that Wilson loops form a complete set of gauge-invariant operators in the large-N gauge theory motivates the first half of the conjecture 2. That the open Wilson lines ought to obey the dipole relation, equation (1.1), will be proven in Section 3. The conjecture 3 constitutes the most important feature regarding the long-distance spectrum and dynamics of generic noncommutative field theories. Evidently, the conjectured duality is strongly reminiscent of the "channel" duality between open and closed strings, and implies that the open Wilson lines are operators associated with "closed" string-like excitations in noncommutative field theories. The conjecture 4 asserts that the open string field theory is a sort of noncommutative field theory and that closed strings are describable entirely in terms of the open Wilson lines made out of open string field.

In this lecture, I take the simplest yet interacting noncommutative field theory, d-dimensional $\lambda[\Phi^3]_\star$ theory, and substantiate the above conjectures 1–4. Specifically, I present relevant results from computation of one- and two-loop effective action. I prove explicitly that the effective action is expressible entirely in terms of *scalar* open Wilson lines, and that interaction among the open Wilson lines is noncommutative and purely geometrical. Viewing the theory as level-zero truncation of the Witten's open string field theory, I also argue that the scalar open Wilson lines act precisely as the closed string field.

2 Flying noncommutative dipole

I begin with a physical situation from which intuitive picture of the noncommutative dipole is developable: Mott exciton in a strong magnetic field. As is well-known, in a strong magnetic field, low-energy excitation of electrons and holes is projected to the lowest Landau level so that, in the quasi-particle's Hamiltonian, the kinetic energy is negligible compared to the residual potential energy such as Coulomb interaction energy. An immediate question is whether the situation repeats for charge-neutral excitons, *viz.*

bound-state consisting of an equal number of electrons and holes. Following the pioneering works [1], I now prove that low-energy excitation of the Mott exciton consists only of rigid translational motion in the plane perpendicular to the magnetic field.

In the non-relativistic limit, the Hamiltonian of a charge-neutral exciton in the background of a uniform electric and magnetic field, $\mathbf{E}$ and $\mathbf{B}$, is given by

$$H = \frac{\left(\mathbf{p}_1 + e\mathbf{A}\left(\mathbf{r}_1\right)\right)^2}{2m} + \frac{\left(\mathbf{p}_2 - e\mathbf{A}\left(\mathbf{r}_2\right)\right)^2}{2m} + e\mathbf{E}\cdot\left(\mathbf{r}_1 - \mathbf{r}_2\right) + V\left(\left|\mathbf{r}_1 - \mathbf{r}_2\right|\right).$$

The velocity operator of the electron and the hole is given by $\mathbf{v}_{1,2} = \partial H/\partial \mathbf{p}_{1,2}$, and obeys the operator equations of motion:

$$m\frac{\mathrm{d}\mathbf{v}_1}{\mathrm{d}t} = -e\mathbf{E} + e\mathbf{B}\wedge\mathbf{v}_1 \qquad \text{and} \qquad m\frac{\mathrm{d}\mathbf{v}_2}{\mathrm{d}t} = +e\mathbf{E} - e\mathbf{B}\wedge\mathbf{v}_2.$$

One readily finds that the total momentum of the excition, $\mathbf{P} := m(\mathbf{v}_1 + \mathbf{v}_2) - e\mathbf{B}\wedge(\mathbf{r}_1 - \mathbf{r}_2)$ is conserved: $\mathrm{d}\mathbf{P}/\mathrm{d}t = [H, \mathbf{P}] = 0$. Take the symmetric gauge $\mathbf{A}(\mathbf{r}) = \frac{1}{2}\mathbf{B}\wedge\mathbf{r}$. The conserved total momentum is then given by

$$\mathbf{P} = \left(\mathbf{p}_1 + \mathbf{p}_2\right) - \frac{e}{2}\mathbf{B}\wedge\left(\mathbf{r}_1 - \mathbf{r}_2\right).$$

In terms of the center-of-mass coordinate, $\mathbf{R} = (\mathbf{r}_1 + \mathbf{r}_2)/2$ and the relative coordinate, $\ell := (\mathbf{r}_1 - \mathbf{r}_2)$, the exciton wave-function is given by

$$\Psi_{\mathbf{P}}\left(\mathbf{r}_1, \mathbf{r}_2\right) = \exp\left(i\mathbf{R}\cdot\mathbf{P} + \frac{ie}{2}\mathbf{R}\wedge\Delta\mathbf{x}\right)\psi_{\mathrm{P}}(\ell),$$

and, after straightforward algebra, one obtains the total Hamiltonian as

$$H = \frac{1}{B^2}\mathbf{P}\cdot\mathbf{E}\wedge\mathbf{B} - \frac{1}{2}M\frac{\mathbf{E}^2}{\mathbf{B}^2} + H_{\mathrm{rel}}, \qquad (M = 2m).$$

To make contact with noncommutative field theories more transparent, introduce noncommutativity parameter $\theta^{ab} = \left(\mathbf{F}^{-1}\right)^{ab}$, and inverse metric $G^{ab} = (-\theta^2)^{ab}$. One then readily derives the operator relations for the exciton center-of-mass velocity $\mathbf{V}$:

$$\mathbf{V}^a = \frac{\partial H}{\partial \mathbf{P}} = \theta^{ab}\mathbf{E}_b,$$

and for the exciton electric dipole moment $\mathbf{d} \equiv e\ell$:

$$\mathbf{d}^a \equiv \frac{\partial H}{\partial \mathbf{E}_a} = MG^{ab}\mathbf{E}_b + \theta^{ab}\mathbf{P}_b.$$

These are precisely the noncommutative dipole relation, equation (1.1).

Moreover, the relative Hamiltonian $H_{\rm rel}$ turns out identical with the standard Landau-level problem for a charged particle (with reduced mass). Thus, in the strong magnetic field limit, approximating the lowest Landau level wavefunction by Dirac delta-function, one obtains the excitation wave-function as:

$$\Psi_{\mathbf{P}}(\mathbf{R}, \mathbf{r}) \sim \exp\left(i\mathbf{R}\cdot\mathbf{P} + \frac{ie}{2}\mathbf{R}\wedge\ell\right)\delta\left(\mathbf{r} - \ell\right). \tag{2.1}$$

The wave-function equation (2.1) proves that the low-energy dynamics of the Mott exciton comprises of rigid translation, whose dipole moment is proportional to the the center-of-mass momentum.

3 Open wilson lines: How and why?

3.1 Open Wilson lines

How, in a given noncommutative field theory, are the open Wilson lines defined, and what are they for? I claim that the answer lies ultimately to the observation alluded above: noncommutative dipoles are present generically as part of theory's low-energy excitation. In noncommutative *gauge* theories, the answer also has to do with gauge-invariant, physical observables, so I will begin with this case first. In [3–5], it has been shown that (part of) gauge orbit is equivalent to the translation along the noncommutative directions. For example, in noncommutative $U(1)$ gauge theory, the gauge potential $\mathbf{A}_\mu(x)$ and the neutral scalar field $\Phi(x)$, both of which give rise to "noncommutative dipoles", transform in "adjoint" representation:

$$\begin{aligned}
\delta_\epsilon \mathbf{A}_\mu(x) &= i\int \frac{\mathrm{d}^2\mathbf{k}}{(2\pi)^2}\,\widetilde{\epsilon}(\mathbf{k})\Big[\Big(\mathbf{A}_\mu(\mathbf{x}+\theta\cdot\mathbf{k}) - \mathbf{A}_\mu(\mathbf{x}-\theta\cdot\mathbf{k})\Big) + i\mathbf{k}_\mu\Big]\mathrm{e}^{i\mathbf{k}\cdot\mathbf{x}}\\
\delta_\epsilon \Phi(x) &= i\int \frac{\mathrm{d}^2\mathbf{k}}{(2\pi)^2}\,\widetilde{\epsilon}(\mathbf{k})\Big[\Phi(\mathbf{x}+\theta\cdot\mathbf{k}) - \Phi(\mathbf{x}-\theta\cdot\mathbf{k})\Big]\mathrm{e}^{i\mathbf{k}\cdot\mathbf{x}},
\end{aligned} \tag{3.1}$$

where the infinitesimal gauge transformation parameter is denoted as

$$\epsilon(\mathbf{x}) = \int \frac{\mathrm{d}^2\mathbf{k}}{(2\pi)^2}\mathrm{e}^{i\mathbf{k}\cdot\mathbf{x}}\,\widetilde{\epsilon}(\mathbf{k}).$$

Because of the peculiarity, for fields transforming in "adjoint" representations under the noncommutative gauge group, there is *no* physical observables local in configuration space. The only physical observable one can construct turn out local in momentum space, and is referred as the "open Wilson line" operators [3–6] and their descendants defined on an open contour C:

$$W_{\mathbf{k}}[\mathbf{A}] = \mathcal{P}_{\rm t} \int \mathrm{d}^2\mathbf{x}\, \exp_\star \left(i \int_0^1 \mathrm{d}t\, \dot{\mathbf{y}} \cdot \mathbf{A}(x+\mathbf{y}) \right) \star \mathrm{e}^{i\mathbf{k}\cdot\mathbf{x}}, \qquad (3.2)$$

$$(\mathcal{O}W)_{\mathbf{k}}[\mathbf{A}] = \mathcal{P}_{\rm t} \int \mathrm{d}^2\mathbf{x} \left[\int_0^1 \mathrm{d}t \mathcal{O}(t) \exp_\star \left(i \int_0^1 \mathrm{d}t\, \dot{\mathbf{y}} \cdot \mathbf{A}(x+\mathbf{y}) \right) \right] \star \mathrm{e}^{i\mathbf{k}\cdot\mathbf{x}}.$$

Here, the $\star$-product is defined with respect to the base point $\mathbf{x}$ of the open contour C. Despite being defined over an open contour, the operator is gauge-invariant *provided* the momentum $\mathbf{k}$ is related to the geodesic distance $\mathbf{y}(1) - \mathbf{y}(0) := \Delta\mathbf{x}$ precisely by the "dipole relation", equation (1.1). In other words, in noncommutative gauge theory, the open Wilson lines (physical observables) are noncommutative dipoles, obeying the dipole relation equation (1.1) as an immediate consequence of the gauge invariance!

The open Wilson lines are actually ubiquitous and are present in generic noncommutative field theories, in which neither gauge invariance nor gauge field is present. This is because, as I have convinced you already, the dipole relation equation (1.1) ought to be a universal relation, applicable for *any* theories defined over noncommuative spacetime. In [7–9], I have shown that the scalar open Wilson line operators $W_{\mathbf{k}}[\Phi]$ and descendants $(\Phi^n W)_{\rm k}[\Phi]$ are given by:

$$W_{\mathbf{k}}[\Phi] := \mathcal{P}_{\rm t}\int \mathrm{d}^2\mathbf{x}\, \exp \left(ig \int_0^1 \mathrm{d}t |\dot{\mathbf{y}}(t)|\, \Phi(x+\mathbf{y}(t)) \right) \star \mathrm{e}^{i\mathbf{k}\cdot\mathbf{x}}$$

$$(\Phi^n W)_{\mathbf{k}}[\Phi] := \left(-i \frac{\partial}{\partial g} \right)^n W_{\mathbf{k}}[\Phi] \qquad (n = 1, 2, 3, \cdots), \qquad (3.3)$$

where λ is an appropriate coupling parameter.

Can one show that the dipole relation equation (1.1) is satisfied for scalar fields, wherein no gauge invariance is present? Consider the following set of operators, so-called Parisi operators [10]:

$$\mathcal{O}_n\left(x_1, \cdots, x_n; \mathbf{k}\right) = \int \mathrm{d}^2\mathbf{z}\, \Phi_1(x_1+\mathbf{z}) \star \Phi_2(x_2+\mathbf{z}) \star \cdots \star \Phi_n\left(x_n+\mathbf{z}\right) \star \mathrm{e}^{i\mathbf{k}\cdot\mathbf{x}},$$

viz. Fourier-transform of a string of elementary scalar fields, $\Phi_k(x)$ ($k = 1, 2, \cdots$). Take the one-point function:

$$G_1(x, \mathbf{k}) := \langle \mathcal{O}_2(x, \mathbf{k}) \rangle = \left\langle \int \mathrm{d}^2\mathbf{z}\, \Phi(\mathbf{z}) \star \Phi(x+\mathbf{z}) \star \mathrm{e}^{i\mathbf{k}\cdot\mathbf{x}} \right\rangle.$$

In terms of Fourier decomposition of the scalar field:

$$\Phi(x) = \int \frac{\mathrm{d}^2\mathbf{k}}{(2\pi)^2}\, \mathrm{e}^{i\mathbf{k}\cdot\mathbf{x}} \widetilde{\Phi}(\mathbf{k}),$$

I obtain that

$$G_1(x, \mathbf{k}) = \int \frac{\mathrm{d}^2\mathbf{l}}{(2\pi)^2} \widetilde{\Phi}(\mathbf{l})\widetilde{\Phi}(-\mathbf{l}+\mathbf{k}) \exp\left[i\mathbf{l}\cdot\left(x + \frac{1}{2}\theta\cdot\mathbf{k}\right)\right]\cdot$$

Consider "wave-packet" of the scalar particle, $\Phi(\mathbf{z}) = \Phi_0\delta^{(2)}(\mathbf{z})$ and $\Phi(x + \mathbf{z}) = \Phi_0\delta^{(2)}(x+\mathbf{z})$ so that $\widetilde{\Phi}(\mathbf{l}) = \Phi_0 \exp(i\mathbf{l}\cdot x)$. From the above equation, I then find that

$$G_1(x, \mathbf{k}) = \Phi_0^2\delta^{(2)}\left(x + \frac{1}{2}\theta\cdot\mathbf{k}\right)\cdot$$

Thus, I find that the stationary point of the correlator is given by $\Delta x^a \sim \theta^{ab}\mathbf{k}_b$, and hence precisely by the "dipole relation", equation (1.1).

I also claim that the open Wilson lines are a sort of "master fields". According to the Weyl-Moyal correspondence, generic noncommutative fields, be they the gauge field $\mathbf{A}$ or the scalar field Φ, are interpretable as $(N\times N)$ matrix-valued fields at $N\to\infty$ limit, living only on commutative directions (if there is any). Hence, from the latter formulation, one can construct Wilson loop operators as the large-N master fields:

$$W_{\mathbf{k}}\left[\widehat{A}\right] = \mathrm{Tr}\exp\left(i\mathbf{k}\cdot\widehat{\mathbf{A}}\right) \qquad \text{or} \qquad W_k\left[\widehat{\Phi}\right] = \mathrm{Tr}\exp\left(ik\widehat{\Phi}\right).$$

In fact, one can readily show that, once expanded around the noncommutative space, these Wilson loop operators turn into the aforementioned open Wilson lines.

3.2 Generalized star products

Computationally, the open Wilson lines originate from resummation of so-called generalized $\star$-products. As such, I will first indicate how the generalized $\star$-products are inherent to the defintion of the open Wilson lines.

Begin with the gauge open Wilson lines. For a straight contour, expanding equation (3.2) in successive powers of the gauge field $\mathbf{A}_\mu$, it was observed [11, 12] that generalized $\star$-product, $\star_n$, a structure discovered first in [13, 14], emerge:

$$W_{\mathbf{k}}[C] = \int \mathrm{d}^2\mathbf{x}\left[1 - (\partial\wedge\mathbf{A}) + \frac{1}{2!}\left(\partial\wedge\mathbf{A}\right)^2_{\star_2} + \cdots\right]\star \mathrm{e}^{i\mathbf{k}\cdot\mathbf{x}}. \qquad (3.4)$$

The generalized $\star_n$ product exhibits different algebraic structures from Moyal's $\star$-product. For instance, the first two, $\star_2, \star_3$ defined as

$$\begin{aligned}
\left[A(x_1)\,B(x_2)\right]_{\star_2} &:= \frac{\sin\left(\frac{1}{2}\partial_1\wedge\partial_2\right)}{\frac{1}{2}\partial_1\wedge\partial_2}A(x_1)\,B(x_2)\\
\left[A(x_1)\,B(x_2)\,C(x_3)\right]_{\star_3} &:= \left[\frac{\sin\left(\frac{1}{2}\partial_2\wedge\partial_3\right)}{\frac{1}{2}(\partial_1+\partial_2)\wedge\partial_3}\frac{\sin\left(\frac{1}{2}\partial_1\wedge(\partial_2+\partial_3)\right)}{\frac{1}{2}\partial_1\wedge(\partial_2+\partial_3)}\right.\\
&\qquad\left.+(1\leftrightarrow 2)\vphantom{\frac{1}{2}}\right]\\
&\qquad\times A(x_1)\,B(x_2)\,C(x_3)
\end{aligned}$$

show that the $\star_n$'s are commutative but non-associative. Despite these distinguishing features, I interpret Moyal's $\star$ product more fundamental than the generalized $\star_n$ products. The open Wilson line is defined in terms of path-ordered $\star$-product, and its expansion in powers of the gauge potential involves the generalized $\star_n$-product at each n-th order. As such, complicated $\star_n$ products arise upon expansion in powers of the gauge potential, and are attributible to dipole nature of the open Wilson line and the gauge invariance therein – each term in equation (3.4) is *not* gauge invariant, as the gauge transformation equation (3.1) mixes terms involving different $\star_n$'s. Indeed, the generalized $\star_n$ products are not arbitrary but obey recursive identities:

$$\begin{aligned}
i\left[\partial_x A\wedge\partial_x B\right]_{\star_2} &= \{A,B\}_\star\\
i\partial_x\wedge\left[A\,B\partial_x C\right]_{\star_3} &= A\star_2\{B,C\}_\star + B\star_2\{A,C\}_\star\,.
\end{aligned}$$

These identities are crucial for ensuring gauge invariance of the power-series expanded open Wilson line operator, equation (3.4).

I can show readily that the same sort of generalized $\star$-products also show up in the scalar open Wilson lines. Consider a simplified form of the scalar open Wilson line with an insertion of a local operator $\mathcal{O}$ at a location $\mathbf{R}$ on the Wilson line contour:

$$(\mathcal{O}_{\mathbf{R}}W)_{\mathrm{k}}[\Phi] := \mathcal{P}_{\mathrm{t}}\int \mathrm{d}^2\mathbf{x}\,\mathcal{O}(x+\mathbf{R})\star\exp\left(ig\int_0^1\mathrm{d}t|\dot{\mathbf{y}}(t)|\Phi\left(x+\mathbf{y}(t)\right)\right)\star\mathrm{e}^{i\mathbf{k}\cdot\mathbf{x}}.$$

Take, for simplicity, a *straight* Wilson line:

$$\mathbf{y}(t)=\mathbf{L}t \qquad \text{where} \qquad \mathbf{L}^a=\theta^{ab}\mathbf{k}_b := (\theta\cdot\mathbf{k})^a\,,\quad L:=|\mathbf{L}|,$$

corresponding to a *uniform* distribution of the momentum $\mathbf{k}$ along the Wilson line. As the path-ordering progresses to the right with increasing t, power-series expansion in $gL\Phi$ yields:

$$(\mathcal{O}_{\mathbf{R}}W)_{\mathrm{k}}[\Phi] = \int \mathrm{d}^2\mathbf{x}\mathrm{e}^{i\mathbf{k}\cdot\mathbf{x}} \star [\mathcal{O}(\mathbf{x}+\mathbf{R}) + igL\int_0^1 \mathrm{d}t\mathcal{O}(x+\mathbf{R})\star\Phi(x+\mathbf{L}t) + (igL)^2\int_0^1 \mathrm{d}t_1\int_{t_1}^1 \mathrm{d}t_2\mathcal{O}(\mathbf{x}+\mathbf{R})\star\Phi(x+\mathbf{L}t_1)\star\Phi(x+\mathbf{L}t_2) + \cdots\;].$$

I can evaluate each term, for instance, by Fourier-transforming $\mathcal{O}$ and Φ's:

$$\mathcal{O}(x) = \int\frac{\mathrm{d}^2\mathbf{k}}{(2\pi)^2}\widetilde{O}(\mathbf{k})\,T_{\mathbf{k}}, \quad \Phi(x) = \int\frac{\mathrm{d}^2\mathbf{k}}{(2\pi)^2}\widetilde{\Phi}(\mathbf{k})\,T_{\mathbf{k}} \quad \text{where} \quad T_{\mathbf{k}} = \mathrm{e}^{i\mathbf{k}\cdot\mathbf{x}},$$

taking the $\star$-products of the translation generators: $T_{\mathbf{k}}\star T_{\mathbf{l}} = \mathrm{e}^{\frac{i}{2}\mathbf{k}\wedge\mathbf{l}}\,T_{\mathbf{k}+\mathbf{l}}$, and then evaluating the parametric $t_1, t_2, \cdots$ integrals. Fourier-transforming back to the configuration space, after straightforward calculations, I obtain

$$\begin{aligned}(\mathcal{O}_{\mathbf{R}}W)_{\mathrm{k}}[\Phi] = &\int \mathrm{d}^2\mathbf{x}\mathrm{e}^{i\mathbf{k}\cdot\mathbf{x}}\mathcal{O}(x+\mathbf{R}) \\ &+ igL\int \mathrm{d}^2\mathbf{x}\,\mathrm{e}^{i\mathbf{k}\cdot\mathbf{x}}\left[\mathcal{O}(x+\mathbf{R})\,\Phi(x)\right]_{\star_2} \\ &+ \frac{1}{2!}(igL)^2\int \mathrm{d}^2\mathbf{x}\,\mathrm{e}^{i\mathbf{k}\cdot\mathbf{x}}\left[\mathcal{O}(x+\mathbf{R})\,\Phi(x)\Phi(x)\right]_{\star_3} + \cdots .\end{aligned} \tag{3.5}$$

I then observe that the products involved, $\star_2, \star_3, \cdots$, are precisely the *same* generalized $\star_n$ products as those appeared prominently in the *gauge* open Wilson line operators.

4 Free and interacting OWLs

To identify degrees of freedom associated with the open Wilson lines and to understand their spectrum and interaction, consider the $\lambda[\Phi^3]_\star$-theory, and study the effective action. The noncommutative Feynman rules of $\lambda[\Phi^3]_\star$-theory are summarized, in the background field method, by the following generating functional:

$$Z[\Phi_0] = Z_0[\Phi_0]\int\mathcal{D}\varphi\exp\left(-\int\mathrm{d}^dx\left[\frac{1}{2}\varphi(x)\mathcal{D}_{\Phi_0}\varphi(x) + \frac{\lambda}{3!}\varphi^3(x)\right]_\star\right). \tag{4.1}$$

Here, $\mathcal{D}_{\Phi_0} = (-\partial_x^2 + m^2 + \lambda\Phi_0(x))$, and $\Phi_0(x)$ and $\varphi(x)$ refer to the background and the fluctuation parts of the scalar field, Φ, respectively. Because of the noncommutativity of the $\star$-product, interactions are classifiable into planar and nonplanar ones. I focus on so-called nonplanar part of the one- and two-loop Feynman diagrams, and, as I am interested primarily in dynamics at long distance, on the low-energy and large noncommutativity limit:

$$\frac{p}{m} = \mathcal{O}\left(\epsilon^{+1}\right), \quad m^2\theta^{ab} = \mathcal{O}\left(\epsilon^{-2}\right), \quad \frac{\lambda\Phi}{m^2} = \mathcal{O}\left(\epsilon^{+1}\right) \quad \text{as} \quad \epsilon \to 0^+. \tag{4.2}$$

4.1 Free OWLs

Begin with effective action at one loop. The nonplanar part of the one-particle-irreducible N-point Green function is given by (see Fig. 1)

$$\Gamma_N\left(\{p_i\},\{q_j\}\right) = \hbar\left(-\frac{\lambda}{2}\right)^N \sum_{N_1+N_2=N} \frac{C_{\{N\}}}{(4\pi)^{d/2}} \int_0^\infty \frac{\mathrm{d}T}{T} T^{-\frac{d}{2}+N} \times \exp\left[-m^2 T - \frac{\ell^2}{4T}\right] J_{N_1}(\ell) J_{N_2}(-\ell).$$

Here, I have denoted an N-dependent combinatoric factor as $C\{_N\}$, divided N external momenta into two groups: $\{p_1, \cdots p_{N_1}\}$ and $\{q_1, \cdots, q_{N_2}\}$, and defined $k = \sum_{i=1}^{N_1} p_i = -\sum_{i=1}^{N_2} q_i$, and $\ell := \theta \cdot k$ (consistent with the noncommutative dipole relation). I have also defined $J_N(\ell)$ by

$$J_{N_1}\left(\ell,\{p_i\}\right) := \int_{-1/2}^{1/2} \mathrm{d}\tau_1 \cdots \mathrm{d}\tau_N \, \exp\left[-i\sum_{i=1}^{N_1} \tau_i p_i \cdot \ell - \frac{i}{2}\sum_{i<j=1}^{N_1} \epsilon\left(\tau_{ij}\right) p_i \wedge p_j\right],$$

where $\tau_{ij} \equiv (\tau_i - \tau_j)$, and similarly $J_{N_2}(-\ell, \{q_j\})$. They are precisely the momentum-space kernel of the generalized $*_N$ product.

In the limit equation (4.2), the T-moduli integral is evaluated accurately *via* the saddle-point approximation. Evidently, $T = |l|/2m$ at the saddle point. Then, the integral is evaluated as

$$\Gamma_N\left[\{p_i\},\{q_j\}\right] = \hbar\left(2\pi\frac{|\ell|}{m}\right)^{-\frac{d}{2}} \left(\frac{2\pi}{m|\ell|}\right)^{1/2} \mathrm{e}^{-m|\ell|} \times \sum_{N_1+N_2=N} C\{_N\}\left\{g^{N_1}|l|^{N_1} J_{N_1}(\ell)\right\}\left\{g^{N_2}|l|^{N_2} J_{N_2}(-\ell)\right\},$$

where $g \equiv -\lambda/4m$. The factorized expression permits exponentiation of the double-sum over N_1, N_2. Indeed, summing over $N = N_1 + N_2$, taking

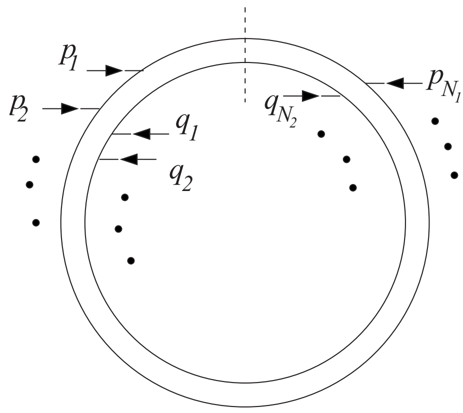


Fig. 1. One-loop Feynman diagram for N-point one-particle-irreducible Green function. The circumference of the vacuum diagram has a length T, while the relative position of each external line from the reference point (dashed vertical) is denoted τ_i's.

carefully into account of the combinatorial factors C_N, I find [7,8]

$$\Gamma = \frac{\hbar}{2}\int \frac{\mathrm{d}^d\ell}{(2\pi)^d} W(\ell)\,\mathcal{K}_{-d}(|\ell|)\,W(-\ell),$$

where

$$\begin{aligned}\mathcal{K}_{-d}(|\ell|) &= \left(2\pi\frac{|\ell|}{m}\right)^{-\frac{d}{2}}\left(\frac{2\pi}{m|\ell|}\right)\mathrm{e}^{-m|\ell|},\\ W(\ell) &= \sum_{N=0}^{\infty}\int\frac{\mathrm{d}^d p_1}{(2\pi)^d}\cdots\int\frac{\mathrm{d}^d p_N}{(2\pi)^d}(2\pi)^d\delta^d\left(p_1+\cdots+p_N-\ell\right)\\ &\quad\times\frac{1}{N!}(-g|\ell|)^N\left[\Phi(p_1)\cdots\Phi(p_N)\cdot J_N(\{p_i\},\ell)\right].\end{aligned}$$

In the last step, I utilized the result of the previous subsection that open Wilson lines are expandable in power-series of $\star_N$ products.

The $N_1 = N_2 = 1$ term was computed previously in [16], where infrared singular behavior of the result was interpreted as manifestation of the UV-IR duality. What remained not understood in the work of [16] was understanding reason behind the UV-IR duality. What I have shown above is that this is answerable by summing over N, *viz.* computing the full effective action, and that the UV-IR duality originates from noncommutative dipole degrees of freedom inherent in any noncommutative field theory.

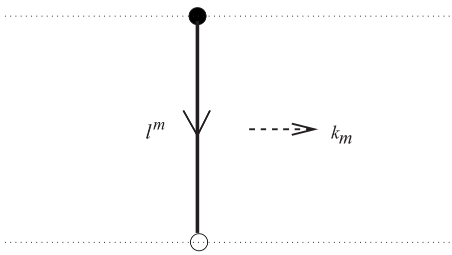


Fig. 2. Spacetime view of open Wilson line propagation. The noncommutativity is turned on the plane spanned by the two vectors, ℓ^m and k_m.

In fact, the one-loop effective action is expressible schematically as

$$\Gamma_2[W] = \frac{1}{2}\mathrm{Tr}_{\mathcal{H}_{\mathrm{dipole}}}\Big(\widehat{W}\cdot\mathbf{K}_2\cdot\widehat{W}\Big), \qquad\qquad \mathbf{K}_2 := \widehat{\mathcal{K}}_{-\frac{d}{2}}.$$

Here, I have used the Weyl-Moyal correspondence and expressed the open Wilson lines and kernel as operators defined on one-dipole Hilbert space $\mathcal{H}_{\mathrm{dipole}}$. Remarkably, the action takes strikingly the same form as the quadratic part of *matter* sector in Witten's cubic open string field theory!

4.2 Interacting OWLs

I next compute the nonplanar part of the two-loop effective action and show that it is expressible as cubic interaction of the open Wilson lines. Begin with the two-loop nonplanar contribution to the N-point one-particle-irreducible Green functions [17]. The Feynman diagram under consideration is depicted in Figure 3. Begin with constructing a *planar* two-loop vacuum diagram[2] constructed by joining three internal propagators *via* two $\lambda[\Phi^3]_\star$ interaction vertices. Denote the internal propagators in double-lines and label them as $a = 1$, 2, and 3 in Figure 3. Moduli parameters T_1, T_2, T_3 refer to the Feynman-Schwinger parameters of the three internal propagators, and range over the moduli space, $\mathcal{M}_{2-\mathrm{loop}} = [0,\infty)\otimes[0,\infty)\otimes[0,\infty)$. We then affix N external lines (background fields), distributed among the three internal propagators as N_a $(a = 1,2,3)$ so that $(N_1 + N_2 + N_3) = N$. Each group of external lines are further classifiable into those affixed from the inner and the outer boundaries. Sum over all possible insertion of the

[2]At two loop and beyond, vacuum diagrams are classifiable into a planar diagram and the rest, nonplanar diagrams. If the number of twist insertion is zero, the vacuum diagram is referred as planar. All other vacuum diagrams, with at least one insertion of the twist, are nonplanar ones. At one loop, by default, the vacuum diagram is planar.

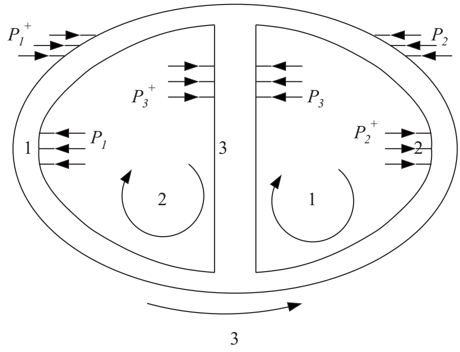


Fig. 3. Two-loop Feynman diagram for N-point Green function. The external lines and the vacuum diagram boundaries are labelled *dual* each other. The external momenta $p_i^{(a)}$, Feynman-Schwinger moduli parameters $\tau_i^{(a)}$, and Moyal's phase-factor signs $\nu_i^{(a)}$ are ordered from top to bottom. With this ordering convention, the $\tau's$ range over $0 < \tau_N < \cdots < \tau_1 < T$ for each connected side of the three internal propagators.

external lines is provided by integration over the moduli parameters $\tau_i^{(a)}$'s over $[0, T_a]$, and $\tau_{ij}^{(a)}$ refers to $\left(\tau_i^{(a)} - \tau_j^{(a)}\right)$.

Momenta of external lines attached at a-th internal propagator are labelled as $\{p_i^{(a)}\}$, where $a = 1, 2, 3$ and $i = 1, 2, \cdots, N_a$. I introduced total momentum injected on a-th internal propagator as:

$$P_a^{\pm} \equiv \sum_{i=1}^{N_a} \frac{1 \pm \nu_i^{(a)}}{2} p_i^{(a)} \qquad \text{and} \qquad P_a \equiv P_a^+ + P_a^-,$$

where $\pm$ refers to the inner and the outer boundaries, respectively, and $\nu_i^{(a)}$ takes ± 1 depending on whether the insertion is made from the "left" or the "right" side. I have also introduced the total momentum inserted on each *worldsheet boundary via* $k_{a+2} = P_a^+ + P_{a+1}^-$. Again, taking the limit equation (4.2) and consuming a lengthy computation, I have obtained:

$$\Gamma_{\rm N}\left(\left\{p_a^{(a)}\right\}\right) = \frac{\hbar^2\lambda^2}{24}\left(-\frac{\lambda}{2}\right)^N \sum_{\{N_a\}=0}^{N} \sum_{\{\nu\}} (2\pi)^d \delta\left(\sum_{a=1}^{3}\sum_{n=1}^{N_a} p_n^{(a)}\right) C_N\, \Gamma_\nu^{(\{N_a\})}.$$

Here, $C_{\{N\}}$ denotes a combinatoric factor, and

$$\Gamma^{(N_1,N_2,N_3)}_{\sigma,\nu} = \int_0^\infty \frac{\mathrm{d}T_1\mathrm{d}T_2\mathrm{d}T_3}{(4\pi)^d}\, \mathrm{e}^{F(T_1,T_2,T_3)}\Delta^{\frac{d}{2}}(T)\left(\prod_{a=1}^{3}\int_0^{T_a}\prod_{i=1}^{N_a}\mathrm{d}\tau_i^{(a)}\right)$$
$$\times \textstyle\prod_{a=1}^{3}\left(\sum_{N_a^+}\frac{T_a^{N_a^+}}{N_a^+!}J_{N_a^+}(+\alpha_a)\right)\left(\sum_{N_a^-}\frac{T_a^{N_a^-}}{N_a^-!}J_{N_a^-}(-\alpha_a)\right)$$

in which

$$\exp F(T_1,T_2,T_3) = \exp\left[-m^2(T_1+T_2+T_3) - \frac{\Delta}{4}\left(T_1\ell_1^2+T_2\ell_2^2+T_3\ell_3^2\right)\right],$$

$$\Delta^{-1}(T) := T_1T_2+T_2T_3+T_3T_1, \tag{4.3}$$

$$\widehat{J}_{(N_a^+,N_{a+1}^-)}(-\alpha_a,\alpha_{a+1}) = \exp\left[\frac{i}{2}t_at_{a+1}k_a\wedge k_{a+1}\right]\widetilde{J}_{N_a^+}\widetilde{J}_{N_{a+1}^-}$$

and $\widetilde{J}$'s are precisely the same as the one-loop $\star_N$ kernel. I have also introduced the following shorthand notations:

$$\begin{aligned} t_a &= \sqrt{\Delta}T_a, \qquad (t_1t_2+t_2t_3+t_3t_1=1)\\ \alpha_1 &= t_1(t_2\ell_2-t_3\ell_3), \quad \alpha_2 = t_2(t_3\ell_3-t_1\ell_1), \quad \alpha_3 = t_3(t_1\ell_1-t_2\ell_2). \end{aligned}$$

Geometrically, for any given nonnegative values of $\{t_a\}$, $\{\alpha_a\}$ split the triangle formed by $\{\ell_a\}$ into three pieces, *viz.*, $\ell_1 = \alpha_3-\alpha_2$, $\ell_2 = \alpha_1-\alpha_3$, $\ell_3 = \alpha_2-\alpha_1$.

The T-moduli integrals are computable by saddle-point conditions. $\partial F/\partial T_a = 0$ yields

$$\Delta^{-1} = \frac{L^2}{4m^2} \quad \text{and} \quad L \equiv |t_1l_1-t_2l_2| = |t_2l_2-t_3l_3| = |t_3l_3-t_1l_1. \tag{4.4}$$

They determine the "size" of the moduli and their relative "angles". Geometrically, the condition equation (4.4) demands that the angle between a pair of α's is $2\pi/3$. Moreover, the value of $F(\{T_a\})$ at the saddle point also has a simple geometric description: $F(\text{saddle}) = -m\big(|\alpha_1|+|\alpha_2|+|\alpha_3|\big)_{\text{saddle}}$. The crucial point is that, at the saddle point, I now have

$$T_a = \Delta^{-1/2}t_a = \frac{L}{2m}t_a = \frac{|\alpha_a|}{2m}, \tag{4.5}$$

Plugging equation (4.5) into equation (4.3), each factor sums up to an open Wilson line

$$\sum_{N_a^+}\left(-\frac{\lambda}{2}\right)^{N_a^+}\frac{T_a^{N_a^+}}{N_a^+!}J_{N_a^+}(\alpha_a) = \sum_{N_a^+}\left(-\frac{\lambda}{4m}|\alpha_a|\right)^{N_a^+}J_{N_a^+}(\alpha_a),$$

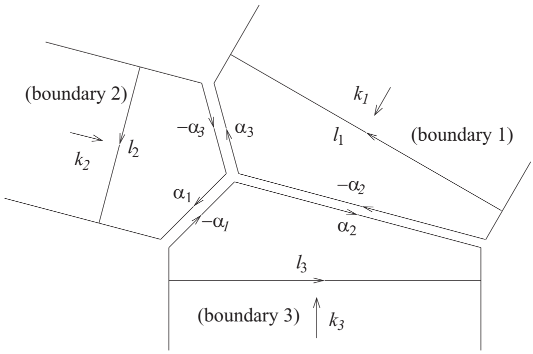


Fig. 4. Spacetime view of interaction among three open Wilson lines. At asymptotic region, the dipoles are described by straight open Wilson lines, while, at interaction region, the dipoles interact locally pairwise by "snapping" the open Wilson lines.

just like the one-loop case, except that the Wilson line contour is now "snapped"! I finally obtain the two-loop effective action as [18]

$$\begin{aligned}\Gamma[\ell] = \frac{1}{3}\lambda^2\hbar^2 \int \frac{\mathrm{d}^d k_1}{(2\pi)^d} \cdots \frac{\mathrm{d}^d k_3}{(2\pi)^d} \delta^{(d)}\left(k_1 + k_2 + k_3\right) \left(\frac{2m}{L}\right)^{d-3} \left(\frac{\delta T}{T}\right)^3 \\ \times \exp\left[-m\left(|\alpha_1| + |\alpha_2| + |\alpha_3|\right)\right] \exp\left(-\frac{i}{2}\sum_{a=1}^{3} \alpha_a \wedge \alpha_{a+1}\right) \\ \times \left[\widehat{W}\left(\alpha_1, -\alpha_2\right)\right]\left[\widehat{W}\left(\alpha_2, -\alpha_3\right)\right]\left[\widehat{W}\left(\alpha_3, -\alpha_1\right)\right].\end{aligned}$$

Again, the two-loop effective action is expressible schematically as:

$$\Gamma_3[W] = \frac{\lambda_c}{3}\mathrm{Tr}_{\mathcal{H}_{\mathrm{dipole}}}\mathbf{K}_3\left(\widehat{W}\,\widehat{\star}\,\widehat{W}\,\widehat{\star}\,\widehat{W}\right) \qquad \text{where} \qquad \lambda_c = (\lambda/2)^2, \tag{4.6}$$

where $\mathbf{K}_3$ represents a weight-factor over one-dipole Hilbert space $\mathcal{H}_{\mathrm{dipole}}$, and the $\widehat{\star}$-product refers to a newly emergent noncommutativity in the algebra of open Wilson lines. I note that the two-loop effective action, once expressed in terms of the snapped open Wilson lines, is remarkably similar to "half-string" picture of the cubic interaction in Witten's open string field theory, again restricted to the matter sector, with a minor difference that location of the triple interaction point is determined *dynamically* by the energy-momentum carried by the asymptotic dipoles. Note also that "soft-dilaton theorem" is obeyed – the interaction strength λ_c is proportional to *square* of the Φ-field coupling parameter, λ.

5 Closed strings out of open strings

The last conjecture of mine concerns extrapolation of the observations made in noncommutative field theories to string theories: *closed strings* are created/annihilated by "open Wilson lines" made out of open string fields [19].

The starting point would be that, at level-zero truncation and in the background of nonzero two-form potential B, Witten's cubic open string field theory is approximated by a noncommutative field theory of the open string tachyon with a cubic interaction. Expand the tachyon potential around local minimum. The resulting theory reduces precisely to the $\lambda[\Phi^3]_\star$-theory I have discussed at length already, but, quite importantly, with the mass-squared $m^2 \to \infty$. According to Sen's conjecture, excitation around the tachyon potential minimum ought to correspond to that of a closed string only and none of the open string. A point to be explored is "can one understand Sen's" conjecture even with zero-level truncation, *viz.* a noncommutative $\lambda[\Phi^3]_\star$ scalar field theory? I now claim that the answer to this question is affirmatively yes.

5.1 Open strings as miniature dipoles

The starting point is the observation that open strings in a constant B-field backgrounds are noncommutative dipoles, albeit of miniature size. In conformal gauge, an open string worldsheet action is

$$S_{\rm open} = \frac{1}{2\ell_{\rm st}^2}\int_\Sigma \left[\left(G_{mn}\delta^{\alpha\beta} + B_{mn}\epsilon^{\alpha\beta}\right)\partial_\alpha X^m \partial_\beta X^n\right] + \int_{\partial_{\rm L}\Sigma - \partial_{\rm R}\Sigma} A_m(X)\dot{X}^m,$$

where $\partial_{\rm L,R}\Sigma$ refers to left or right boundary of the worldsheet Σ. The B-field is locally exact, and can be gauged away *via* $U(1)$ transformation: $B_2 \to B_2 + {\rm d}\Lambda_1$, $A_1 \to A_1 - \ell_{\rm st}^{-2}\Lambda_1$. This results in

$$S_{\rm open} = \frac{1}{2\ell_{\rm st}^2}\int_\Sigma G_{mn}\left(\dot{X}^m\dot{X}^n - X'^m X'^n\right) + \int_{\partial_{\rm L}\Sigma - \partial_{\rm R}\Sigma} A_m(X)\dot{X}^m.$$

I now latticize the open string by two points, separated by worldsheet length $2\ell_{\rm st}$. Let $X(0,\tau) = X_{\rm L}(\tau)$ and $X(2\ell_{\rm st},\tau) = X_{\rm R}(\tau)$. The open string action is then approximatable as

$$S_{\rm open} \to \int {\rm d}\tau \left[\frac{1}{2}m\left(\dot{X}_{\rm L}^2 + \dot{X}_{\rm R}^2\right) - \frac{1}{2}m\omega^2(X_{\rm L} - X_{\rm R})^2\right] + \int {\rm d}\tau\left[qB_{mn}\left(\dot{X}_{\rm L}^m X_{\rm L}^n - \dot{X}_{\rm R}^m X_{\rm R}^n\right)\right],$$

where $m = \omega = 1/\ell_{\rm st}$ and $q = 1/\ell_{\rm st}^2$. Alternatively, in terms of center-of-mass and relative coordinates, $\mathbf{R} = (X_{\rm L} + X_{\rm R})/2$ and $\ell = (X_{\rm L} - X_{\rm R})$,

$$S_{\rm open} = \int {\rm d}\tau \left[\frac{1}{2} M \dot{\mathbf{R}}^2 + \frac{1}{2}\mu\dot{\ell}^2 - \frac{1}{2}\mu\omega^2\ell^2 + \frac{q}{2} B_{mn} \left(\dot{\mathbf{R}}^m \ell^n - \dot{\ell}^m \mathbf{R}^n \right) \right],$$

where $M = 2m$, $\mu = m/2$. In either form, it shows that the two-point latticized open string is literally identical to the Mott exciton or the noncommutative dipole. From the action, I also infer the boundary conditions:

$$\begin{aligned} m\omega^2 \left(X_{\rm L} - X_{\rm R}\right)_m + qF_{mn}\left(X_{\rm L}\right)\dot{X}_{\rm L}^n &= 0, \\ m\omega^2 \left(X_{\rm R} - X_{\rm L}\right)_m - qF_{mn}\left(X_{\rm R}\right)\dot{X}_{\rm R}^n &= 0. \end{aligned} \tag{5.1}$$

I now quantize the open string, *viz.* first-quantize the two point particles inside the dipole. In doing so, the boundary conditions equation (5.1) needs to be imposed as constraints. The dipole coordinates then obey exactly the same commutation relations as those obeyed by the Mott exciton:

$$[X_{\rm L}^m, X_{\rm L}^n] = +i\theta^{mn}, \qquad [X_{\rm R}^m, X_{\rm R}^n] = -i\theta^{mn}, \qquad [X_{\rm L}^m, X_{\rm R}^n] = 0,$$

or, equivalently, in the notation adopted for the Mott exciton,

$$[\mathbf{R}^m, \ell^n] = i\theta^{mn}, \qquad [\mathbf{R}^m, \mathbf{R}^n] = 0 = [\ell^m, \ell^n].$$

These commutation relations indicate that the open string is a noncommutative dipole, obeying the dipole relation: one-dipole Hilbert space is simply a tensor product of two one-particle Hilbert spaces: $\mathcal{H}_{\rm dipole} = \mathcal{H}_{\rm L} \otimes \mathcal{H}_{\rm R}$. In light of the latticization employed, one-dipole Hilbert space should be identified with first-quantized string Hilbert space.

5.2 Witten's $\star_{\rm w}$-product is Moyal's $\star_{\rm m}$-product

Utilizing the dipole picture of latticized open string, I now argue that the $\star$-product defining Witten's open string theory is identifiable as Moyal's $\star$-product. Actually, in the absence of constant B-field background, the isomorphism holds for string oscillator modes, but not for zero mode. Once B-field is turned on, the isomorphism hold exactly including the zero mode. This is the reason why we started with Witten's open string field theory in B-field background.

Schematically, Witten's $\star_{\rm w}$ product is defined on string Hilbert space $\mathcal{H}_{\rm L} \otimes \mathcal{H}_{\rm R}$ by

$$\left(|x_{\rm L}\rangle \otimes |x_{\rm R}\rangle\right) \star_{\rm w} \left(|y_{\rm L}\rangle \otimes |y_{\rm R}\rangle\right) \longrightarrow \langle x_{\rm R}|y_{\rm L}\rangle \left(|x_{\rm L}\rangle \otimes |y_{\rm R}\rangle\right). \tag{5.2}$$

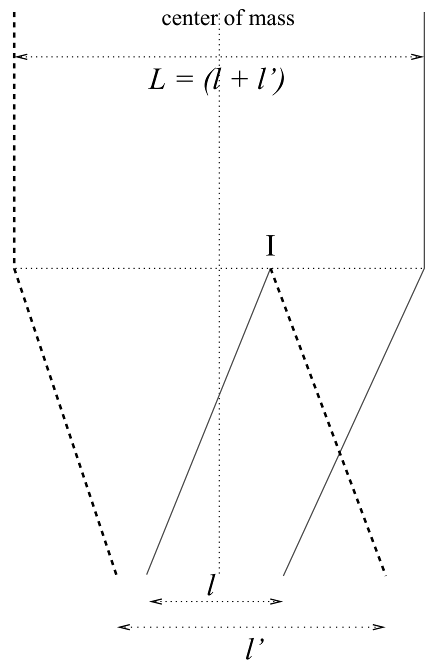


Fig. 5. Interaction of two miniature dipoles in support of equivalence between Moyal's $\star_{\rm m}$ product and Witten's $\star_{\rm w}$ product. Interaction point I can be anywhere and should be summed over.

Because of the dipole relation, I advocate the viewpoint treating the dipole configuration space $(X_{\rm L}, X_{\rm R})$ as a one-particle phase-space $(\mathbf{R}, \mathbf{P})$ associated with dipole's center-of-mass, where $\mathbf{P} = \theta^{-1} \cdot \ell$. Then, *via* Weyl-Moyal correspondence, Moyal's $\star_{\rm m}$ product in $(\mathbf{R}, \mathbf{P})$ space ought to be equivalent to matrix product in $(X_{\rm L}, X_{\rm R})$ space. It then follows that Moyal's $\star_{\rm m}$ product equals to *Fourier transform* of Witten's $\star_{\rm w}$ product.

Explicitly, start with Moyal's $\star_{\rm m}$-product in $(\mathbf{R}, \mathbf{P})$ space:

$$[A \star_{\rm m} B]\,(\mathbf{R}, \mathbf{P}) = A(\mathbf{R}, \mathbf{P}) \exp\left[\frac{i}{2}\left(\overleftarrow{\partial_{\mathbf{R}}} \cdot \overrightarrow{\partial_{\mathbf{P}}} - \overleftarrow{\partial_{\mathbf{P}}} \cdot \overrightarrow{\partial_{\mathbf{R}}}\right)\right] B(\mathbf{R}, \mathbf{P}). \quad (5.3)$$

I next "Fourier transform" with respect to $\mathbf{P}$ and express all in terms of dipole's relative distance ℓ's:

$$A(\mathbf{R}, \mathbf{P}) = \int {\rm d}\ell {\rm e}^{-i\mathbf{P}\cdot\ell} \widetilde{A}(\mathbf{R}, \ell) \quad \text{and} \quad B(\mathbf{R}, \mathbf{P}) = \int {\rm d}\ell {\rm e}^{-i\mathbf{P}\cdot\ell} \widetilde{B}(\mathbf{R}, \ell).$$

Substituting so, equation (5.3) is considerably simplified after a change of variables: $L = (\ell + \ell')$, $L' = (\ell' - \ell)$. Fourier transforming back the whole expression in equation (5.3) with respect to $\mathbf{P}$, I obtain [20]:

$$\left[\widetilde{A \star_{\rm m} B}\right](\mathbf{R}, L) = \int_{-\infty}^{+\infty} \frac{\mathrm{d}L'}{2} \widetilde{A}\,(\mathbf{R}+L/2, \mathbf{R}+L'/2)\, \widetilde{B}\,(\mathbf{R}+L/2, \mathbf{R}-L'/2)\,. \quad (5.4)$$

The emerging picture is that a miniature dipole $\widetilde{A}$ at center $\mathbf{R}$ and of length ℓ and another $\widetilde{B}$ at center $\mathbf{R}$ and of length ℓ' come into contact. When interacting, $\widetilde{A}, \widetilde{B}$ shift their centers to $\mathbf{R} + \ell'$ and $\mathbf{R} - \ell$, respectively. The final dipole $\widetilde{A \star_{\rm m} B}$ is then centered at $\mathbf{R}$ and of length $L = (\ell + \ell')$. See Figure 5. Evidently, the dipole interaction equation (5.4) defined *via* Moyal's $\star_{\rm m}$ product yields is algebraically equivalent to the string field interaction equation (5.2) defined *via* Witten's $\star_{\rm w}$ product.

5.3 Closed strings as OWLs

Recall that I have identified the scalar field Φ with the level-zero mode of the open string field. If I focus on low-energy and low-momentum excitation below a fixed cutoff, $p^2 \leq \Lambda^2$, as $m^2 \to \infty$, excitation of the Φ-quanta is entirely suppressed. This is clearly counterpart of half of Sen's conjecture: "around the tachyon potential minimum, there is no open string excitation". The regime $\Lambda^2 \ll m^2$ is also of considerable relevance to the effective action computation in $\lambda[\Phi^3]_\star$-theory, which I have not discussed at all so far. The point is that, in addition to the nonplanar diagram contribution, there also exists the planar diagram contribution to the effective action. The planar part is actually sensitive to the UV cutoff. If I identify the UV cutoff with the fixed cutoff Λ and take the conventional limit $m^2 \ll \Lambda^2$, the planar part of the effective action yields a sort of Coleman-Weinberg type potential (plus derivative corrections) – *viz.* exponentiation of the scalar field Φ takes place. On the other hand, if I take the opposite limit, $\Lambda^2 \ll m^2$, I have found that the planar diagram contribution turns remarkably into the same functional form as the nonplanar diagram contribution, except that (some of) the open Wilson lines carry nearly zero momentum. The point is that, even for planar diagrams, the scalar field Φ is exponentiated into open Wilson lines, albeit miniature ones, provided the cutoff condition obeys $\Lambda^2 \ll m^2$. While quite bizzare from the standard quantum field theory viewpoint, to our delight, this cutoff condition is precisely what is dictated by Witten's open string field theory!

The other half of Sen's conjecture – closed string out of open string tachyon vacuum – is then readily inferred from the results of previous sections. The open Wilson line formed out of the tachyon field Φ is precisely the interpolating operator creating and annihilating a closed string. The fact that open Wilson lines are Moyal formulation counterpart of the

Wilson loop in Weyl formulation adds an another supporting evidence for this claim. There is one peculiar aspect, though. First of all, the spacetime structure of the open Wilson lines is literally open, *viz.* the two ends are situated at distinct points in the target space. Moreover, the cubic interaction of the open Wilson lines, equation (4.6), involves newly emergent $\widehat{\star}$-product. As both are the aspects inherently associated with traditional open strings, one might feel suspicious to my conjecture of identifying the open Wilson lines as closed strings. I claim that a resolution can be drawn from the well-known fact that closed string is formed by joining two ends of open string(s). In the absence of the two-form potential, $B_{mn} = 0$, size of the open string is characteristically of string scale, and is too small to be probed by the level-zero truncated tachyon field. If the two-form potential is nonzero, $B_{mn} \neq 0$, the open string is polarized to a size much bigger than the string scale, and behaves essentially like a rigid rod. Because of that, joining and splitting of the two end of open string(s) would never form a closed string. In other words, open Wilson lines are precisely what the open strings can do the best for forming closed strings out of themselves! Reverting the logic, utility of turning on the B-field and hence noncommutativity for the open string is to render closed strings as much the same as open strings. That the open Wilson lines interaction is governed by a newly emergent $\widehat{\star}$-product (see Eq. (4.6)) would then constitute a nontrivial prediction of the conjectures I put forward [19].

This work was supported in part by BK-21 Initiative in Physics (SNU – Project 2), KOSEF Interdisciplinary Research Grant 98-07-02-07-01-5 and KOSEF Leading Scientist Program.

References

[1] N. Read, *Semicond. Sci. Technol.* **9** (1994) 1859; *Surf. Sci.* **361** (1996) 7; V. Pasquier, (unpublished); R. Shankar and G. Murthy, *Phys. Rev. Lett.* **79** (1997) 4437; D.-H. Lee, *Phys. Rev. Lett.* **80** (1998) 4745; V. Pasquier and F.D.M. Haldane, *Nucl. Phys. B* **516** [FS] (1998) 719; A. Sterm, B.I. Halperin, F. von Oppen and S. Simon, *Phys. Rev. B* **59** (1999) 12547.

[2] D. Bigatti and L. Susskind, *Phys. Rev. D* **66** (2000) 066004 [hep-th/9908056].

[3] S.-J. Rey and R. von Unge, *Phys. Lett. B* **499** (2001) 215 [hep-th/0007089].

[4] S.R. Das and S.-J. Rey, *Nucl. Phys. B* **590** (2000) 453 [hep-th/0008042].

[5] D.J. Gross, A. Hashimoto and N. Itzhaki, *Adv. Theor. Math. Phys.* **4** (2000) 893 [hep-th/0008075].

[6] N. Ishibashi, S. Iso, H. Kawai and Y. Kitazawa, *Nucl. Phys. B* **573** (2000) 573 [hep-th/9910004].

[7] Y. Kiem, S.-J. Rey, H.-T. Sato and J.-T. Yee, *Phys. Rev. D* **65** (2002) 026002 [hep-th/0106121].

[8] Y. Kiem, S.-J. Rey, H. Sato and J.-T. Yee, *Eur. Phys. J. C* **22** (2002) 757 [hep-th/0107106].

[9] Y. Kiem, S.-J. Rey, H.-T. Sato and J.-T. Yee, *Eur. Phys. J. C* **22** (2002) 781.

[10] G. Parisi, *Phys. Lett. B* **112** (1982) 463.

[11] T. Mehen and M.B. Wise, *J. High-Energy Phys.* **0012** (2000) 008 [`hep-th/0010204`].

[12] H. Liu [`hep-th/0011125`].

[13] M.R. Garousi, *Nucl. Phys. B* **579** (2000) 209 [`hep-th/9909214`].

[14] H. Liu and J. Michelson [`hep-th/0008205`].

[15] H. Liu [`hep-th/0011125`].

[16] S. Minwalla, M. Van Raamsdonk and N. Seiberg, *JHEP* **0002** (2000) 020; M. Van Raamsdonk and N. Seiberg, *JHEP* **0003** (2000) 035.

[17] Y. Kiem, S.-S. Kim, S.-J. Rey and H.-T. Sato [`hep-th/0110066`].

[18] Y. Kiem, S. Lee, S.-J. Rey and H.-T. Sato, *Phys. Rev. D* **65** (2002) 046003 [`hep-th/0110215`].

[19] S.-J. Rey, *Noncommutative Closed Strings out of Noncommutative Open Strings*, to appear.

[20] S.-J. Rey, unpublished note (2001). See also I. Bars and S.-J. Rey, *Phys. Rev. D* **64** (2001) 046005.

SEMINAR 6

OPEN-STRING MODELS WITH BROKEN SUPERSYMMETRY

A. SAGNOTTI

Dipartimento di Fisica,
Università di Roma "Tor Vergata",
I.N.F.N., Sezione di Roma II,
Via della Ricerca Scientifica 1,
00133 Roma, Italy

Contents

OPEN-STRING MODELS WITH BROKEN SUPERSYMMETRY

A. Sagnotti

Abstract

I review the salient features of three classes of open-string models with broken supersymmetry. These suffice to exhibit, in relatively simple settings, the two phenomena of "brane supersymmetry" and "brane supersymmetry breaking". In the first class of models, to lowest order supersymmetry is broken both in the closed and in the open sectors. In the second class of models, to lowest order supersymmetry is broken in the closed sector, but is *exact* in the open sector, at least for the low-lying modes, and often for entire towers of string excitations. Finally, in the third class of models, to lowest order supersymmetry is *exact* in the closed (bulk) sector, but is broken in the open sector. Brane supersymmetry breaking provides a natural solution to some old difficulties met in the construction of open-string vacua.

1 Broken supersymmetry and type-0 models

In this talk I would like to review the key features of some open-string models with broken supersymmetry constructed in [1–4]. These models may be derived in a systematic fashion from corresponding models of oriented closed strings [5], and once more display a surprising richness compared to them. Since the relevant techniques have been discussed at length in the original papers, I will not present any explicit derivations. Rather, referring to some of the resulting vacuum amplitudes, I will try to illustrate how supersymmetry can be broken at tree level in the bulk, on some branes or everywhere.

Closed-string models with broken supersymmetry were among the first new examples considered in the last decade. In particular, the type-0 models [6] provided the first non-trivial instances of modified GSO projections

This work is supported by EU grants HPRN-CT-2000-00122 and HPRN-CT-2000-00148, and by the INTAS contract 55-1-590.

compatible with modular invariance. In order to describe their partition functions, I will begin by introducing some notation that will be used repeatedly in the following, defining the four level-one $SO(8)$ characters

$$O_8 = \frac{\vartheta_3^4 + \vartheta_4^4}{2\eta^4}, \quad V_8 = \frac{\vartheta_3^4 - \vartheta_4^4}{2\eta^4},$$
$$S_8 = \frac{\vartheta_2^4 - \vartheta_1^4}{2\eta^4}, \quad C_8 = \frac{\vartheta_2^4 + \vartheta_1^4}{2\eta^4}, \tag{1.1}$$

where the ϑ_i are Jacobi theta functions and η is the Dedekind function. In terms of these characters, and leaving aside the contribution of the eight transverse bosonic coordinates, the type II models are described by

$$\mathcal{T}_{\rm IIA} = (V_8 - S_8)(\bar{V}_8 - \bar{C}_8), \tag{1.2}$$
$$\mathcal{T}_{\rm IIB} = |V_8 - S_8|^2, \tag{1.3}$$

while the type-0A and type-0B models are described by

$$\mathcal{T}_{\rm 0A} = |O_8|^2 + |V_8|^2 + S_8\bar{C}_8 + C_8\bar{S}_8, \tag{1.4}$$
$$\mathcal{T}_{\rm 0B} = |O_8|^2 + |V_8|^2 + |S_8|^2 + |C_8|^2. \tag{1.5}$$

In these expressions, the characters (O_8, V_8, S_8, C_8) depend on $q=\exp(i2\pi\tau)$, with τ the modulus of the torus, while their conjugates depend of $\bar{q}$. All these characters have power series expansions of the type

$$\chi(q) = q^{h-c/24} \sum_{n=0}^{\infty} d_n\, q^n, \tag{1.6}$$

where the d_n are integers. The low-lying spectra, essentially manifest in this notation, include in all cases the universal triple ($g_{\mu\nu}$, $B_{\mu\nu}$, ϕ). The corresponding states fill a generic transverse matrix, the direct product of the ground states of the V_8 module and of its conjugate $\bar{V}_8$. In addition, the type-IIA superstring has a Majorana gravitino and a Majorana spinor from the NS-R and R-NS sectors and a vector and a three-form from the R-R sector. The fermions result from pairs of Majorana-Weyl spinors of opposite chiralities, ground states of $V_8\bar{C}_8$ and $S_8\bar{V}_8$, while the nature of the R-R bosons is determined by the direct product of the two inequivalent spinor representations of $SO(8)$. These are the ground states of S_8 and $\bar{C}_8$, and the product $8_{\rm s} \times 8_{\rm c}$ indeed decomposes into a vector and a three-form. A similar reasoning shows that the type-IIB superstring has a complex Majorana-Weyl gravitino and a complex Weyl fermion from the NS-R and R-NS sectors, and a scalar, a two-form and a self-dual four-form from the R-R sector. The spectra of the type-0 models are purely bosonic, and can be essentially deduced from these. Aside from the universal triple ($g_{\mu\nu}$, $B_{\mu\nu}$, ϕ), their low-lying excitations include a tachyon, the ground state of the $O_8\bar{O}_8$ sector, while their R-R sectors are two copies of the previous ones, and include a pair of vectors and a pair of three-forms for the 0A model, and a pair of scalars, a pair of two-forms and an unconstrained four-form for the 0B model. Both type-0 models are clearly non-chiral, and are thus free of gravitational anomalies.

Let us now turn to the open descendants of the type-0 models. Their structure is essentially determined by the Klein bottle projection. Leaving aside the contributions of the transverse bosons, the conventional choice, originally discussed in [1], corresponds to

$$\mathcal{K}_{0\mathrm{A}} = \frac{1}{2}(O_8 + V_8), \tag{1.7}$$

$$\mathcal{K}_{0\mathrm{B}} = \frac{1}{2}(O_8 + V_8 - S_8 - C_8). \tag{1.8}$$

It eliminates the NS-NS two-form $B_{\mu\nu}$, but does not affect the tachyon. The effect of $\mathcal{K}$ can be simply summarized recalling that a *positive* (*negative*) sign implies a *symmetrization* (*antisymmetrization*) of the sectors fixed under left-right interchange. This unoriented projection is frequently called Ω. For instance, $\mathcal{K}_{0\mathrm{B}}$ symmetrizes the two NS-NS sectors, described by $|O_8|^2$ and $|V_8|^2$, and antisymmetrizes the two R-R sectors, eliminating the NS-NS $B_{\mu\nu}$ and leaving only a pair of R-R two-forms. In addition, the projected spectrum generally includes invariant combinations of all pairs of sectors interchanged by Ω, that do not contribute to $\mathcal{K}$. Thus, the low-lying spectrum of the projected 0A model includes also a R-R vector and a R-R three-form. As usual, the characters in these direct-channel Klein-bottle amplitudes depend on $q\bar{q} = \exp(-4\pi\tau_2)$.

The open sector of the 0A model, described by [1]

$$\mathcal{A}_{0\mathrm{A}} = \frac{n_\mathrm{B}^2 + n_\mathrm{F}^2}{2}(O_8 + V_8) \ - \ n_\mathrm{B} n_\mathrm{F}(S_8 + C_8), \tag{1.9}$$

$$\mathcal{M}_{0\mathrm{A}} = - \ \frac{n_\mathrm{B} + n_\mathrm{F}}{2}\hat{V}_8 \ - \ \frac{n_\mathrm{B} - n_\mathrm{F}}{2}\hat{O}_8, \tag{1.10}$$

is not chiral and involves *two* different "real" charges, corresponding to orthogonal or symplectic groups. These enter the partition functions *via* the dimensions, here n_B and n_F, of the corresponding fundamental representations, and in general are subject to linear relations originating from (massless) tadpole conditions. The low-lying modes are simply identified from the contributions of $\mathcal{A}$ and $\mathcal{M}$. For instance, the model contains two sets of $n_\mathrm{B}(n_\mathrm{B}-1)/2$ and $n_\mathrm{F}(n_\mathrm{F}-1)/2$ vectors (corresponding to V_8), enough to fill the adjoint representations of a pair of orthogonal groups, tachyons (corresponding to O_8) in doubly (anti)symmetric representations and fermions (corresponding to the R characters S_8 and C_8) in bi-fundamental representations. This description of open-string spectra is also useful in Conformal Field Theory, where it provides a convenient encoding of the spectrum of boundary operators in a generating function of their multiplicities. In this case, if one insists on demanding the cancellation of all NS-NS tadpoles, the result is the family of gauge groups $SO(n_{\rm B}) \times SO(n_{\rm F})$, with $n_{\rm B} + n_{\rm F} = 32$. Here $\mathcal{A}$ depends on $(q\bar{q})^{1/4} = \exp(-\pi\tau_2)$. On the other hand, $\mathcal{M}$ depends on $-(q\bar{q})^{1/4} = \exp(-\pi\tau_2 + i\pi)$, but the "hatted" characters are redefined by suitable phases, and are thus real.

The open sector of the 0B model, described by [1]

$$\mathcal{A}_{0\rm B} = \frac{n_{\rm o}^2 + n_{\rm v}^2 + n_{\rm s}^2 + n_{\rm c}^2}{2} V_8 \ + \ (n_{\rm o} n_{\rm v} + n_{\rm s} n_{\rm c}) O_8 \\ - \ (n_{\rm v} n_{\rm s} + n_{\rm o} n_{\rm c}) S_8 \ - \ (n_{\rm v} n_{\rm c} + n_{\rm o} n_{\rm s}) C_8, \tag{1.11}$$

$$\mathcal{M}_{0\rm B} = - \ \frac{n_{\rm o} + n_{\rm v} + n_{\rm s} + n_{\rm c}}{2} \hat{V}_8, \tag{1.12}$$

involves four different "real" charges, and is *chiral* but free of anomalies, as a result of the R-R tadpole conditions $n_{\rm o} = n_{\rm v}$ and $n_{\rm s} = n_{\rm c}$. All irreducible gauge and gravitational anomalies cancel as a result of the R-R tadpole conditions, while the residual anomaly polynomial requires a generalized Green-Schwarz mechanism [2, 28]. If one insists on demanding the cancellation of all NS-NS tadpoles, not related to anomalies [8] as the previous ones, one obtains the family of gauge groups $SO(n_{\rm o}) \times SO(n_{\rm v}) \times SO(n_{\rm s}) \times SO(n_{\rm c})$, with $n_{\rm o} + n_{\rm v} + n_{\rm s} + n_{\rm c} = 64$. Despite their apparent complication, these open-string models are actually simpler than the previous ones, since the 0B torus amplitude corresponds to the "charge-conjugation" modular invariant. This circumstance implies a one-to-one correspondence between types of boundaries and types of bulk sectors typical of the "Cardy case" of boundary CFT [9]. In equivalent terms, this model has four types of boundary states, in one-to-one correspondence with the chiral sectors of the bulk spectrum. The boundary states of the 0A model are a bit subtler, since they are proper combinations of these that do not couple to the R-R states, that cannot flow in the transverse channel compatibly with 10D Lorentz invariance. Indeed, the product of the two spinor representations $8_{\rm s}$ and $8_{\rm c}$ does not contain the identity, and consequently a right-moving S state cannot reflect into a left-moving C state at a Lorentz-invariant boundary.

The modified Klein bottle projection

$$\mathcal{K}'_{0\rm B} = \frac{1}{2}(-O_8 + V_8 + S_8 - C_8), \tag{1.13}$$

first proposed in [2] as an amusing application of the results of [7], removes the tachyon and the NS-NS two-form $B_{\mu\nu}$ from the closed spectrum, and leaves a *chiral* unoriented closed spectrum that comprises the $(g_{\mu\nu}, \phi)$ NS-NS pair, together with a two-form, an additional scalar and a self-dual four-form from the R-R sectors. The resulting open spectrum, described by

$$\begin{aligned} \mathcal{A}'_{0\mathrm{B}} &= -\frac{n^2 + \bar{n}^2 + m^2 + \bar{m}^2}{2} C_8 + (n\bar{n} + m\bar{m}) V_8 \\ &\quad + (n\bar{m} + m\bar{n}) O_8 - (mn + \bar{m}\bar{n}) S_8, \end{aligned} \tag{1.14}$$

$$\mathcal{M}'_{0\mathrm{B}} = \frac{m + \bar{m} - n - \bar{n}}{2} C_8, \tag{1.15}$$

involves the “complex” charges of a pair of unitary groups, subject to the R-R tadpole constraint $m - n = 32$, that eliminates all (non-Abelian) gauge and gravitational anomalies. The notation resorts to pairs of multiplicities [1], say m and $\bar{m}$, to emphasize the different roles of the fundamental and conjugate fundamental representations of a unitary group $U(m)$. The tadpole conditions identify the numerical values of m and $\bar{m}$, but once again one can read the low-lying spectrum directly and conveniently from the amplitudes written in this form. The choice $n = 0$ selects a $U(32)$ gauge group, with a spectrum that is free of tachyons both in the closed and in the open sectors. All irreducible (non-Abelian) gauge and gravitational anomalies cancel, while the residual anomaly polynomial requires a generalized Green-Schwarz mechanism [28]. On the other hand, the $U(1)$ factor is anomalous, and is thus lifted by a ten-dimensional generalization of the mechanism of [10], so that the effective gauge group of this model is $SU(32)$. Here one does not have the option of eliminating all the NS-NS tadpoles, and as a result a dilaton potential is generated.

These models have also been studied in some detail in [11], first with the aim of connecting them to the bosonic string, and more recently with the aim of relating them to (non-supersymmetric) reductions of M theory. This last approach goes beyond the perturbative analysis, and therefore has the potential of discriminating between the various options. According to [11], both types of 0B descendants admit a non-perturbative definition, while the 0A descendants do not. It would be interesting to take a closer look at this relatively simple model and try to elicit some manifestation of this phenomenon.

Let us now spend a few words to summarize the key features of these descendants, where supersymmetry is broken both in the closed and in the open sectors. Whereas the first two models have tachyons both in the closed and in the open sectors, the last results from a non-tachyonic brane configuration of impressive simplicity. This feature actually extends to lower-dimensional compactifications, as first shown by Angelantonj [12]. These type 0 models, and in particular the non-tachyonic one, have interesting applications [13] in the framework of the AdS/CFT correspondence [14].

A simple generalisation of this setting allows one to describe the branes allowed in these ten-dimensional models and in their "parent" oriented closed models. These results, originally obtained by a number of authors, can be efficiently described in this formalism as in [16].

2 Scherk-Schwarz deformations and brane supersymmetry

We may now turn to the second class of models. These rest on elegant extensions of the Kaluza-Klein reduction, known as Scherk-Schwarz deformations [17], that allow one to induce the breaking of supersymmetry from the different behaviors of fermionic and bosonic modes in the internal space. This setting, as adapted to the entire perturbative spectra of models of oriented closed strings in [18], is the starting point for the constructions in [3, 19]. I will confine my attention to particularly simple examples, related to the reduction of the type IIB superstring on a circle of radius R where the momenta or the windings are subjected to 1/2-shifts, compatibly with modular invariance, in such a way that all massless fermions are lifted in mass. In these models, supersymmetry is completely broken, but several more complicated open-string models with partial breaking of supersymmetry are discussed in [3, 22].

This Scherk-Schwarz deformation generically introduces tachyons, in the first case *(momentum shifts)* for $R < \sqrt{\alpha'}$, and in the second case *(winding shifts)* for $R > \sqrt{\alpha'}$. The former choice [3,19] is essentially a Scherk-Schwarz deformation of the low-energy field theory, here lifted to the entire string spectrum. On the other hand, the latter [3] is a bit more subtle to interpret from a field theory perspective, and indeed the resulting deformation of the spectrum is removed in the limit of small radius R, that strictly speaking is inaccessible to the field theory description. Naively, in the first case the open descendants should not present new subtleties. One would expect that the momentum deformations be somehow inherited by the open spectrum, and this is indeed what happens. On the other hand, naively the open spectrum should be insensitive to winding deformations, simply because the available Neumann strings have only momentum excitations. Here the detailed analysis settles the issue in an interesting way. The open spectrum is indeed affected, although in a rather subtle way, and supersymmetry is effectively broken again, at the compactification scale $1/R$, but is *exact* for the massless modes. In order to appreciate this result, let us present the closed-string amplitudes for the two cases, here written in the Scherk-Schwarz basis,

$$\begin{aligned}\mathcal{T}_1 = {} & Z_{m,2n}(V_8\bar{V}_8 + S_8\bar{S}_8) + Z_{m,2n+1}(O_8\bar{O}_8 + C_8\bar{C}_8) \\ & - Z_{m+1/2,2n}(V_8\bar{S}_8 + S_8\bar{V}_8) - Z_{m+1/2,2n+1}(O_8\bar{C}_8 + C_8\bar{O}_8)\end{aligned} \quad (2.1)$$

and

$$\begin{aligned}\mathcal{T}_2 &= Z_{2m,n}(V_8\bar{V}_8 + S_8\bar{S}_8) + Z_{2m+1,n}(O_8\bar{O}_8 + C_8\bar{C}_8)\\ &\quad - Z_{2m,n+1/2}(V_8\bar{S}_8 + S_8\bar{V}_8) - Z_{2m+1,n+1/2}(O_8\bar{C}_8 + C_8\bar{O}_8),\end{aligned} \tag{2.2}$$

and the corresponding Klein bottle projections

$$\mathcal{K}_1 = \frac{1}{2}\,(V_8 - S_8)\; Z_m, \tag{2.3}$$

$$\mathcal{K}_2 = \frac{1}{2}\,(V_8 - S_8)\; Z_{2m} + \frac{1}{2}\,(O_8 - C_8)\; Z_{2m+1}. \tag{2.4}$$

In these expressions, $Z_{m,n}$ denotes the usual Narain lattice sum for the circle

$$Z_{m,n} = \sum_{m,n} q^{\frac{\alpha'}{4}\left(\frac{m}{R}+\frac{nR}{\alpha'}\right)^2}\;\bar{q}^{\frac{\alpha'}{4}\left(\frac{m}{R}-\frac{nR}{\alpha'}\right)^2}, \tag{2.5}$$

while, for instance, $Z_{2m,n}$ denotes the sum restricted to even momenta. In a similar fashion, Z_m in (2.4) denotes the restriction of the sum to the momentum lattice.

In writing the corresponding open sectors, I will now eliminate several contributions, restricting the charge configurations in such a way that no tachyons are introduced. This is, to some extent, in the spirit of the previous discussion of the 10D $U(32)$ model, but here one can also cancel all NS-NS tadpoles. Moreover, I will take into account the infrared subtlety discussed in [20], that in the model with winding shifts leads the emergence of additional tadpoles in the singular limit $R \to 0$, where whole towers of massive excitations collapse to zero mass. With this proviso, the corresponding open spectra are described by

$$\begin{aligned}\mathcal{A}_1 &= \frac{n_1^2+n_2^2}{2}(V_8 Z_m - S_8 Z_{m+1/2}) + n_1 n_2 (V_8 Z_{m+1/2} - S_8 Z_m),\\ \mathcal{M}_1 &= -\frac{n_1+n_2}{2}(\hat{V}_8 Z_m - \hat{S}_8 Z_{m+1/2}),\end{aligned} \tag{2.6}$$

and

$$\begin{aligned}\mathcal{A}_2 &= \frac{n_1^2+n_2^2}{2}(V_8 - S_8) Z_m + n_1 n_2 (O_8 - C_8) Z_{m+1/2},\\ \mathcal{M}_2 &= -\frac{n_1+n_2}{2}\left(\hat{V}_8 - (-1)^m \hat{S}_8\right) Z_m.\end{aligned} \tag{2.7}$$

As anticipated, the first model, with $n_1 + n_2 = 32$, is essentially a conventional Scherk-Schwarz deformation of the type-I superstring. It can also describe the type I spectrum at a finite temperature related to the internal radius R. On the other hand, the second model, where $n_1 = n_2 = 16$, is more interesting, and displays the first novel phenomenon reviewed here, "brane supersymmetry": although supersymmetry is broken at the compactification scale by the Scherk-Schwarz deformation, the massless modes of the open sector fill complete supersymmetry multiplets. I would like to stress that the breaking of supersymmetry in the massive spectrum of the second model can also be regarded as a deformation, now resulting from the unpairing of the Chan-Paton representations for bosonic and fermionic modes with lattice excitations at alternate massive levels. This is the simplest instance of the phenomenon that, following [21], we can call "brane supersymmetry". Here the residual supersymmetry is present only for the massless modes, but in more complicated models it extends to entire sectors of the open spectrum, as first shown in [22].

T-dualities turn these descriptions into equivalent ones, that can often have more intuitive appeal [23]. This is particularly rewarding in the second case: a T-duality along the circle can turn the winding deformation into a momentum deformation orthogonal to the brane responsible for the open-string excitations. It is then perhaps simpler to accept the previous result that the deformation, now a momentum shift orthogonal to the brane, does not affect its massless excitations. A little more work [3] results in a duality argument that associates the (now momentum) shift to the eleventh dimension of M theory, thus realizing the proposal of [24]. Thus, as is often the case, a simple perturbative type I phenomenon has a non-perturbative origin in the heterotic string (and *vice versa*).

3 Brane supersymmetry breaking

I will now conclude by reviewing the third possibility afforded by these constructions. Here I will follow [3], concocting a six-dimensional analogue of "discrete torsion" [25]. The construction of the resulting closed string model is another application of the methods of [7], in the same spirit as the construction of the 10D $U(32)$ model. Starting from the T^4/Z_2 $U(16) \times U(16)$ model [26, 27], one can revert the Klein-bottle projection for all twisted states. This results in an unoriented closed spectrum with (1,0) supersymmetry, whose massless excitations, aside from the gravitational multiplet, comprise 17 tensor multiplets and 4 hypermultiplets. In [4] it is shown how this choice, described by

$$\mathcal{T} = \frac{1}{2}|Q_\mathrm{o} + Q_\mathrm{v}|^2 \Lambda + \frac{1}{2}|Q_\mathrm{o} - Q_\mathrm{v}|^2 \left|\frac{2\eta}{\theta_2}\right|^4 + \frac{1}{2}|Q_\mathrm{s} + Q_\mathrm{c}|^2 \left|\frac{2\eta}{\theta_4}\right|^4 + \frac{1}{2}|Q_\mathrm{s} + Q_\mathrm{c}|^2 \left|\frac{2\eta}{\theta_3}\right|^4 , \tag{3.1}$$

$$\mathcal{K} = \frac{1}{4}\left\{(Q_\mathrm{o} + Q_\mathrm{v})(P + W) - 2 \times 16 (Q_\mathrm{s} + Q_\mathrm{c})\right\} , \tag{3.2}$$

does not allow a consistent supersymmetric solution of the tadpole conditions. A consistent solution does exist [4], but requires the introduction of anti-branes, with the end result that supersymmetry, exact to lowest order in the bulk, is necessarily broken on their world volume. Hence the name "brane supersymmetry breaking" for this peculiar phenomenon, that has the attractive feature of confining the breaking of supersymmetry, and the resulting contributions to the vacuum energy, to a brane, or to a collection of branes, that float in a bath of supersymmetric gravity. In writing these expressions, I have introduced the $(1,0)$ supersymmetric characters [1]

$$\begin{aligned} Q_\mathrm{o} &= V_4 O_4 - C_4 C_4, \qquad & Q_\mathrm{v} &= O_4 V_4 - S_4 S_4, \\ Q_\mathrm{s} &= O_4 C_4 - S_4 O_4, \qquad & Q_\mathrm{c} &= V_4 S_4 - C_4 V_4. \end{aligned} \tag{3.3}$$

The two untwisted ones, Q_o and Q_v, start with a vector multiplet and a hypermultiplet, and are Z_2 orbifold breakings of $(V_8 - S_8)$. Out of the two twisted ones Q_s and Q_c, only Q_s describes massless modes, that in this case correspond to a half-hypermultiplet. The breaking of supersymmetry is demanded by the consistency of String Theory. This can be seen rather neatly from the dependence of the transverse-channel Klein bottle amplitude at the origin of the lattices on the sign ϵ associated to the twisted states

$$\tilde{\mathcal{K}}_0 = \frac{2^5}{4}\left\{ Q_\mathrm{o}\left(\sqrt{v} \pm \frac{1}{\sqrt{v}}\right)^2 + Q_\mathrm{v}\left(\sqrt{v} \mp \frac{1}{\sqrt{v}}\right)^2 \right\} , \tag{3.4}$$

where the upper signs would correspond to the conventional case of the $U(16) \times U(16)$ model, while the lower signs correspond to the model of equation (3.2). Since the terms with different powers of $\sqrt{v}$ are related by tadpole conditions to the multiplicities of the N and D charge spaces, a naive solution of the model corresponding to the upper signs would require a negative multiplicity, $D = -32$.

The open sector is described by

$$
\begin{aligned}
\mathcal{A} = \frac{1}{4}\Big\{ & (Q_{\mathrm{o}} + Q_{\mathrm{v}})(N^2 P + D^2 W) + 2ND(Q'_{\mathrm{s}} + Q'_{\mathrm{c}})\left(\frac{\eta}{\theta_4}\right)^2 \\
& + (R_{\mathrm{N}}^2 + R_{\mathrm{D}}^2)(Q_{\mathrm{o}} - Q_{\mathrm{v}})\left(\frac{2\eta}{\theta_2}\right)^2 \\
& + 2R_{\mathrm{N}}R_{\mathrm{D}}(-O_4S_4 - C_4O_4 + V_4C_4 + S_4V_4)\left(\frac{\eta}{\theta_3}\right)^2\Big\} \\
\mathcal{M} = -\frac{1}{4}\Big\{ & NP(\hat{O}_4\hat{V}_4 + \hat{V}_4\hat{O}_4 - \hat{S}_4\hat{S}_4 - \hat{C}_4\hat{C}_4) \\
& - DW(\hat{O}_4\hat{V}_4 + \hat{V}_4\hat{O}_4 + \hat{S}_4\hat{S}_4 + \hat{C}_4\hat{C}_4) \\
& - N(\hat{O}_4\hat{V}_4 - \hat{V}_4\hat{O}_4 - \hat{S}_4\hat{S}_4 + \hat{C}_4\hat{C}_4)\left(\frac{2\hat{\eta}}{\hat{\theta}_2}\right)^2 \\
& + D(\hat{O}_4\hat{V}_4 - \hat{V}_4\hat{O}_4 + \hat{S}_4\hat{S}_4 - \hat{C}_4\hat{C}_4)\left(\frac{2\hat{\eta}}{\hat{\theta}_2}\right)^2\Big\}.
\end{aligned}
\tag{3.5}
$$

Supersymmetry is broken on the antibranes, and indeed the amplitudes involve new characters Q'_{s} and Q'_{c}, that describe supermultiplets of a chirally flipped supercharge and may be obtained from equation (3.3) upon the interchange of S_4 and C_4, as well as other non-supersymmetric combinations. The tadpole conditions determine the gauge group $[SO(16) \times SO(16)]_9 \times USp(16) \times USp(16)]_{\bar{5}}$, and the 99 spectrum is supersymmetric, with (1,0) vector multiplets for the $SO(16) \times SO(16)$ gauge group and a hypermultiplet in the $(\mathbf{16}, \mathbf{16}, \mathbf{1}, \mathbf{1})$. On the other hand, the $\bar{5}\bar{5}$ spectrum is not supersymmetric and, aside from the $[USp(16) \times USp(16)]$ gauge vectors, contains quartets of scalars in the $(\mathbf{1}, \mathbf{1}, \mathbf{16}, \mathbf{16})$, right-handed Weyl fermions in the $(\mathbf{1}, \mathbf{1}, \mathbf{120}, \mathbf{1})$ and $(\mathbf{1}, \mathbf{1}, \mathbf{1}, \mathbf{120})$ and left-handed Weyl fermions in the $(\mathbf{1}, \mathbf{1}, \mathbf{16}, \mathbf{16})$. Finally, the ND sector, also not supersymmetric, comprises doublets of scalars in the $(\mathbf{16}, \mathbf{1}, \mathbf{1}, \mathbf{16})$ and in the $(\mathbf{1}, \mathbf{16}, \mathbf{16}, \mathbf{1})$, and additional (symplectic) Majorana-Weyl fermions in the $(\mathbf{16}, \mathbf{1}, \mathbf{16}, \mathbf{1})$ and $(\mathbf{1}, \mathbf{16}, \mathbf{1}, \mathbf{16})$. These fields are a peculiar feature of six-dimensional space time, where one can define Majorana-Weyl fermions, if the Majorana condition is supplemented by the conjugation in a pseudo-real representation. All irreducible gauge and gravitational anomalies cancel also in this model, while the residual anomaly polynomial requires a generalized Green-Schwarz mechanism [28].

The radiative corrections in this model are quite interesting, since they convey the (soft) breaking to the gravitational sector. At any rate, the situation in which a model requires the simultaneous presence of branes and antibranes presents itself in at least two other instances [4,29], the four dimensional $Z_2 \times Z_2$ model with discrete torsion and the four-dimensional Z_4 model. In both cases, brane supersymmetry breaking allows a solution of all tadpole conditions and a consistent definition of the open descendants.

One can actually enrich these constructions, allowing for the simultaneous presence of branes and antibranes of the same type [4, 30, 31]. These configurations are generically unstable, and their instability reflects itself in the possible presence of tachyonic excitations, a feature that we have already confronted in our analysis of the ten-dimensional type-0 models. The internal lattice can be used to lift in mass the tachyons, at least within certain ranges of parameters for the internal geometry, that is actually partly *stabilized*, as a result of the different scaling behavior ($O(\sqrt{v})$ and $O(1/\sqrt{v})$) of the contributions of the different $D_{\rm p}$ branes. Thus, for instance, starting from the type-IIB superstring, one can introduce both branes and antibranes at the price of having tachyonic excitations in the open spectrum [31]. In addition, even with special tachyon-free configurations, simply waiving the restriction to configurations free of NS-NS tadpoles often gives new interesting models with broken supersymmetry. The simplest setting is provided again by the type-IIB superstring that, aside from the type-I superstring, has an additional chiral tachyon-free descendant, free of gauge and gravitational anomalies, but with broken supersymmetry and a $USp(32)$ gauge group. In lower-dimensional models, more possibilities are afforded by the internal lattice, that may be used to lift in mass some tachyons, leading to stable vacuum configurations including both branes and antibranes. Some of these [32], related to the four-dimensional Z_3 orientifold of [33], appear particularly interesting.

I would like to conclude by mentioning that brane configurations similar to these have also been studied by several other authors over the last couple of years, from a different vantage point, following Sen [15]. Brane studies are reminiscent of monopole studies in gauge theories of a classical Electrodynamics of charge probes in a given external field, and are an interesting enterprise in their own right. Open-string vacua are particular brane configurations that are also vacuum configurations for a perturbative construction, and thus become exact solutions in the limit of vanishing string coupling. This, in retrospect, makes them particularly attractive and instructive, and makes their study particularly rewarding. In addition, they have very amusing applications, in particular to issues related to the AdS/CFT correspondence, that we are only starting to appreciate.

I am grateful to the Organizers for their kind invitations and to C. Angelantonj, I. Antoniadis, M. Bianchi, G. D' Appollonio, E. Dudas, G. Pradisi, Ya.S. Stanev for stimulating discussions and very enjoyable collaborations on the topics reviewed in this talk, originally prepared for the QFTHEP Meeting, held in Moscow in June 1999, and for the QG99 Meeting, held in Villasimius in September 1999. With slight corrections

in the text and in the references, this short review is now being contributed to the Proceedings of the 2001 Como Conference on "Statistical Field Theories", of the 2001 Les Houches Summer Institute on "Gravity, gauge theory and strings" and of the 2001 Johns Hopkins Meeting, since my presentations in all cases are within its spirit of this short review. A forthcoming review article [34] will provide a comprehensive derivation of these world-sheet constructions more easily accessible to interested readers.

References

[1] M. Bianchi and A. Sagnotti, *Phys. Lett. B* **247** (1990) 517.

[2] A. Sagnotti [hep-th/9509080]; *Nucl. Phys. Proc. Suppl. B* **56** (1997) 332 [hep-th/9702093].

[3] I. Antoniadis, E. Dudas and A. Sagnotti, *Nucl. Phys. B* **544** (1999) 469 [hep-th/9807011]; I. Antoniadis, G. D'Appollonio, E. Dudas and A. Sagnotti, *Nucl. Phys. B* **553** (1999) 133 [hep-th/9812118]; *Nucl. Phys. B* **565** (2000) 123 [arXiv:hep-th/9907184]; A.L. Cotrone, *Mod. Phys. Lett. A* **14** (1999) 2487 [arXiv:hep-th/9909116].

[4] I. Antoniadis, E. Dudas and A. Sagnotti, *Phys. Lett. B* **464** (1999) 38 [hep-th/9908023]; C. Angelantonj, *Nucl. Phys. B* **566** (2000) 126 [arXiv:hep-th/9908064]; G. Aldazabal and A.M. Uranga, *JHEP* **9910** (1999) 024 [hep-th/9908072]; G. Aldazabal, L.E. Ibanez and F. Quevedo [hep-th/9909172]; C. Angelantonj, I. Antoniadis, G. D'Appollonio, E. Dudas and A. Sagnotti [hep-th/9911081].

[5] A. Sagnotti, *Non-Perturbative Quantum Field Theory*, edited by G. Mack *et al.* (Pergamon Press, 1988), p. 521; G. Pradisi and A. Sagnotti, *Phys. Lett. B* **216** (1989) 59; M. Bianchi and A. Sagnotti, in [1]; *Nucl. Phys. B* **361** (1991) 519; M. Bianchi, G. Pradisi and A. Sagnotti, *Nucl. Phys. B* **376** (1992) 365.

[6] L.J. Dixon and J.A. Harvey, *Nucl. Phys. B* **274** (1986) 93; N. Seiberg and E. Witten, *Nucl. Phys. B* **276** (1986) 272.

[7] D. Fioravanti, G. Pradisi and A. Sagnotti, *Phys. Lett. B* **321** (1994) 349 [hep-th/9311183]; G. Pradisi, A. Sagnotti and Y. S. Stanev, *Phys. Lett. B* **354** (1995) 279 [hep-th/9503207]; *Phys. Lett. B* **356** (1995) 230 [hep-th/9506014]; *Phys. Lett. B* **381** (1996) 97 [hep-th/9603097]. For short reviews see G. Pradisi, *Nuovo Cim. B* **112** (1997) 467 [arXiv:hep-th/9603104]; A. Sagnotti and Y.S. Stanev, *Fortsch. Phys.* **44** (1996) 585; [*Nucl. Phys. Proc. Suppl. B* **55** (1996) 200] [arXiv:hep-th/9605042].

[8] J. Polchinski and Y. Cai, *Nucl. Phys. B* **296** (1988) 91.

[9] J.L. Cardy, *Nucl. Phys. B* **324** (1989) 581.

[10] E. Witten, *Phys. Lett. B* **149** (1984) 351.

[11] O. Bergman and M.R. Gaberdiel, *Nucl. Phys. B* **499** (1997) 183 [hep-th/9701137]; *JHEP* **9907** (1999) 022 [hep-th/9906055].

[12] C. Angelantonj, *Phys. Lett. B* **444** (1998) 309 [hep-th/9810214]; R. Blumenhagen, A. Font and D. Lust [hep-th/9904069]; R. Blumenhagen and A. Kumar, *Phys. Lett. B* **464** (1999) 46 [hep-th/9906234]; K. Forger, *Phys. Lett. B* **469** (1999) 113 [hep-th/9909010].

[13] I.R. Klebanov and A.A. Tseytlin, *Nucl. Phys. B* **546** (1999) 155 [hep-th/9811035]; *JHEP* **9903** (1999) 015 [hep-th/9901101]. An interesting discussion of non-tachyonic type-0 descendants in this context can be found in C. Angelantonj and A. Armoni [hep-th/9912257].

[14] J. Maldacena, *Adv. Theor. Math. Phys.* **2** (1998) 231 [hep-th/9711200]; S.S. Gubser, I.R. Klebanov and A.M. Polyakov, *Phys. Lett. B* **428** (1998) 105 [hep-th/9802109]; E. Witten, *Adv. Theor. Math. Phys.* **2** (1998) 253 [hep-th/9802150]. For a review see O. Aharony, S.S. Gubser, J. Maldacena, H. Ooguri and Y. Oz [hep-th/9905111].

[15] A. Sen, *JHEP* **9806** (1998) 007 [hep-th/9803194]; *JHEP* **9808** (1998) 010 [hep-th/9805019]; *JHEP* **9808** (1998) 012 [hep-th/9805170]; *JHEP* **9809** (1998) 023 [hep-th/9808141]; *JHEP* **9812** (1998) 021 [hep-th/9812031]. For recent reviews, see A. Sen [hep-th/9904207]; A. Lerda and R. Russo [hep-th/9905006].

[16] See E. Dudas, J. Mourad and A. Sagnotti, *Nucl. Phys. B* **620** (2002) 109 [arXiv:hep-th/0107081], and references therein

[17] J. Scherk and J.H. Schwarz, *Nucl. Phys. B* **153** (1979) 61.

[18] R. Rohm, *Nucl. Phys. B* **237** (1984) 553; C. Kounnas and M. Porrati, *Nucl. Phys. B* **310** (1988) 355; S. Ferrara, C. Kounnas, M. Porrati and F. Zwirner, *Nucl. Phys. B* **318** (1989) 75; C. Kounnas and B. Rostand, *Nucl. Phys. B* **341** (1990) 641; I. Antoniadis and C. Kounnas, *Phys. Lett. B* **261** (1991) 369; E. Kiritsis and C. Kounnas, *Nucl. Phys. B* **503** (1997) 117 [hep-th/9703059].

[19] J.D. Blum and K.R. Dienes, *Phys. Lett. B* **414** (1997) 260 [hep-th/9707148]; J.D. Blum and K.R. Dienes, *Nucl. Phys. B* **516** (1998) 83 [hep-th/9707160].

[20] J. Polchinski and E. Witten, *Nucl. Phys. B* **460** (1996) 525 [hep-th/9510169].

[21] Z. Kakushadze and S.H. Tye, *Nucl. Phys. B* **548** (1999) 180 [hep-th/9809147].

[22] R. Blumenhagen and L. Gorlich, *Nucl. Phys. B* **551** (1999) 601 [hep-th/9812158]; C. Angelantonj, I. Antoniadis and K. Forger, *Nucl. Phys. B* **555** (1999) 116 [hep-th/9904092].

[23] P. Horava, *Nucl. Phys. B* **327** (1989) 461; P. Horava, *Phys. Lett. B* **231** (1989) 251; J. Dai, R.G. Leigh and J. Polchinski, *Mod. Phys. Lett. A* **4** (1989) 2073.

[24] E. Dudas and C. Grojean, *Nucl. Phys. B* **507** (1997) 553 [hep-th/9704177]; I. Antoniadis and M. Quiros, *Nucl. Phys. B* **505** (1997) 109 [hep-th/9705037].

[25] C. Vafa, *Nucl. Phys. B* **273** (1986) 592; C. Vafa and E. Witten, *J. Geom. Phys.* **15** (1995) 189 [hep-th/9409188]; J.D. Blum, *Nucl. Phys. B* **486** (1997) 34 [hep-th/9608053]; M.R. Douglas [hep-th/9807235]; S. Mukhopadhyay and K. Ray [hep-th/9909107].

[26] M. Bianchi and A. Sagnotti, in [5].

[27] E.G. Gimon and J. Polchinski, *Phys. Rev. D* **54** (1996) 1667 [hep-th/9601038].

[28] M.B. Green and J.H. Schwarz, *Phys. Lett. B* **149** (1984) 117; A. Sagnotti, *Phys. Lett. B* **294** (1992) 196 [hep-th/9210127].

[29] M. Bianchi, Ph.D. Thesis, preprint ROM2F-92/13; A. Sagnotti [hep-th/9302099]; G. Zwart, *Nucl. Phys. B* **526** (1998) 378 [hep-th/9708040]; Z. Kakushadze, G. Shiu and S.H. Tye, *Nucl. Phys. B* **533** (1998) 25 [hep-th/9804092]; G. Aldazabal, A. Font, L.E. Ibanez and G. Violero, *Nucl. Phys. B* **536** (1998) 29 [hep-th/9804026].

[30] G. Aldazabal and A.M. Uranga, in [4].

[31] S. Sugimoto, *Prog. Theor. Phys.* **102** (1999) 685 [hep-th/9905159].

[32] G. Aldazabal, L.E. Ibanez and F. Quevedo, in [4].

[33] C. Angelantonj, M. Bianchi, G. Pradisi, A. Sagnotti and Y.S. Stanev, *Phys. Lett. B* **385** (1996) 96 [hep-th/9606169].

[34] C. Angelantonj and A. Sagnotti, *Phys. Reports* **371** (2002) 1 [hep-th/0204089].

SEMINAR 7

ON A FIELD THEORY OF OPEN STRINGS, TACHYON CONDENSATION AND CLOSED STRINGS

S.L. SHATASHVILI

Department of Physics,
Yale University, New Haven,
CT 06520-8120, U.S.A.

ON A FIELD THEORY OF OPEN STRINGS, TACHYON CONDENSATION AND CLOSED STRINGS

S.L. Shatashvili

Abstract

I review the physical properties of different vacua in the background independent open string field theory.

One of the most popular candidates for field theory of open strings is the classical action proposed by Witten in 1985 [1] – cubic **CS** action. The first test that any string field theory action shall be subject to is to recover all tree level as well as loop amplitudes (which are independently known exactly from world-sheet approach) by standard field theory methods, and cubic action does produce correct tree level amplitudes [2]. It seems that if tree level amplitudes are recovered by unitarity one can reconstruct all perturbative, loop, amplitudes. But, it is known that one loop diagram in any field theory of open strings shall contain the closed string pole (the cylinder diagram can be viewed as one loop in open string theory, or equivalently as a tree level propagator for closed strings; thus there shall be a pole for all on-shell closed string momenta); from the point of view of open string field theory these new (closed string) poles violate unitarity since corresponding degrees of freedom are not (at least in any obvious way) present in the particle spectrum of classical Lagrangian. This situation is similar to the one in anomalous gauge theory, but in the latter one can choose the representation for fermions such that anomaly cancels. In open string field theory case one can make arrangements when closed string poles decouple (for example in topological or non-commutative setup) but it seems very interesting to study other possibilities.

One can think of two options: 1) find the closed string degrees of freedom as "already being present" in given classical open string field theory Lagrangian, or 2) introduce them as additional degrees of freedom (for example by adding corresponding string field together with its Lagrangian plus the interaction term with open string field).

In the lines of the option 2 the solution to the above problem has been found many years ago by B. Zwiebach (see recent version [3] which also contains references on old work). In this approach one shall take care of the problem of multiple counting when closed string field and its Lagrangian is added – the same Feynman diagram will come both from open and closed string sectors; thus one shall make sure that each diagram is properly counted only once; the solution of this problem is rather complicated and requires the detailed knowledge of the string amplitudes to all loop order from world-sheet approach[1].

It is very interesting to explore option 1 instead. In order to do so one needs to have a truly background independent open string field theory. Unfortunately such a theory has not yet been written although the theory which doesn't depend (at least formally) on the choice of open string background is known [5–8] (it also passes the test of reproducing all tree level on-shell amplitudes in a very simple way since corresponding action on-shell is given by world-sheet partition function on the disk as it was explained in [6, 8]).

One might try to make option 1 more precise by exploring the idea of closed string degrees of freedom being some kind of classical, solitonic, solutions of open string field theory. More concretely: start with open string field theory (background independent) and find the new background – closed string. This seems to be a natural way of "reversing the arrow" which describes D-branes (open strings in case of space-filling branes) as solitons in closed string theory [9].

The above line of thoughts suggests that we shall change the point of view about branes in general and think about them as solitons in open string field theory rather than in closed one. In fact, if one considers the maximal dimensional brane (or brane-anti-brane system) it is very easy to think [10] about lower-dimensional branes as solutions of classical equations of motion for corresponding open string field theory action in the formalism of background independent open string field theory. The latter is defined via *the action $S(t)$ – a functional on the space of boundary conditions for bosonic string on the disc with critical points $t = t_* \rightarrow$ being the conformal boundary conditions.* Since for the trivial bulk backgrounds (Δ operator on world-sheet) mixed Dirichlet-Neumann conditions (D-branes) are conformal (in fact in this case these are only conformal boundary conditions) they shall correspond to critical points of space-time action for open strings [6, 10].

Another motivation for the search of closed strings in open string field theory (at least for the present author) comes from Matrix Strings [11].

[1]One shall note that adding closed string fields to open string field theory Lagrangian is very similar to the formalism of [4] for quantization of anomalous gauge theories.

If we take the soliton corresponding to N $D1$-strings in open string field theory and look on the dynamics of collective modes we find (in strong coupling for 2d theory on $D1$ and large N) the spectrum of closed strings in the same space-time where open strings live, thus we might ask the question whether these closed string degrees of freedom are already present in original theory of open strings where we had $D1$'s as solitons in the beginning. This observation also might help in making contact with option 1 described above. One shall mention that the search for closed strings in open string field theory has a long history and was originated in [12]; in the above context, more in the lines of current developments related to D-branes, the interest has been revived in [10].

The conjectures put forward by A. Sen [13,14] made it possible to study these questions in much more precise terms. For simplicity one considers the open bosonic string in 26 dimensions ($D25$, or any $Dp, p < 26$) which contains tachyon and is unstable. Three conjectures made by A. Sen are:

1. Tachyon potential takes the form:

$$V(T) = Mf(T), \tag{1}$$

with M-mass of D-brane and f – universal function independent of the background where brane is embedded. The conjecture of Sen states that $f(T)$ has a stationary point (local minimum) at some $T = T_{\mathrm{c}} < \infty$ such that

$$f(T_{\mathrm{c}}) = -1 \tag{2}$$

and thus $M + V(T_{\mathrm{c}}) = M(1 + f(T_{\mathrm{c}})) = 0$.

2. There are soliton configurations on unstable D-branes which correspond to lower-dimensional branes.

3. New vacuum, at T_{c}, is a closed string vacuum and in addition there are no open string degrees of freedom.

One might be tempting to amplify the **Conjecture 3** a bit [10] and claim that in properly defined open string field theory there shall be an expansion around new critical point which will describe the theory of closed strings (without open string sector; of course in this theory of closed strings we again can discuss open strings as solitonic branes). *A priori* there is no reason to assume that such expansion should exist since the potential not necessarily shall be analytic, but one just can hope to see whole closed string sector and not just vacuum by starting from classical Lagrangian for open strings. We will comment on this question at the end of this talk. One shall note that the picture described below together with the corresponding tachyon potential is very attractive from the point of view of applications of string theory to the phenomenology related to branes and also to stringy cosmology.

We will address these problems in the formalism of [5–8] and present the exact tachyon Lagrangian up to two derivatives in tachyon field, which provides the important tool in verifying Sen's conjectures; more detailed discussion and references can be found in [15–18] for the bosonic case and [19] for superstring[2]. The important questions related to the description of multiple-branes in the formalizm of background independent open string field theory and unified treatment of **RR** couplings is studied in [22].

Following [5] one starts with world-sheet description of critical bosonic string theory on disk. Consider the map of the disk D to space-time M:

$$X(z,\bar{z}): \qquad D \to M. \tag{3}$$

In general one can consider any critical 2d CFT coupled to 2d gravity on the disk but it is interesting to study the particular case of critical bosonic string with flat 26 dimensional space-time $M = R^{1,25}$.

Two-dimensional quantum field theory on the string world-sheet is given by the path integral:

$$\langle ... \rangle = \int [\mathrm{d}X][\mathrm{d}b][\mathrm{d}c] \quad \mathrm{e}^{-I_0(X,b,c)} ... \tag{4}$$

$$I_0 = \int_D \sqrt{g}\left[g^{\alpha\beta}\partial_\alpha X^\mu \partial_\beta X_\mu + b^{\alpha\beta} D_\alpha c_\beta\right]. \tag{5}$$

Define BRST operator through the current J_{BRST} and contour C (note this is a closed string BRST operator):

$$Q_{\mathrm{BRST}} = \int_C J_{\mathrm{BRST}}; \qquad Q^2_{\mathrm{BRST}} = 0. \tag{6}$$

Denote the limit when contour C approaches the boundary ∂D by Q: $Q = \int_{C\to\partial D} J_{\mathrm{BRST}}$. Now we consider the local operator $\mathcal{V}(X,b,c)$ of the form

$$\mathcal{V} = b_{-1}\mathcal{O}(X,b,c), \qquad b_{-1} = \int_{C\to\partial D} v^i b_{ij}\epsilon^j_k \mathrm{d}x^k \tag{7}$$

and deform the world-sheet action:

$$I_{\mathrm{ws}} = I_0 + \int_{C\to\partial D} \mathcal{V}(X(\sigma)) \tag{8}$$

[2]The same tachyon Lagrangian, which we will present below for bosonic string and its analog for supersymmetric case was proposed in [20] as a toy model that mimics the properties of tachyon condensation. Very impressive progress has been achieved in verifying Sen's conjectures in the cubic string field theory of [1]; see contributions of A. Sen, B. Zwiebach and W. Taylor in the proceedings of Strings (2001) conference. One should note that the world-sheet approach to the problem was discussed previously in [21].

$$\langle ... \rangle = \int [\mathrm{d}X][\mathrm{d}b][\mathrm{d}c] \mathrm{e}^{-I_{\mathrm{ws}}} ... \tag{9}$$

The simplest case is when ghosts decouple: $\mathcal{O} = cV(X)$. The boundary term in the action modifies the boundary condition on the map $X^\mu(z, \bar{z})$ from the Neumann boundary condition (this follows from I_0) $\partial_{\mathrm{r}} X^\mu(\sigma) = 0$ to "arbitrary" non-linear condition:

$$\partial_{\mathrm{r}} X^\mu(\sigma) = \frac{\partial}{\partial X^\mu(\sigma)} \int_{\partial D} V(X) \tag{10}$$

I_{ws} defines the family of boundary 2d quantum field theories on the disk. The action $S(\mathcal{O})$ is defined on this space (more precisely – on the space of $\mathcal{O}$'s) and is formally independent of the choice of particular open string background:

$$\mathrm{d}S = \left\langle \mathrm{d} \int_{\partial D} \mathcal{O} \quad \left\{ Q, \quad \int_{\partial D} \mathcal{O} \right\} \right\rangle. \tag{11}$$

Since $\mathrm{d}\mathcal{O}$ is arbitrary all solutions of the equation $\mathrm{d}S = 0$ correspond to conformal, exactly marginal boundary deformations with $\{Q, \mathcal{O}\} = 0$, and thus to valid string backgrounds.

A very important question at this stage is to understand *what is the space of deformations given by* $V(X(\sigma))$. An obvious assumption (which is also a very strong restriction, see the comment at the end of this talk) is – V can be expanded into "Taylor series" in the derivatives of $X(\sigma)$:

$$\begin{aligned} V(X) = T(X(\sigma)) + A_\mu(X(\sigma))\partial X^\mu(\sigma) + C_{\mu\nu}(X(\sigma))\partial X^\mu(\sigma)\partial X^\nu(\sigma) \\ + D_\mu(X(\sigma))\partial^2 X^\mu(\sigma) + ... \end{aligned} \tag{12}$$

Thus the action now becomes the functional of coefficients: $S = S(T(X(\sigma)),$ $A_\mu(X(\sigma)), ...)$. It is almost obvious that the above assumption singles out the open string degrees of freedom from very large space of functionals of the map $X^\mu(\sigma) - \partial D \to M$. The goal is to write S as an integral over the space-time X (constant mode of $X(\sigma)$: $X(\sigma) = X + \phi(\sigma)$, $\int \phi(\sigma) = 0$) of some "local" functional of fields $T(X), A(X), ...$ and their derivatives.

In a more general setup one can introduce some coordinate system $\{t^1, t^2, ...\}$ in the space of the boundary operators – $\mathcal{O} = \mathcal{O}(t, X(\sigma)); V = V(t, X(\sigma))$:

$$\mathrm{d}\mathcal{O} = \sum \mathrm{d}t^i \frac{\partial}{\partial t^i} \mathcal{O}(t), \qquad \mathrm{d}V(t, X(\sigma)) = \sum \mathrm{d}t^i \frac{\partial}{\partial t^i} V(t, X(\sigma)).$$

At the origin, $t^i = 0$, we have an un-deformed theory and the linear term in the deformation is given by an operator $\int_{\partial D} V_i$ of dimension Δ_i from the spectrum:

$$I_{\rm ws} = I_0 + t^i \int_{\partial D} V_i + O(t^2) = I_0 + t^i \frac{\partial}{\partial t^i} \int_{\partial D} V(t)_{|_{t=0}} + O(t^2). \qquad (13)$$

For the general boundary term $\int_{\partial D} V$ one might worry that the two-dimensional theory on the disk is not renormalizable; it makes sense as a cutoff theory, but it turns out that if one perturbes by a complete set of operators from the spectrum the field theory action is still well-defined (see discussion at the end [8]; for the tachyon T and gauge field A world-sheet theory is obviously renormalizable). It has been proven that the action (11) can written in terms of world-sheet β-function and partition function [6, 8]:

$$S(t) = -\beta^i(t) \frac{\partial}{\partial t^i} Z(t) + Z(t) \qquad (14)$$

here β^i is the beta function for coupling t^i, a vector field in the space of boundary theories, and $Z(t)$ – the matter partition function on the disk.

Since equations of motion $\mathrm{d}S = 0$ coincide with the condition that deformed 2d theory is exactly conformal we have

$$\frac{\partial}{\partial t^i} S(t) = G_{ij}(t) \beta^j(t) \qquad (15)$$

with some non-degenerate Zamolodchikov metric $G_{ij}(t)$ (the equation $\partial S(t) = 0$ shall be equivalent to $\beta^i = 0$). In addition we see that on-shell

$$S_{\rm on-shell}(t) = Z(t)$$

and as a result classical action on solutions of equation motion will generate correct tree level string amplitudes.

It is important to note that in general the total derivatives don't decouple inside the correlation function and we have coupling dependent BRST operator $Q(t)$. In fact the same contact terms contributes to β-function. More precisely one can fix the contact terms from the condition that the definition (11) is self-consistent after contact terms are included:

$$Q = Q(t); \quad \langle .... \int \partial_\theta ... \rangle \neq 0 \qquad (16)$$

$$\mathrm{d} \quad \left[\quad \mathrm{d}S = \left\langle \mathrm{d} \int_{\partial D} \mathcal{O} \quad \left\{ Q, \quad \int_{\partial D} \mathcal{O} \right\} \right\rangle \quad \right] = 0. \qquad (17)$$

If one assumes that Q is constant and total derivatives decouple – (17) is a simple Ward identity; otherwise it is a condition which relates the choice of contact terms with t dependence of corresponding modes of Q [7,8]. Usually, in quantum field theory one has to choose the contact terms (regularization) based on some (symmetry) principle (an example from recent years is the Donaldson theory which rewritten in terms of Seiberg-Witten IR description requires the choice of the contact terms based on Seiberg-Witten modular invariance together with topological Q symmetry [23] and dependence of Q on moduli is fixed from this consistency principle); here we have (15), (17) as guiding principle.

The principle (17) leads to the formula (14) for the action with all non-linear terms included: $\beta^i = (\Delta_i - 1)t^i + c^i_{jk}t^j t^k + ...$ and in addition guarantees that Zamolodchikov metric in (15) is non-degenerate. In second order all terms except those that satisfy the resonant condition $\Delta_j - \Delta_i + \Delta_k = 1$ can be removed by redefinition of couplings; obviously these correspond to logarithmic divergences and thus if the theory is perturbed with the complete set of operators – it is logarithmic. More generally – from Poincare-Dulac theorem (which of course is for the finite-dimensional case but we will assume it is correct for infinite-dimensional space also) one can show that all coefficients are cutoff-independent after an appropriate choice of coordinates is made. Let us repeat in regard to the choice of coordinates – as it follows from the above discussion, any choice of coordinates is good as long as *equations of motion lead to β-function equations with invertible Zamolodchikov metric G_{ij}.*

First we turn on only tachyon: $V(X(\sigma)) = T(X(\sigma))$. It is not difficult to find the action $S(T)$ as an expansion in derivatives; for example – exact in T and second order in derivatives ∂. We know the derivative expansion of β and Z:

$$\beta^T(X) = [2\Delta T + T] + a_0(T) + a_1(T)(\partial T) + a_2(T)(\partial^2 T) + a_3(T)(\partial T)^2 + ... \tag{18}$$

$$Z(T) = \int \mathrm{d}X \mathrm{e}^{-T}(1 + b(T)(\partial T)^2 + ...) \tag{19}$$

with appropriate conditions for unknown coefficients dictated by the perturbative expansion. In this concrete case the basic relation becomes:

$$S(T) = -\int \mathrm{d}X \beta^T(X) \frac{\partial}{\partial T(X)} Z(T) + Z(T). \tag{20}$$

The condition for β in (18) to be an equation of motion for $S(T)$ (20) with Z given by (19) in lowest order in T (around $T = 0 \to 2\Delta T + T = 0$) fixes the two derivative action (and relevant unknowns $a_i(T), b(T)$):

$$S(T) = \int \mathrm{d}X \left[\mathrm{e}^{-T}(\partial T)^2 + \mathrm{e}^{-T}(1 + T)\right] \tag{21}$$

with equations of motion, β^T and metric $G = \mathrm{e}^{-T}$ in (15):

$$\partial_T S = \mathrm{e}^{-T}\beta^T = \mathrm{e}^{-T}(2\Delta T + T - (\partial T)^2) = 0 \tag{22}$$

$$\beta^T = 2\Delta T + T - (\partial T)^2; \quad Z(T) = \int \mathrm{e}^{-T}. \tag{23}$$

This answer was deduced from the consistency condition for the expansion (18), (19) and basic relation (20), but one can compute it directly from world-sheet definition, practically repeating the computations (in this particular case) leading to general relation (14). We first write the world-sheet path integral only in terms of boundary data $\int \mathrm{d}X(\theta)\mathrm{e}^{-I_{\mathrm{ws}}} = \int \mathrm{d}X\mathrm{d}\phi(\theta)\mathrm{e}^{-I_{\mathrm{ws}}}$ (we use the notation $H(\theta,\theta') = \frac{1}{2}\sum_k \mathrm{e}^{ik(\theta-\theta')}|k|$; bulk part decouples since arbitrary map $X(z,\bar{z})$ can be written as the sum of two terms: one has zero value on the boundary, another coincides with $X(\sigma)$ on boundary and is harmonic in the bulk):

$$X(\theta) = X + \phi(\theta); \qquad \int \phi(\theta) = 0; \tag{24}$$

$$\begin{aligned} I_{\mathrm{ws}} &= \int\int \mathrm{d}\theta\mathrm{d}\theta' X^\mu(\theta)H(\theta,\theta')X_\mu(\theta') + \int \mathrm{d}\theta T(X(\theta)) \\ &= T(X) + \int \phi^\mu(\theta)[H(\theta-\theta')\delta_{\mu\nu} + \delta(\theta-\theta')\partial_\mu\partial_\nu T(X)]\phi^\nu(\theta') \\ &\qquad + O(\partial^3 T(X)). \end{aligned} \tag{25}$$

In the two derivative approximation contribution comes only from[3]:

$$Z(T) = \int \mathrm{d}X\mathrm{e}^{-T(X)}\det{}'\left[H + \partial^2 T\right]^{-\frac{1}{2}}. \tag{26}$$

This can be computed exactly. For $\partial_\mu\partial_\nu T = \delta^{\mu\nu}\partial_\mu^2 T$ it is [6] (in some regularization):

$$Z = \int \mathrm{d}X\mathrm{e}^{-T(X)}\prod_\mu \sqrt{\partial_\mu^2 T(X)}\mathrm{e}^{\gamma\partial_\mu^2 T(X)}\Gamma\left(\partial_\mu^2 T(X)\right) \tag{27}$$

[3]If we turn on other fields from (12) it immediately follows from corresponding expression of the type (25) that only tachyon will condense; see discussion in [16].

and in the two-derivative approximation this gives[4]:

$$\begin{aligned} Z(T) &= \int \mathrm{d}X \mathrm{e}^{-T(X)}(1+b(T))(\partial T)^2) \\ \beta^T &= 2\Delta T + T \\ b(T) &= 0, \quad a_0(T) = a_1(T) = a_2(T) = 0, \quad a_3(T) = -1. \end{aligned} \tag{28}$$

Thus the action (14) is:

$$S(T) = \int \mathrm{d}X \left[\mathrm{e}^{-T} 2(\partial T)^2 + \mathrm{e}^{-T}(1+T)\right] \tag{29}$$

with equations of motion:

$$\mathrm{e}^{-T}\left(T + 4\Delta T - 2(\partial T)^2\right) = 0. \tag{30}$$

Now we see that Zamolodchikov metric G becomes:

$$G(\delta_1 T, \delta_2 T) = \int \mathrm{d}X \mathrm{e}^{-T}(\delta_1 T \delta_2 T - 2(\partial_\mu \delta_1 T)(\partial_\mu \delta_2 T)) \tag{31}$$

and equations of motion (30) can be written in this approximation as $G\beta = 0$:

$$\mathrm{e}^{-T}(1 + 2\Delta - 2\partial_\mu T \partial_\mu + ...)(2\Delta T + T) = 0. \tag{32}$$

The linear form of β for arbitrary $T(X)$ looks strange since we miss a possible higher order in T (but second order in ∂T) terms which shall come from a 3-point function. In addition this metric is rather complicated and is not an obvious expansion of some invertible metric in the space of fields. Thus, according the principle for the choice of coordinates, we need to choose new coordinates such that the metric is invertible. Such new variables are given by:

$$T \to T - \partial^2 T + (\partial T)^2. \tag{33}$$

This modifies the β function (without changing its linear part) and metric to (22), (23); in new coordinates action is given by:

$$S(T) = \int \mathrm{d}X \left[\mathrm{e}^{-T}(\partial T)^2 + \mathrm{e}^{-T}(1+T)\right].$$

[4]In fact (27) can be used only for two-derivative approximation for the action since the contribution of the last term in (25) will mix with higher order terms coming from Γ function in (27) after integration by parts due to the presence of universal exponential e^{-T} factor in the action; this is very similar to what happens in the attempts to write non-Abelian version of Born-Infeld action.

The potential in this action has unstable extremum at $T = 0$ (tachyon) and stable at $T = \infty$. The difference between the values of this potential is 1, exactly as predicted by A. Sen in **Conjecture 1**.

In a new variable with the canonical kinetic term $\Phi = \mathrm{e}^{-\frac{T}{2}}$:

$$S(\phi) = \int \left[4(\partial\Phi)^2 - \Phi^2 \log \frac{\Phi^2}{e} \right] \tag{34}$$

(interestingly this is also an exact action, see [15], for a p-adic string for $p \to 1$). In the unstable vacuum $T = 0, \Phi = 1; m^2 = -\frac{1}{2}$; in new vacuum – $T = \infty, \Phi = 0$ and one could naively think that $m^2 = +\infty$, but since the action is non-analytic at this point the meaning is unclear.

DN boundary conditions, $p \le 25$:

$$\partial_{\mathrm{r}} X^a(\sigma) = 0, \quad a = 1, ..., p; \qquad X^i(\sigma) = 0, \quad i = p+1, ..., 26 \tag{35}$$

are obviously conformal. We conclude that they give critical points of string field theory action. In addition since the value of classical action is always $S(t_{\mathrm{c}}) = Z(t_{\mathrm{c}})$ – we conclude that these solitons are in fact Dp-branes; *e.g.* one can take [5]:

$$T(X) = a + u_\mu (X^\mu)^2 \Rightarrow \partial_{\mathrm{r}} X^\mu = u_\mu X^\mu; \quad u_i \to \infty, \quad u_a \to 0. \tag{36}$$

This verifies the **conjecture 2**.

Conjecture 3 is in the heart of the discussion we had in the beginning, and is also most difficult one. In order to address it we add the gauge field. The action in two derivative approximation can be constructed using the same basic relation (14), [18]. Here we will also introduce the closed string fields G and B (for covariance):

$$\begin{aligned} &S(G, B, A, T) = S_{\mathrm{closed}}(G, B) \\ &+ \int \mathrm{d}^{26} X \sqrt{G} \left[\mathrm{e}^{-T}(1+T) + \mathrm{e}^{-T} ||\mathrm{d}T||^2 + \frac{1}{4} \mathrm{e}^{-T} ||B - \mathrm{d}A||^2 + \cdots \right] . \end{aligned} \tag{37}$$

One can choose a different regularization and obtain the action which is an expansion of **BI** action $\int V(T) \sqrt{\det(G - B + \mathrm{d}A)}$ but it would lead to a complicated and not obviously non-degenerate metric, exactly like in purely tachyon case discussed above (32)).

In $\Phi = \mathrm{e}^{-\frac{T}{2}}$ coordinates:

$$\begin{aligned} &S(G, B, A, \Phi) = \\ &S_{\mathrm{closed}} + \int \mathrm{d}^{26} X \sqrt{G} \left[\Phi^2 (1 - 2 \log \Phi) + 4 ||\mathrm{d}\Phi||^2 + + \frac{1}{4} \Phi^2 ||B - \mathrm{d}A||^2 + \cdots \right] . \end{aligned} \tag{38}$$

The latter immediately suggests the analogy with an Abelian Higgs model for complex scalar and gauge field in angular coordinates $H\mathrm{e}^{i\phi}, \mathcal{A}$:

$$S(H,\phi,\mathcal{A}) = S_{\mathrm{YM}}(\mathcal{A}) + \int \left[\lambda(H^2 - H_0^2)^2) + |\mathrm{d}H|^2 + +H^2|\mathcal{A} - \mathrm{d}\phi|^2\right] \tag{39}$$

with identifications:

$$B \to \mathcal{A}, \qquad A \to \phi, \qquad \Phi = \mathrm{e}^{-\frac{1}{2}T} \to H. \tag{40}$$

Exactly like in a symmetric point for an Abelian Higgs model where phase ϕ is not a good coordinate – in string theory the gauge field A becomes ill-defined in new vacuum, $\Phi = 0, T = \infty$; the same is true for all open string modes [18, 24]; all open string degrees of freedom, except the zero mode of tachyon, T_0, are angular variables with T_0 being only the radial one. In addition, as it has been explained in [18], using the Zwiebach's open/closed string field theory, in the symmetric vacuum when the closed string gauge symmetries are restored, all open string degrees of freedom are gauge parameters for closed string gauge transformation and thus disappear from the spectrum. Thus, as long as we know that new vacuum is invariant under closed string gauge transformation we can safely conclude that there are no open string degrees of freedom.

We shall note that in an Abelian Higgs model instead of assuming that background gauge field is a dynamical variable one can also choose the Cartesian coordinates, but in the stringy case there is no possibility to do so since the "phase" (gauge field) and absolute value ($\Phi = \mathrm{e}^{-\frac{T}{2}}$) carry different space-time spin, thus we can only conclude that the theory is smooth and consistent with closed string modes becoming dynamical in new vacuum. This should be compared to the restriction made for the boundary functional $V(X(\sigma))$ in (12) – we see now that our assumption was too restrictive and one shall consider in addition truly non-local functionals which most likely shall correspond to dynamical closed string modes. At the same time one can introduce non-local string field theory wave-function:

$$\Psi(X(\sigma)) = \mathrm{e}^{-\int_C T(X(\sigma)) + A_\mu(X(\sigma))\partial X^\mu(X(\sigma)) + \dots} \tag{41}$$

which can be considered as a formal analog of complex scalar (Cartesian coordinates) field and 2-form field B gives natural connection on the space of such functionals (this Ψ now depends on the choice of contour C in space-time; the creation and annihilation operators of A do not make sense anymore and only loop operator Ψ describes the physical degree of freedom). It seems to be closely related to string field that enters in cubic string field theory [1]: consider wave-function/string field $\Psi(X_*(\sigma))$ given by world-sheet path integral for disk and divide it in two equal parts with the first

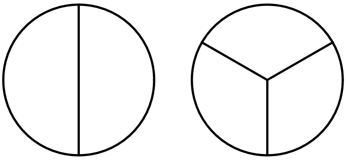


Fig. 1.

half carrying the fixed boundary conditions $X(z,\bar{z})_{\partial D} = X_*(\sigma)$ and on the other half the operator (41) inserted. One shall think about this path integral as a very non-linear and non-local change of variables from $V(X(\sigma))$ to Ψ of cubic string field theory:

$$\Psi_{V(\sigma)}(X_*(\sigma)) = \int \mathrm{e}^{-\int_0^\pi V(X(\sigma))}. \tag{42}$$

For tachyon zero mode it leads to $\mathrm{e}^{-\frac{T}{2}} = 1 + T_0^{\text{cubic}}$. One can map the disk to "1/3 of pizza" (see Fig. 1) and glue three such wave-functions in order to get a cubic term in cubic **CS** string field theory which becomes the disk partition function with the operator (41) inserted everywhere on the boundary of the disk (the second term in the action (14)); as far as first term with β-function it should be possible to obtain from cubic string field theory action kinetic term $(\Psi, Q\Psi)$ *via* the conformal map of the disk to the half of the disk (note this Q is now the open string BRST operator as opposed to the one which enters in the definition of background independent string field theory action). This proposal needs serious investigation; unfortunately a more direct relation is difficult to demonstrate since we considered the restrictive situation when ghosts decouple from matter in the boundary perturbation.

One can wonder what is the description of the above space-time picture from the point of view of world-sheet theory on disk. Since condensation takes place for infinite value of tachyon field, $T \to \infty$, we see that the typical size of the boundary which will contribute to world-sheet path integral has to be small and thus at the end-point of condensation the boundary of the disk shrinks to the zero size (see the Fig. 2). Thus there is no boundary anymore and the operator insertion at the boundary in fact becomes the insertion at the points on the bulk. More careful treatment shows that these operators, which by definition were the operators defined in the bulk and taken to the boundary, are integrated over the whole sphere (with Liouville factor properly included) and thus deformation on the boundary becomes deformation on the sphere. It is tempting to think that this is in fact the correct world-sheet way of thinking about the end-point of condensation and string field theory action for open strings (14) becomes the action for closed string degrees of freedom.

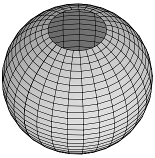


Fig. 2.

I would like to thank A. Gerasimov for collaboration during the recent studies in [15,18] and Rutgers New High Energy Theory Group for hospitality. This work is supported by OJI award from DOE.

References

[1] E. Witten, *Nucl. Phys. B* **268** (1986) 235.

[2] D. Gross, A. Jevicki, *Nucl. Phys. B* **283** (1987) 1; D. Gross and A. Jevicki, *Nucl. Phys. B* **287** (1987) 225.

[3] B. Zwiebach, *Annals Phys.* **267** (1998) 193 [hep-th/9705241].

[4] L. Faddeev and S. Shatashvili, *Theor. Math. Phys.* **60** (1985) 770; *Teor. Mat. Fiz.* **60** (1984) 206; L. Faddeev and S. Shatashvili, *Phys. Lett. B* **167** (1986) 225.

[5] E. Witten, *Phys. Rev. D* **46** (1992) 5467 [hep-th/9208027].

[6] E. Witten, *Phys. Rev. D* **47** (1993) 3405 [hep-th/9210065].

[7] S. Shatashvili, *Phys. Lett. B* **311** (1993) 83 [hep-th/9303143].

[8] S. Shatashvili, *Algebra Anal.* **6** (1994) 215 [hep-th/9311177].

[9] J. Polchinski, *Phys. Rev. Lett.* **75** (1995) 4724 [hep-th/9510017].

[10] S. Shatashvili, unpublished; edited by A. Losev, G. Moore and S. Shatashvili, *Nucl. Phys. B* **522** (1998) 105 [hep-th/9707250]; talk at IHES in summer 1997.

[11] R. Dijkgraaf, E. Verlinde and H. Verlinde, *Nucl. Phys. B* **500** (1997) 43 [hep-th/9703030].

[12] A. Strominger, *Phys. Rev. Lett.* **58** (1987) 629.

[13] A. Sen, *JHEP* **9808** (1998) 012 [hep-th/9805170].

[14] A. Sen, *JHEP* **9912** (1999) 027 [hep-th/ 9911116].

[15] Anton A. Gerasimov and Samson L. Shatashvili, *JHEP* **0010** (2000) 034 [hep-th/0009103].

[16] D. Kutasov, M. Marino and G. Moore, *JHEP* **0010** (2000) 045 [hep-th/0009148].

[17] D. Ghoshal and A. Sen, *Normalisation of the background independent open string field theory* [hep-th/0009191].

[18] Anton A. Gerasimov and Samson L. Shatashvili, *JHEP* **0101** (2001) 019 [hep-th/0011009].

[19] D. Kutasov, M. Marino and G. Moore, *Remarks on Tachyon Condensation in Superstring Field Theory* [hep-th/0010108].

[20] J. Minahan and B. Zwiebach, *JHEP* **0009** (2000) 029 [hep-th/0008231]; J. Minahan and B. Zwiebach, *JHEP* **0103** (2001) 038 [hep-th/0009246].

[21] J. Harvey, D. Kutasov and E. Martinec, *On the relevance of tachyons* [hep-th/ 0003101].

[22] E. Akhmedov, A. Gerasimov and S. Shatashvili, *Non-abelain strunctures in the open string field theory*, edited by E. Akhmedov, A. Gerasimov and S. Shatashvili, On unification of **RR** couplings, to be published.

[23] A. Losev, N. Nekrasov and S. Shatashvili, *Nucl. Phys. B* **534** (1998) 549 [`hep-th/9711108`].

[24] A. Sen, *JHEP* **0011** (2000) 035.

SEMINAR 8

EXCEPTIONAL MAGIC

S.L. SHATASHVILI

Department of Physics,
Yale University, New Haven,
CT 06520-8120, U.S.A.

Contents

EXCEPTIONAL MAGIC

S.L. Shatashvili

Abstract

I review the properties of superstrings on the manifolds of exceptional holonomy.

Supersymmetric sigma models in two dimensions have been the source of many interesting ideas in the interplay between quantum field theories and geometry and topology of manifolds. In the context of superstring theories, viewing strings moving on a manifold leads to the use of sigma models as the building blocks of string vacua. To be a string vacuum the sigma model must lead to a conformal theory in two dimensions. Moreover to lead to spacetime supersymmetry, which is the only class of superstrings we know which are perturbatively stable, the manifold should admit covariantly constant spinors which can be used to define the supersymmetry transformation. It turns out that having a covariantly constant spinor already guarantees conformal invariance to one loop order in the sigma model perturbation theory and perhaps to all orders (with appropriate adjustments of the metric), so the study of manifolds admitting covariantly constant spinors seems like an important question for string theory. In general if we have an n dimensional Riemannian manifold the holonomy group is in $SO(n)$; however having a covariantly constant spinor the holonomy group is smaller and it is (a subgroup of) the little group which leaves a spinor of $SO(n)$ invariant.

Since superstrings live in 10 dimensions and we in 4, the most important physical case to study is the 6 dimensional manifolds with covariantly constant spinors. If we require only one spacetime supersymmetry, this means we need only one covariantly constant spinor (for a fixed chirality) and this leads to the manifolds of $SU(3)$ holonomy, *i.e.* the Calabi-Yau manifolds [2]. These manifolds have been investigated a great deal (for all

The talk is based on the old work of author with C. Vafa [1] and was presented on this workshop due to recent interest to the subject. The developments after 1994 are not reflected.

dimensions and not just 6) with interesting physical results. Among these one could mention that many classically singular Calabi-Yau manifolds lead to non-singular sigma models. Also there is a mirror phenomenon which means that strings on two inequivalent manifolds can lead to the same sigma model. Moreover there is a topological ring in these theories known as the chiral ring [3] which captures the deformation structure of the Calabi-Yau manifold.

One could consider odd dimensional manifolds with a minimal number of covariantly constant spinors by considering the product of Calabi-Yau with a circle. However, if one is interested in the minimum number of supersymmetries allowed, sigma models on Calabi-Yau manifolds and their products with a circle misses two special cases (for a review see [4]): first of all, in manifolds of 7 dimensions the minimum number of supersymmetries is given by a manifold of G_2 holonomy which has only one covariantly constant spinor as opposed to a manifold of $SU(3)$ holonomy times a circle which has 2. Furthermore in dimension 8 an $SU(4)$ holonomy manifold leads to 2 covariantly constant spinors whereas for an 8 dimensional manifold of $Spin(7)$ holonomy there will be only 1 covariantly constant spinor. The possible existence of these two special cases had been known for a long time [5]. It is very amusing that these two special cases can be used in physical models simply because the dimensions where they occur is less than 10, which means that if we were to compactify superstrings down to 3 or 2 dimensional Minkowski space and ask which manifolds would lead to minimal number of nonvanishing supersymmetries (1 for heterotic strings and 2 for type II strings) we would have to study sigma models on 7d manifolds of G_2 holonomy and 8d manifolds of $Spin(7)$ holonomy. The study of these two classes of sigma models is the subject of the present paper. This is a shortened version of a paper [6] to which we refer the interested reader for details.

Before we go on to describe some general properties of manifolds with G_2 and $Spin(7)$ holonomy, motivated by the success of the mirror conjecture for CY target spaces [7] let us make a conjecture which will prove helpful in clarifying the observations we shall make later:

Generalized Mirror Conjecture: the degree of ambiguity left by being unable to decipher all the topological aspects of the target manifold using the algebraic formulation of quantum field theories is precisely explained by having topologically inequivalent manifolds allowed by the ambiguity to lead to the same quantum field theory up to deformation in the moduli of the quantum field theory.

Until very recently the only known examples of manifolds of G_2 holonomy and $Spin(7)$ holonomy were non-compact manifolds [8]. The situation has dramatically changed recently due to the work of Joyce [9] who constructed the first compact examples of manifolds with G_2 and $Spin(7)$ holonomy, which we denote by M^7 and M^8 respectively. Just as in the Calabi-Yau case where the fact that the manifold has $SU(n)$ holonomy leads to the existence of a unique non-vanishing holomorphic covariantly constant n-form (and of course its conjugate), in these two exceptional cases a similar thing happens (see [10] Chapters 11 and 12): in the case of G_2 manifolds, there is a canonical 3 form ϕ and its dual which is a 4 form $*\phi$ which are covariantly constant and in the case of $Spin(7)$ there is a self-dual 4-form Ω. They can be locally written as follows: we choose a local vielbein so that the metric is $\sum e_i \otimes e_i$ where e_i are one forms and for the G_2 case, i runs from 1 to 7 and for the $Spin(7)$ case i runs from 1 to 8. Then these forms can be written as

$$\begin{aligned}
\phi &= e_1 \wedge e_2 \wedge e_7 + e_1 \wedge e_3 \wedge_6 + e_1 \wedge e_4 \wedge e_5 + \\
&\quad e_2 \wedge e_3 \wedge e_5 - e_2 \wedge e_4 \wedge e_6 + e_3 \wedge e_4 \wedge e_7 + \\
&\quad e_5 \wedge e_6 \wedge e_7 = \sum_{ijk} f_{ijk} e_i \wedge e_j \wedge e_k, \qquad (0.1)\\
*\phi &= e_1 \wedge e_2 \wedge e_3 \wedge e_4 + e_1 \wedge e_2 \wedge e_5 \wedge e_6 - \\
&\quad e_1 \wedge e_3 \wedge e_5 \wedge e_7 + e_1 \wedge e_4 \wedge e_6 \wedge e_7 + \\
&\quad e_2 \wedge e_3 \wedge e_6 \wedge e_7 + e_2 \wedge e_4 \wedge e_5 \wedge e_7 + \\
&\quad e_3 \wedge e_4 \wedge e_5 \wedge e_6, \qquad (0.2)\\
\Omega &= e_8 \wedge \phi - *\phi. \qquad (0.3)
\end{aligned}$$

Metric $\sum e_i \otimes e_i$ can be uniquely reconstructed from ϕ and $*\phi$ (or Ω in the case of $Spin7$).

Moreover it is true [9,11] that the dimension of moduli space of deformation of manifolds of G_2 holonomy is $b_3(M^7)$ and the dimension of the moduli space of deformation of manifolds of $Spin(7)$ holonomy is $b_4^-(M^8) + 1$, where $b_4^{\pm}$ denote the self-dual/anti-self-dual dimensions of $H^4(M^8)$. The simplest class of examples considered by Joyce involve toroidal orbifolds. In both the G_2 case and the $Spin(7)$ case he finds that there are in general many inequivalent ways of desingularizing the manifold, which we will be able to explain physically as a consequence of the generalized mirror conjecture stated above. In fact it is crucial to note that the dimension of the moduli space of the conformal theory is actually *bigger* than that predicted geometrically. The reason for this is that the possibility of using the antisymmetric two form to add a phase to the action has no geometrical analog. Therefore we have

$$\text{dim. moduli}_{\text{physical}} = \text{dim. moduli}_{\text{geometrical}} + b_2.$$

In particular for the G_2 case the dimension of sigma model moduli is b_2+b_3 and not b_3 and for the case of $Spin(7)$ the dimension of sigma model moduli is $b_4^-+b_2+1$.

Now we will unravel the extended symmetry algebras which underlie sigma models with $N=1$ superconformal symmetry on manifolds of G_2 and $Spin(7)$ holonomy.

Perhaps to make some aspects of the algebra that we obtain a little less mysterious it would be helpful to see *a priori* what we should expect to play the role that $U(1)$ plays for sigma models on manifolds of $SU(n)$ holonomy[1]. If we start with a sigma model on a Kähler manifold we have *a priori* a $U(n)$ symmetry. Having a holonomy in $SU(n)$ means that part of the symmetry is broken but we are left with an unbroken $U(1)=U(n)/SU(n)$. Similarly in the case of 7 dimensional manifolds of G_2 holonomy, *a priori* we have $SO(7)$ symmetry (more precisely $SO(7)$ current algebra at level 1). The holonomy of the manifold being in G_2 means that we are left with the residual symmetry $SO(7)/G_2$, which in geometrical terms is no longer a group, however, from the viewpoint of conformal theory it is a coset model. Computing its central charge we see that since $SO(7)$ at level 1 has central charge 7/2 and G_2 at level 1 has central charge 14/5, the central charge of the residual system is $\frac{7}{2}-\frac{14}{5}=\frac{7}{10}$, which is thus a tri-critical Ising model [12]! Similarly for the case of $Spin(7)$ manifolds one considers $SO(8)/Spin(7)$ which gives a central charge $4-7/2=1/2$, which is just the Ising model. Below we shall recover these facts directly as well as find out that these symmetries mix in a very interesting way with the $N=1$ superconformal algebra to obtain the extended symmetry algebra of our models.

1 G_2

We start from the flat 7 dimensional space, and construct the chiral operators which we expect to exist even after we perturb the metric to obtain a non-trivial G_2 holonomy. We of course expect to have the energy momentum tensor T and its superpartner G. Moreover the fact that a three form ϕ exists even after the perturbation suggests that one can add to $N=1$ superconformal generators $T=T_\mathrm{b}+T_\mathrm{f}=\frac{1}{2}\sum_1^7{:}J^iJ^i{:}-\frac{1}{2}\sum_1^7{:}\psi^i\partial\psi^i$ and $G=\sum_1^7{:}J^i\psi^i{:}$ a new spin 3/2 operator

$$\Phi=f_{ijk}\psi^i\psi^j\psi^k, \tag{1.1}$$

with the coefficients f_{ijk} defined by the G_2 invariant three form ϕ from the previous section; $J^i=\partial x^i$ with x^i being a bosonic sigma model coordinate.

[1] This line of thought was developed following a suggestion of E. Martinec.

$N=1$ generators are invariant under the rotation group $SO(7)$ and Φ is invariant only under the G_2 subgroup of $SO(7)$. If we compute the operator expansion of new generator Φ with itself we obtain:

$$\Phi(z)\Phi(w) = -\frac{7}{(z-w)^3} + \frac{6}{z-w}X(w), \tag{1.2}$$

where operator X has spin 2

$$X = - * \Phi + T_{\rm f}, \tag{1.3}$$

and is a linear combination of "dual" operator $*\Phi$ (defined by dual form $*\phi$) and a fermionic stress-tensor. Next step is to compute operator expansion of the operators X and Φ. We find that $\Phi = \frac{i}{\sqrt{15}} G_{\rm I}$ and $1/5X = T_{\rm I}$ form an $N=1$ superconformal algebra with Virasoro central charge 7/10 – tri-critical Ising model. This is not the end of story because now we need to deal with superpartners of new generators with respect to original $N=1$ algebra. This introduces two new operators of spins 2 and $\frac{5}{2}$ into the game; we will denote them by K and M respectively: $K=\alpha_{ijk}J^i\psi^j\psi^k, M = \beta_{ijkl}J^i\psi^j\psi^k\psi^l$, with α and β being $+1$ or -1 (obviously, these coefficients are uniquely defined by f_{ijk}):

$$G(z)\Phi(w) = \frac{1}{z-w}K(w), \tag{1.4}$$

$$G(z)X(w) = -\frac{1}{2}\frac{1}{(z-w)^2}G(w) + \frac{1}{z-w}M(w). \tag{1.5}$$

A nontrivial fact deeply related to "G_2 structure" is that operator expansion algebra formed by these six operators T, G, Φ, X, K and M closes [6]. We will use the properties of this algebra later and will present relevant commutation relations when necessary. Here we just mention that extended chiral algebra contains quadratic combinations in the right hand side and thus reminds one of W-algebra (for detalied description see [6])[2].

After extended chiral algebra is derived we can forget about the free field picture recalling that the perturbation will destroy the fact that the theory is free, but assume the existence of the algebra beyond free realization and study the corresponding conformal field theory. As a first step we have to find the spectrum of low lying states and in particular the spectra of Ramond ground states which carry the geometrical information about the

[2]The existence of extended symmetry for $N=1$ sigma model on G_2 manifold in classical approximation was previously mentioned in [13].

manifold. In this study it is extremely useful to note that our extended algebra contains two (non-commutative) $N = 1$ superconformal subalgebras: 1. Original $N = 1$ generated by G and T, and 2. $N = 1$ superconformal algebra generated by $G_{\rm I} = \frac{i}{\sqrt{15}}\Phi$ and $T_{\rm I} = -\frac{1}{5}X$. The latter is a very interesting one – it has a Virasoro central charge $\frac{7}{10}$ as predicted in the beginning of this section and is the tri-critical Ising model which is the only bosonic minimal model in the list of $N = 1$ superconformal minimal models. In addition a simple observation that

$$T_{\rm I}(z)T_{\rm r}(w) = 0, T = T_{\rm I} + T_{\rm r} \tag{1.6}$$

allows us to classify the highest weight representations of our algebra using two numbers: tri-critical Ising highest weight and eigenvalue of the zero mode of the remaining stress-tensor $T_{\rm r}$.

Now, at the beginning we consider only chiral sector (left-movers say). The theory is supersymmetric and thus we have two sectors – Neveu-Schwarz and Ramond. We shall see below that the $(-1)^F$ for the full theory can be identified with the $(-1)^{F_{\rm I}}$ which is the Z_2 symmetry of the tri-critical Ising model viewed as an $N = 1$ superconformal system. From the observation that total stress tensor can be written as a sum of two commutative Virasoro generators where one is tri-critical Ising, we conclude that unitary highest weight representations should have following tri-critical Ising dimensions:

$$NS\colon [0]_{\rm Vir}, \quad \left[\frac{1}{10}\right]_{\rm Vir}, \quad \left[\frac{6}{10}\right]_{\rm Vir}, \quad \left[\frac{3}{2}\right]_{\rm Vir}; \tag{1.7}$$

or in $N = 1$ terms

$$NS\colon \qquad [0], \qquad \left[\frac{1}{10}\right] \tag{1.8}$$

and

$$R\colon \qquad \left[\frac{7}{16}\right], \qquad \left[\frac{3}{80}\right]. \tag{1.9}$$

Supersymmetry requires that Ramond vacuum for the full theory has dimension $\frac{d}{16} = \frac{7}{16}$, and this leads to the following unitary highest weight representations of extended chiral algebra in the Ramond ground state (we use the notation $[\Delta_{\rm I}, \Delta_{\rm r}]$ for operators that correspond to Virasoro highest weights $|\Delta_{\rm I}, \Delta_{\rm r}\rangle$ with first dimension being the dimension of tri-critical Ising part and the second the dimension of the remaining Virasoro algebra $T_{\rm r}$):

$$R\colon \qquad \left|\frac{7}{16}, 0\right\rangle, \qquad \left|\frac{3}{80}, \frac{2}{5}\right\rangle. \tag{1.10}$$

It is one of the most remarkable facts for this theory that there exists a ground state in the Ramond sector which is entirely constructed out of the tri-critical Ising sector, namely the $|\frac{7}{16},0\rangle$ state. It is as if the tri-critical Ising model "knows" about the fact that the dimension of the manifold of interest is 7. As we will see this is crucially related to having an $N=1$ spacetime supersymmetry as well as the possibility of twisting the theory. In many ways the operator corresponding to this ground state plays the same role as the spectral flow operator in $N=2$ theories which is also entirely built out of the $U(1)$ piece of $N=2$. To have one spacetime supersymmetry we would be interested in realization of this algebra which has exactly one Ramond ground state of the form $|\frac{7}{16},0\rangle$ (we shall make this statement a little bit more precise when we talk about putting left- and right-movers together). In this regard it is crucial to note that in the tri-critical Ising model we have unique fusion rules for the operator $[\frac{7}{16}]$

$$\left[\frac{7}{16}\right]\left[\frac{7}{16}\right] = [0]_{\mathrm{Vir}} + \left[\frac{3}{2}\right]_{\mathrm{Vir}} = [0], \tag{1.11}$$

$$\left[\frac{7}{16}\right]\left[\frac{3}{80}\right] = \left[\frac{1}{10}\right]_{\mathrm{Vir}} + \left[\frac{6}{10}\right]_{\mathrm{Vir}} = \left[\frac{1}{10}\right]. \tag{1.12}$$

The existence of this operator in the Ramond sector allows us to *predict* the existence of certain states in the NS sector. This follows from the fact that it sits entirely in the tri-critical Ising part of the theory and its OPE with other fields depend only on the tri-critical Ising content of other state and thus by considering the OPE of the operator corresponding to $|\frac{7}{16},0\rangle$ state with the other states in the Ramond sector we end up with certain special NS states. From (1.11) we conclude that Ising spin field $[\frac{7}{16}]$ maps Ramond ground state $|\frac{7}{16},0\rangle$ to NS vacuum $|0,0\rangle$ and *vice versa*. More importantly when we consider the OPE of the $|\frac{7}{16},0\rangle$ state with $|\frac{3}{80},\frac{2}{5}\rangle$ we end up with a primary state in the NS sector of the form $|\frac{1}{10},\frac{2}{5}\rangle$, which has total dimension $\frac{1}{2}$ and is primary. This procedure can be repeated in opposite direction: tri-critical Ising model spin field $[\frac{7}{16}]$ maps primary field of NS sector $[\frac{1}{10},\frac{2}{5}]$ to an R ground state $|\frac{3}{80},\frac{2}{5}\rangle$. This leads to the prediction of existence of the following special states in NS sector:

$$NS\colon \qquad |0,0\rangle\,, \qquad \left|\frac{1}{10},\frac{2}{5}\right\rangle. \tag{1.13}$$

Note in particular that since the T_{r} part of the theory is un-modified as we go from the R sector to the NS sector. It is again quite remarkable that the state in the NS sector corresponding to $|\frac{1}{10},\frac{2}{5}\rangle$ is a primary field of dimension $1/2$ and so $G_{-1/2}$ acting on it is of dimension 1, preserving $N=1$

supersymmetry and thus a candidate for exactly marginal perturbation in the theory. We will use the extended chiral algebra below to show that indeed they lead to exactly marginal directions. Again the fact that this state has dimension 1/2 is a consequence of a miraculous relation between the dimension of tri-critical Ising model states. If one traces back one finds that it comes from the fact that $\frac{7}{16}-\frac{3}{80}+\frac{1}{10}=\frac{1}{2}$. In the above discussion we assumed that Z_2 fermion number assignment on any state is equal to the Z_2 grading for its tri-critical part alone which in particular implies that in the NS sector of the full theory only NS dimensions of tri-critical model show up and similarly in the R sector. Let us now discuss how this comes about. Our chiral algebra has three bosonic T, X, K and three fermionic G, Φ, M generators. We have the following tri-critical Z_2 assignments: $[0]^+, [\frac{1}{10}]^-, [\frac{6}{10}]^+, [\frac{3}{2}]^-$. To prove that $(-1)^F=(-1)^{F_{\mathrm{I}}}$ it suffices to derive tri-critical Ising dimensions of our generators and see if the two Z_2 assignments agree. Here we have to use relations presented in Appendix 2 of [6]; we have

$$\begin{aligned}
Ł_{-2}\,|0,0\rangle &= |2,0\rangle^+ + |0,2\rangle^+\,,\\
X_{-2}\,|0,0\rangle &= |2,0\rangle^+\,,\\
K_{-2}\,|0,0\rangle &= \left|\frac{6}{10},\frac{14}{10}\right\rangle^+\,, \qquad (1.14)\\
G_{-3/2}\,|0,0\rangle &= \left|\frac{1}{10},\frac{14}{10}\right\rangle^-\,, M_{-5/2}\,|0,0\rangle\\
&= a\left|\frac{1}{10},\frac{24}{10}\right\rangle^- + b\left|\frac{1}{10}+1,\frac{14}{10}\right\rangle^-\,. \qquad (1.15)
\end{aligned}$$

We see that in the assignment in above expressions $(-1)^F=(-1)^{F_{\mathrm{I}}}$ and thus we can use tri-critical gradings for the whole theory.

Now we are ready to discuss the non-chiral, left-right sector. We claim that only states in (R, R) ground state are:

$$\begin{aligned}
(R,R)\colon\qquad &\left|\left(\frac{7}{16},0\right)_L;\left(\frac{7}{16},0\right)_R;\pm\right\rangle,\\
&\left|\left(\frac{3}{80},\frac{2}{5}\right)_L;\left(\frac{3}{80},\frac{2}{5}\right)_R;\pm\right\rangle. \qquad (1.16)
\end{aligned}$$

where the significance of $\pm$ will be explained momentarily. We had two other possibilities of left-right combinations: $|(\frac{7}{16},0)_L;(\frac{3}{80},\frac{2}{5})_R;\pm\rangle$ and the same with exchange of L with R. The reason we didn't put these states in the list (1.16) is simple. If we use fusion rules (1.11) and (1.12) we see that primary field corresponding to first ground state in (1.16) acting on these additional states will lead (according to tri-critical Ising model fusion rules) to the highest weight state $|(0,0)_L;(\frac{1}{10},\frac{2}{5})_R\rangle$ in the Neveu-Schwarz sector. But, this operator has total dimension $\frac{1}{2}$ and is chiral, so, we get an additional chiral operator of half-integer spin in the theory which is not present in our original extended chiral algebra. This means that these additional states aren't present in the case of generic theory (which is assumed to have only chiral operators described in the beginning of this section). The $\pm$ signs next to the states are a reflection of the fact that since acting on the ground states we have $\{\Phi_0, \bar{\Phi}_0\} = 0, \Phi_0^2 = \bar{\Phi}_0^2 = \frac{6}{15}$, they form a 2 dimensional representation. The $\pm$ sign therefore reflects states with 2 different $(-1)^F$ assignments. Thus, Ramond ground states are coming in pairs – $\Phi_0|(\frac{7}{16},0,)_L;\ (\frac{7}{16},0)_R;\ +\rangle\ = |(\frac{7}{16},0)_L;\ (\frac{7}{16},0)_R;\ -\rangle, \Phi_0|(\frac{3}{80},\frac{2}{5})_L;\ (\frac{3}{80},\frac{2}{5})_R;\ +\rangle\ =\ |(\frac{3}{80},\frac{2}{5})_L;\ (\frac{3}{80},\frac{2}{5})_R;-\rangle$. Now we can better describe the relation of Ramond ground states with the cohomology of the manifold. Recall [14] that the number of ground states in the theory are exactly equal to the number of harmonic forms:

$$\mathrm{Tr}\ \exp(-\beta \mathrm{H})\Big|_{\beta\to\infty} = \sum_{i=0}^{n} b_i.$$

(Note that even though the number of ground states are equal to the number of cohomology elements of M there is no canonical correspondence.) The fact that states come in pairs is a consequence of the fact that in odd dimension the dual of every cohomology state is another cohomology state with different degree mod 2. So the Ramond + states correspond to even cohomology elements and $-$ to the odd ones. So now concentrating on the even cohomology elements in principle we could have $b_0 = 1, b_2, b_4$ as the elements (note that having no extra supersymmetry leads to having $b_6 = b_1 = 0$ which is correlated with the fact that we assume the $|(\frac{7}{16},0)_L;(\frac{7}{16},0)_R,+\rangle$ is unique). We see that we can only compute one extra number, and not two, which is the number of ground states involving the $\frac{3}{80}$ tri-critical piece for both left- and right-movers which we identify with $b_2 + b_4$.

Let us discuss the special NS states taking into account both the left- and right-moving degrees of freedom. Acting on all + Ramond ground states with the state $|(\frac{7}{16},0)_L;(\frac{7}{16},0)_R,+\rangle$ leads to (NS,NS) states

$$(NS,NS): \qquad \begin{array}{l} |(0,0)_L;(0,0)_R\rangle \\ \left|\left(\dfrac{1}{10},\dfrac{2}{5}\right)_L;\left(\dfrac{1}{10},\dfrac{2}{5}\right)_R\right\rangle. \end{array} \tag{1.17}$$

where the number of $|(\frac{1}{10},\frac{2}{5})_L;(\frac{1}{10},\frac{2}{5})_R\rangle$ states are the same as the states

$|(\frac{3}{80},\frac{2}{5})_L;(\frac{3}{80},\frac{2}{5})_R\rangle$ which is equal to b_2+b_4[3]. Moreover as we will argue later in this section each of all such NS operators are exactly marginal operators preserving the G_2 structure. This agrees with the geometrical facts discussed above – the dimension of conformal moduli space is $b_2+b_4=b_2+b_3$.

Before we address the question of marginal deformations of our conformal field theory let us discuss the relation of the above construction to 10-dimensional Superstring Theory compactified down to 3-dimensions. It is easy to show that if corresponding compact 7-dimensional manifold is a G_2-manifold we will have $N=2$ supersymmetry for type II strings and $N=1$ supersymmetry for heterotic strings in 3-dimension. Let us construct the corresponding supersymmetry generators using all the information that we already obtained. We have:

$$J_{L,R}=\mathrm{e}^{-\frac{\phi_{\mathrm{gh}}}{2}}S_3^{\alpha}\sigma^{L,R}_{\frac{7}{16}}. \tag{1.18}$$

Here ϕ_{gh} is a bosonized 10-dimensional ghost field, S_3^{α} are 3-dimensional spin fields and σ is tri-critical Ising model spin field that we had already discussed many times[4]. First we notice that J has dimension 1; dimension of 10-dimensional ghost part doesn't depends on compactification and always is equal to $\frac{3}{8}$, dimension of 3d spin field is 3. $\frac{1}{16}=\frac{3}{16}$ and dimension of sigma by definition is $\frac{7}{16}$, and all add up to 1. If we remember that σ has a unique OPE with vacuum [0] in the right hand side we can consider J as a chiral operator and this explains subscript L,R in (1.18). Now we can define 3d supersymmetry generators: $Q_{L,R}=\oint J_{L,R}$ and standard computation leads to supersymmetry algebra. Also, one finds that one of the supersymmetry transforms of $(\frac{3}{80},\frac{2}{5},)_{L,R}$ which is accompanied by spacetime spinor field and ghost degrees of freedom is simply the state $(\frac{1}{10},\frac{2}{5})_{L,R}$.

Now we would like to consider marginal deformations of our theory. As mentioned before we will show that marginal deformations are given by perturbation with dimension 1 operators of the form $G^L_{-1/2}G^R_{-1/2}[(1/10,2/5)_L;(1/10,2/5)_R]$; the dimension of this moduli space is b^2+b^3. In addition to showing that they preserve $N=1$ superconformal symmetry we need to show that they do not have any tri-critical piece in them, which would otherwise destroy the existence of the extended algebra in question. This follows because the full algebra was generated by the $N=1$ algebra together with the supersymmetry operator Φ of the tri-critical model. We will first show this fact by studying the content of above operator with respect to

[3]Also in principle we will get other higher dimension states such as $|\frac{3}{2},0)_L;(\frac{3}{2},0)_R\rangle$ or $|(\frac{6}{10},\frac{2}{5})_L;(\frac{6}{10},\frac{2}{5})_R\rangle$.

[4]This is a standard ansatz for target space supersymmetry current, see [15].

tri-critical Ising model. For this we have to apply the operator X_0. We have (it is enough to consider only chiral sector):

$$X_0 G_{-1/2}\left|\frac{1}{10},\frac{2}{5}\right\rangle = G_{-1/2}X_0\left|\frac{1}{10},\frac{2}{5}\right\rangle +$$
$$[X_0, G_{-1/2}]\left|\frac{1}{10},\frac{2}{5}\right\rangle =$$
$$\left(-\frac{1}{2}G_{-1/2} - M_{-1/2}\right)\left|\frac{1}{10},\frac{2}{5}\right\rangle = P. \tag{1.19}$$

It turns out that the right hand side of this equation is identically zero in our theory: $P = 0 - P$ is a null vector and thus $G_{-1/2}|\frac{1}{10},\frac{2}{5}\rangle$ is of the type $[(0,1)_L;(0,1)_R]$. So all we are left to show is that the deformation preserves conformal invariance.

For simplicity we will denote our perturbation by $G^L_{-1/2}A(z,\bar{z})$ (we will work with the chiral part below and thus will suppress $\bar{z}$ dependence and $G^R_{-1/2}$). The following proof is based on two facts:

1. Dixon [16] has shown using just $N = 1$ superconformal algebra that perturbation with a dimension 1 operator of the form $G_{-1/2}A$ is marginal if

$$\begin{aligned} F \;=\; & \langle G_{-1/2}A(z_1)G_{-1/2}A(z_2)G_{-1/2}A(z_3) \\ & G_{-1/2}A(z_4)...G_{-1/2}A(z_n)\rangle \end{aligned} \tag{1.20}$$

is a total derivative with respect to coordinates z_i, $i > 3$. Perturbation is truly marginal if n-point correlation function (1.20) integrated over all points, except the first three, is zero (first three points are fixed by $SL(2,C)$ invariance on sphere) and Dixon has shown that in $N = 1$ super conformal theory the integrand can be regulated in such a fashion that if it is a total derivative there are no contact term contributions[5].

2. As we have seen above $A(z)$ has a null vector (1.19) $P = 0$; in addition we need several relations between the generators of the extended algebra acting on $A(z)$ [6]:

$$M_{1/2}G_{-1/2}A(z) = -2X_0A(z) = A(z), \tag{1.21}$$
$$M_{-1/2}G_{-1/2}A(z) = \left(-X_{-1} + \frac{1}{2}L_{-1}\right)A(z), \tag{1.22}$$
$$M_{-3/2}G_{-1/2}A(z) = -L_{-1}X_{-1}A(z). \tag{1.23}$$

[5]If there is a total derivative in holomorphic variable by symmetry we get total derivative both in holomorphic and antiholomorphic coordinates $\partial_{z_i}\partial_{\bar{z}_j}$ and this is crucial in showing that there are no contact term contributions.

In fact, from [16] it follows that it is enough to prove that

$$I_0 = \langle G_{-1/2}A(z_1)A(z_2)A(z_3)G_{-1/2}A(z_4) \\ \ldots G_{-1/2}A(z_n)\rangle \tag{1.24}$$

is a total derivative $\frac{\partial}{\partial z_i}$, $i > 3$, of something. Our main strategy is to use the null vector condition (1.19) and contour deformation argument first for $G_{-1/2}A(z_1)$ in I and then the same argument but now replacing $G_{-1/2}A(z_1)$ by $-2M_{-1/2}A(z_1)$. First we insert $\oint_\infty (w - z_l)G(w)$ with contour around infinity in the correlator $\langle A(z_1)A(z_2)A(z_3)(\int G_{-1/2}A(z))^{n-3}\rangle$ and place the zero z_l at z_3 and z_2. After the contour deformation we get:

$$(z_1 - z_3)\langle G_{-1/2}A(z_1)A(z_2)A(z_3)\left(\int G_{-1/2}A(z)\right)^{n-3}\rangle + \\ (z_2 - z_3)\langle A(z_1)G_{-1/2}A(z_2)A(z_3)\left(\int G_{-1/2}A(z)\right)^{n-3}\rangle = 0, \tag{1.25}$$

$$(z_1 - z_2)\langle G_{-1/2}A(z_1)A(z_2)A(z_3)\left(\int G_{-1/2}A(z)\right)^{n-3}\rangle\rangle + \\ (z_3 - z_2)\langle A(z_1)A(z_2)G_{-1/2}A(z_3)\left(\int G_{-1/2}A(z)\right)^{n-3}\rangle = 0. \tag{1.26}$$

A similar formula can be written for M, which has dimension $5/2$, and thus we need to insert $\oint_\infty v(z)M(z)$ with v now having three zeros. Placing zeros at points z_1, z_2, z_3 we get:

$$(z_1 - z_2)(z_1 - z_3)\langle M_{-1/2}A(z_1)A(z_2)A(z_3)\left(\int G_{-1/2}A(z)\right)^{n-3}\rangle \\ + (z_2 - z_1)(z_2 - z_3)\langle A(z_1)M_{-1/2}A(z_2)A(z_3)\left(\int G_{-1/2}A(z)\right)^{n-3}\rangle \\ + (z_3 - z_1)(z_3 - z_2)\langle A(z_1)A(z_2)M_{-1/2}A(z_3)\left(\int G_{-1/2}A(z)\right)^{n-3}\rangle \\ + (n-3)\langle A(z_1)A(z_2)A(z_3)\int \mathrm{d}^2 z_4 (z_4 - z_1$$

$$+ z_4 - z_2 + z_4 - z_3)M_{1/2}G_{-1/2}A(z_4)\left(\int G_{-1/2}A(z)\right)^{n-4}\rangle$$

$$+ (n-3)\langle A(z_1)A(z_2)A(z_3)\int \mathrm{d}^2 z_4[(z_4 - z_1)(z_4 - z_2) + (z_4 - z_1)(z_4 - z_3)$$

$$+ (z_4 - z_2)(z_4 - z_3)]M_{-1/2}G_{-1/2}A(z_4)\left(\int G_{-1/2}A(z)\right)^{n-4}\rangle$$

$$+ (n-3)\langle A(z_1)A(z_2)A(z_3)\int \mathrm{d}^2 z_4(z_4 - z_1)(z_4 - z_2)(z_4 - z_3)$$

$$M_{-3/2}G_{-1/2}A(z_4)\left(\int G_{-1/2}A(z)\right)^{n-4}\rangle = 0. \tag{1.27}$$

Now we use relations (1.21), (1.22) and (1.23), and simply find that the last three terms combined lead to the integral of total derivative in z_4. More concretely, we write $L_{-1} = \partial$ and integrating by part in last term of (1.27) using (1.23) we cancel contribution of X_{-1} from (1.22) in the previous term; similarly, after integration by parts, second term from (1.22) kills the contribution of X_0 from (1.21)[6]. Thus, we drop these terms and replace $M_{-1/2}$ by $-\frac{1}{2}G_{-1/2}$. Combined with the identities (1.25) and (1.26) we see that $-\frac{3}{2}I = 0$. This leads to the proof of the statement that our perturbation is truly marginal. It is very satisfying that we used many different aspects of the extended chiral algebra for this proof.

2 *Spin*(7)

The story is completely parallel to the previous case and so we will be brief. As before, we take *Spin*(7) 4-form and replace e by target space fermions; thus we get a spin 2 operator – $\tilde{X}$:

$$\tilde{X} = \psi^8\Phi - X + 1/2\partial\psi^8\psi^8. \tag{2.1}$$

Pleasantly we find that the operator $T_{\mathrm{I}} = \frac{1}{8}\tilde{X}$ forms a Virasoro algebra with central charge $\frac{1}{2}$ and this means that the tri-critical Ising model that we had in the previous case is replaced by the ordinary, bosonic Ising model as predicted at the beginning of this section. As before, we have to check operator expansion with original $N = 1$ generators and we immediately find that $\tilde{X}$ has a super partner – $\tilde{M}$:

$$G(z)\tilde{X}(w) = 1/2(z-w)^2 G(w) + 1/(z-w)\tilde{M}(w), \tag{2.2}$$

[6]The terms that have been ignored here are total derivatives only if $2L_0 A = -2X_0 A = A$, and this condition is exactly satisfied by our choice of A.

with

$$\tilde{M} = J^8\Phi - \psi^8 K - M + 1/2\partial J^8\psi^8 - 1/2 J^8\partial\psi^8. \tag{2.3}$$

This operator has dimension $\frac{5}{2}$ and will play the role of the operator M. It turns out that these four operators, $G, T, \tilde{X}$ and $\tilde{M}$, form a closed operator expansion algebra, which again is a quadratic W-type algebra [6]. From this extended symmetry algebra it follows that one can again

decompose original stress-tensor as a sum of two commutative Virasoro generators:

$$T = T_{\mathrm{I}} + T_{\mathrm{r}}, \tag{2.4}$$

and we can classify our states again by two numbers: ising model highest weight and the eigenvalue of the zero mode of T_{r}: $|\Delta_{\mathrm{I}}, \Delta_{\mathrm{r}}\rangle$.

In chiral (left-mover) sector above observation immediately leads to the following content:

$$|0, \Delta_{\mathrm{r}}\rangle \quad \left|\frac{1}{2}, \Delta_{\mathrm{r}}\right\rangle \quad \left|\frac{1}{16}, \Delta_{\mathrm{r}}\right\rangle. \tag{2.5}$$

This means that in the Ramond sector, where we have to have dimension of ground state equal to $\frac{8}{16} = \frac{1}{2}$, (this follows from the requirement of supersymmetry – dimension of the Ramond ground state has to be equal to $\frac{c}{24}$) we should have the following highest weight states:

$$R\colon \quad \left|\frac{1}{2}, 0\right\rangle \quad \left|0, \frac{1}{2}\right\rangle \quad \left|\frac{1}{16}, \frac{7}{16}\right\rangle. \tag{2.6}$$

Amazingly enough there is again a unique state in the ground state built purely from the Ising piece, which is the $|\frac{1}{2}, 0\rangle$ state. This will now play an identical role to that of spin operator of tri-critical Ising model $[\frac{7}{16}]$ that mapped Ramond ground state to NS sector and *vice versa*; the specific property this operator had was that it had unique fusion rules with itself and other operator from Ramond ground state. In the $Spin(7)$ model this operator is replaced by the Ising model energy operator $\epsilon = [\frac{1}{2}]$; it has unique fusion rules and maps the Ramond ground state to a certain special NS highest weight states and *vice versa*:

$$NS\colon \quad |0, 0\rangle \quad \left|\frac{1}{2}, \frac{1}{2}\right\rangle \quad \left|\frac{1}{16}, \frac{7}{16}\right\rangle. \tag{2.7}$$

Here we are using Ising model fusion rules: $[\epsilon][\epsilon] = [0], [\epsilon][\sigma] = [\sigma], [\sigma][\sigma] = [0] + [\epsilon], [\sigma] = [\frac{1}{16}]$. The operator $(\frac{1}{16}, \frac{7}{16})$ has total dimension $\frac{1}{2}$ and clearly is a candidate for marginal deformation after acting by $G_{-1/2}$ on it. Again the fact that the dimension of this operator is $\frac{1}{2}$ is magical and related to the existence of spacetime supersymmetry.

In the Ising sector we have Z_2 symmetry: $\sigma \to -\sigma; 1, \epsilon \to 1, \epsilon$. We would like to show that corresponding $(-1)^{F_{\rm I}}$ is again identified with total $(-1)^F$. As in the G_2 case we have to compute Ising content of the generators of the chiral algebra. We have:

$$\begin{aligned} L_{-2}\left|0,0\right\rangle &= \left|2,0\right\rangle^+ + \left|0,2\right\rangle^+ , \\ \tilde{X}_{-2}\left|0,0\right\rangle &= \left|2,0\right\rangle^+ , \end{aligned} \tag{2.8}$$

$$\begin{aligned} G_{-3/2}\left|0,0\right\rangle &= \left|\frac{1}{16},\frac{23}{16}\right\rangle^- , \tilde{M}_{-5/2}\left|0,0\right\rangle \\ &= a\left|\frac{1}{16}+1,\frac{23}{16}\right\rangle^- + b\left|\frac{1}{16},\frac{39}{16}\right\rangle^- , \end{aligned} \tag{2.9}$$

and we had used the commutation relations from Appendix 1 of [6]. Now we see that $(-1)^{F_{\rm I}} = (-1)^F$. Thus we use Ising model fermion number assignment.

Let us now discuss non-chiral sector putting left and right sectors together. We claim that the content of RR ground state is given by the following combinations:

$$\begin{aligned} RR\colon \quad & \left|\left(\frac{1}{2},0\right)_L;\left(\frac{1}{2},0\right)_R\right\rangle, \left|\left(0,\frac{1}{2}\right)_L;\left(0,\frac{1}{2}\right)_R\right\rangle, \\ & \left|\left(0,\frac{1}{2}\right)_L;\left(\frac{1}{16},\frac{7}{16}\right)_R\right\rangle, \left|\left(\frac{1}{16},\frac{7}{16}\right)_L;\right. \\ & \left.\left(0,\frac{1}{2}\right)_R\right\rangle \left|\left(\frac{1}{16},\frac{7}{16}\right)_L;\left(\frac{1}{16},\frac{7}{16}\right)_R\right\rangle . \end{aligned} \tag{2.10}$$

Other possible combinations can be ruled out by similar arguments as in the G_2 case – using Ising model fusion rules they lead to existence of chiral half-integer spin operators that are not present in extended chiral algebra and thus such combinations can't appear in the ground state of a generic model.

We now wish to connect the above states as much as possible with the cohomology of the manifold. As far as even degrees are concerned they come from first, second and last state which all have $(-1)^F = +1$. Moreover we will connect all the NS versions of the last state with exactly marginal deformations, and so as discussed in Section 2 there are $1 + b_2 + b_4^-$ of them. Moreover the condition of having exactly one supersymmetry means that the first state is unique. So the second states are as many as $b_6 + b_4^+$.

The second and third state correspond to odd cohomology elements and each one are in number equal to $b_3 = b_5$.

Using the unique analog of spectral flow the above content of $(R,\ R)$ ground state after mapping to $(NS,\ NS)$ sector due to Ising model energy operator leads to following special states

$$
\begin{aligned}
&(NS,NS)\colon |(0,0)_L;(0,0)_R\rangle\,, \left|\left(\frac{1}{2},\frac{1}{2}\right)_L;\left(\frac{1}{2},\frac{1}{2}\right)_R\right\rangle,\\
&\left|\left(\frac{1}{2},\frac{1}{2}\right)_L;\left(\frac{1}{16},\frac{7}{16}\right)_R\right\rangle,\left|\left(\frac{1}{16},\frac{7}{16}\right)_L;\left(\frac{1}{2},\frac{1}{2}\right)_R\right\rangle,\\
&\left|\left(\frac{1}{16},\frac{7}{16}\right)_L;\left(\frac{1}{16},\frac{7}{16}\right)_R\right\rangle.
\end{aligned}
\tag{2.11}
$$

As we already mentioned operator $G^L_{-1/2}G^R_{-1/2}\ [(\frac{1}{16},\ \frac{7}{16})_L;\ (\frac{1}{16},\ \frac{7}{16})_R]$ is a candidate for marginal perturbation. Again we wish to show that the Ising structure is not affected by this perturbation. In other words we will show that this operator has zero dimension in Ising part. To demonstrate this fact we have to show that it is annihilated by $\tilde{X}_0$ (again we will keep only chiral part in this computation):

$$
\begin{aligned}
&\tilde{X}_0 G_{-1/2}\left|\frac{1}{16},\frac{7}{16}\right\rangle = G_{-1/2}\tilde{X}_0\left|\frac{1}{16},\frac{7}{16}\right\rangle +\\
&[\tilde{X}_0, G_{-1/2}]\left|\frac{1}{16},\frac{7}{16}\right\rangle = \left(\frac{1}{2}G_{-1/2} -\right.\\
&\left.\tilde{M}_{-1/2}\right)\left|\frac{1}{16},\frac{7}{16}\right\rangle = \tilde{P}.
\end{aligned}
\tag{2.12}
$$

$\tilde{P}$ is a null vector, $\tilde{P} = 0$, similar to the one in G_2 case (1.19); so, we see that $G_{-1/2}[\frac{1}{16},\frac{7}{16}]$ is of the type $(0,1)$ and if it is truly marginal it will preserve also extended $Spin(7)$ symmetry. In addition we got a very important null vector that will allow us to prove exact marginality as in the case of G_2.

In fact, the only information from extended chiral algebra we had used in the G_2 case to prove exact marginality was null vector condition (relation between $G_{-1/2}A$ and $M_{-1/2}A$) and commutation relation (1.21), (1.22), (1.23). Null vector condition $\tilde{P} = 0$ is practically the same (relative coefficient in $\tilde{P}$ doesn't play a key role) and analog of (1.21), (1.22), (1.23) are given by:

$$\tilde{M}_{1/2}G_{-1/2}A = -2\tilde{X}_0 A = -A, \tag{2.13}$$

$$\tilde{M}_{-1/2}G_{-1/2}A = (-\frac{1}{2}L_{-1} - \tilde{X}_{-1})A, \tag{2.14}$$

$$\tilde{M}_{-3/2}G_{-1/2}A = -L_{-1}\tilde{X}_{-1}A; \tag{2.15}$$

we use the notation $A = G^R_{-1/2}[(\frac{1}{16}, \frac{7}{16})_L; (\frac{1}{16}, \frac{7}{16})_R]$. Now the argument presented in the case of G_2 can be repeated identically with the same conclusion– our perturbation is truly marginal to all orders.

3 Topological twist

Let us briefly discuss the possibility of topological twisting (below we will describe the topological twist only for the case of G_2 manifolds; $Spin7$ case is very similar, see [6]). We have already seen that G_2 and $Spin(7)$ compactifications are very similar to $N = 2$ superconformal theories corresponding to $SU(n)$ or $N = 4$ corresponding to $Sp(n)$ holonomy. In particular they both lead to $N = 1$ spacetime supersymmetry upon heterotic compactification. In $N = 2$ (and similarly in the $N = 4$ [17]) there is a topological side to the story, which is deeply connected to spacetime supersymmetry in the compactified theory. Basically the spectral flow operator, which is the same operator used to construct spacetime supersymmetry operator is responsible for the twisting. Twisting is basically the same as insertions of $2g - 2$ of these operators at genus g. The spectral flow operator is constructed entirely out of the $U(1)$ piece of the $N = 2$ theory and since the spectral flow operator can be written as

$$\sigma = \exp(i\rho/2) \qquad J = \partial\rho,$$

the twisting becomes equivalent to modifying the stress tensor by

$$T \to T + \frac{\partial^2\rho}{2},$$

where J is the $U(1)$ current of $N = 2$. With this change in the energy momentum tensor the central charge of the theory becomes zero. Once one does this twisting the chiral fields which are related by spectral flow operator to the ground states of the Ramond sector become dimension 0 and form a nice closed ring known as the chiral ring [3]. Given the similarities to $N = 2$ we would like to explore analogous construction for G_2 and $Spin(7)$. In the $N = 2$ case the main modification in the theory was in the $U(1)$ piece of the theory. Therefore also here we expect the main modifications to be in the tri-critical Ising piece for the G_2 and in the Ising piece for the $Spin(7)$ case.

Let us concentrate on the sphere.

As noted above abstractly, on the sphere one can define twisted correlation functions by insertion of two spin fields ($\sigma_{\frac{7}{16}}$ in G_2 case and $\sigma_{\frac{1}{2}}$ in $Spin(7)$ case) in NS sector:

$$\begin{aligned}&\langle V_1(z_1,\bar{z}_1)...V_n(z_n,\bar{z}_n)\rangle_{\text{twisted}} = \\ &\langle \sigma(0)V_1(z_1,\bar{z}_1)...V_n(z_n,\bar{z}_n)\sigma(\infty)\rangle_{\text{untwisted}}.\end{aligned} \tag{3.1}$$

Let us check this idea by bosonizing Ising sector. First we discuss G_2. Bosonized tri-critical Ising supercurrent and stress tensor have the form:

$$\Phi = \mathrm{e}^{\frac{3i}{\sqrt{5}}\varphi}, X = (\partial\varphi)^2 + \frac{1}{4\sqrt{5}}\partial^2\varphi. \tag{3.2}$$

At the same time we can write down the chiral primaries in terms of boson φ:

$$\begin{aligned} &[0] = I, \left[\frac{1}{10}\right] = \mathrm{e}^{\frac{i}{\sqrt{5}}\varphi}, \left[\frac{6}{10}\right] = \mathrm{e}^{\frac{2i}{\sqrt{5}}\varphi}, \\ &\left[\frac{7}{16}\right] = \mathrm{e}^{\frac{-5i}{4\sqrt{5}}\varphi}, \left[\frac{3}{80}\right] = \mathrm{e}^{\frac{-i}{4\sqrt{5}}\varphi}. \end{aligned} \tag{3.3}$$

Background charge is $-2\alpha_0 = -\frac{1}{2\sqrt{5}}$ and one can check that central charge is correct $c = 1 - 24\alpha_0^2 = \frac{7}{10}$. Insertion of spin fields according to (3.3) is equivalent to a change in background charge $-2\alpha_0 \to -2\tilde{\alpha}_0 = -\frac{3}{\sqrt{5}}$, and thus new stress-tensor that replaces X is $X_{\mathrm{tw}} = (\partial\varphi)^2 - \frac{3}{2\sqrt{5}}\partial^2\varphi$ with central charge $\tilde{c}_{\mathrm{tw}} = 1 - 24\tilde{\alpha}_0^2 = -\frac{98}{10}$. If we compute total central charge (we don't touch remaining sector T_{r} by our twist) since the central charge of T_{r} is equal to $21/2 - 7/10 = 98/10$ and we have not changed it by the twisting we get: $c_{\mathrm{twist}} = -98/10 + 98/10 = 0$. This is indeed remarkable! It is the strongest hint for the existence of a topological theory. Obviously, before twisting we have a minimal model and correct vertex operators are given by above formulas dressed by screening operators (see [18–20] charges are: $\alpha_+ = \frac{5}{2\sqrt{5}}, \alpha_- = -\frac{2}{\sqrt{5}}$. At the same time after twisting we get a model which is not a minimal model and if now correlation functions of above operators aren't non-zero they can't be screened. Thus, after twisting when we calculate correlation functions we could forget about dressing by screening operators and do just naive computation. This simplifies the story. Vertex operators are the same, but their dimensions are now different. We have:

$$\left[\frac{1}{10}\right] \longrightarrow \left[-\frac{2}{5}\right], \left[\frac{6}{10}\right] \longrightarrow \left[-\frac{2}{5}\right], \left[\frac{3}{2}\right] \longrightarrow [0]. \tag{3.4}$$

Note that in particular we learn that the special states we get in the NS sector have total dimension zero in the topological theory:

$$\left|\frac{1}{10}, \frac{2}{5}\right\rangle \to \left|-\frac{2}{5}, \frac{2}{5}\right\rangle, \left|\frac{6}{10}, \frac{2}{5}\right\rangle \to \left|-\frac{2}{5}, \frac{2}{5}\right\rangle, \left|\frac{3}{2}, 0\right\rangle \to |0, 0\rangle \cdot \tag{3.5}$$

Which is what one would expect of topological observables. Moreover they do seem to form a ring under multiplication [6].

The expressions for the shift in the dimension of the tri-critical piece together with the fact that we have already discussed the tri-critical content of the generators of the chiral algebra means that we can deduce their twisted dimension. We find that they all have shifted to integer dimensions, another hallmark of topological theories: $G - \text{dim}.1, \Phi - \text{dim}.0, M - \text{dim}.2$, plus we got dimension 1 bosonic operator K. Thus, after twisting, G is a candidate for BRST current of the topological theory and M – for antighost. To prove the last statement we need to show that OPE's of G with itself, as well as M with itself don't have simple poles (or at least do not contribute to the amplitudes) and in addition, G with M have the modified stress-tensor as a residue of simple pole. This would need to be verified. It should also be verified that with this sense of topological BRST invariance the above special states in the NS sector indeed are BRST invariant.

References

[1] S. Shatashvili and C. Vafa, *Selecta Matem., New Ser.* **1**(2) (1995) 347 [hep-th/9407025].

[2] P. Candelas, G. Horowitz, A. Strominger and E. Witten, *Nucl. Phys. B* **258** (1985) 46.

[3] W. Lerche, C. Vafa and N. Warner, *Nucl. Phys. B* **324** (1989) 427.

[4] McKenzie Y. Wang, *Ann. Global Anal. Geom.* **7** (1989) 59.

[5] M. Berger, *Bull. Soc. Math. France* **83** (1955) 279.

[6] S.L. Shatashvili and C. Vafa, *Superstrings and Manifolds of Exceptional Holonomy*, Preprint HUTP–94/A016, IASSNS-HEP-94/47 [hep-th/9407025].

[7] *Essays on Mirror Manifolds*, edited by S.-T. Yau (International Press, 1992).

[8] R.L. Bryant, *Ann. Math.* **126** (1987) 525; R.L. Bryant and S.M. Salamon, *Duke Math. J.* **58** (1989) 829.

[9] D.D. Joyce, *Compact 7-manifolds with holonomy* G_2,I,II, IAS preprints 1994; *Compact Riemannian 8-manifolds with Exceptional Holonomy Spin*(7), in preparation.

[10] S.L. Salamon, *Riemannian geometry and holonomy groups*, Pitman Research notes in mathematics series No. 201 (published by Longman, Harlow, 1989).

[11] R.L. Bryant and F.R. Harvey, unpublished.

[12] P. Goddard and D. Olive, *Nucl. Phys. B* **257** (1985) 226.

[13] P.S. Howe and G. Papadopolous, *Comm. Math. Phys.* **151** (1993) 467.

[14] E. Witten, *J. Diff. Geometry* **17** (1982) 661.

[15] T. Banks, L. Dixon, D. Friedan and E. Martinec, *Nucl. Phys. B* **299** (1988) 613.

[16] L. Dixon, *Some world sheet properties of superstring compactifications, on orbifolds and otherwise*, lecture given at the 1987 ICTP Summer Workshop (Trieste, Italy, 1987).

[17] N. Berkovits and C. Vafa, *N=4 Topological Strings*, Preprint HUTP-94/A018, KCL-TH-94-12, [hep-th/9407190].

[18] B. Feigin and D. Fucks, *Func. Anal. i ego Priloz.* **17** (1983) 241.

[19] V. Dotsenko, V. Fateev, *Nucl. Phys. B* **240** [FS12] (1984) 312.

[20] G. Felder, *Nucl. Phys. B* **324** (1989) 548.