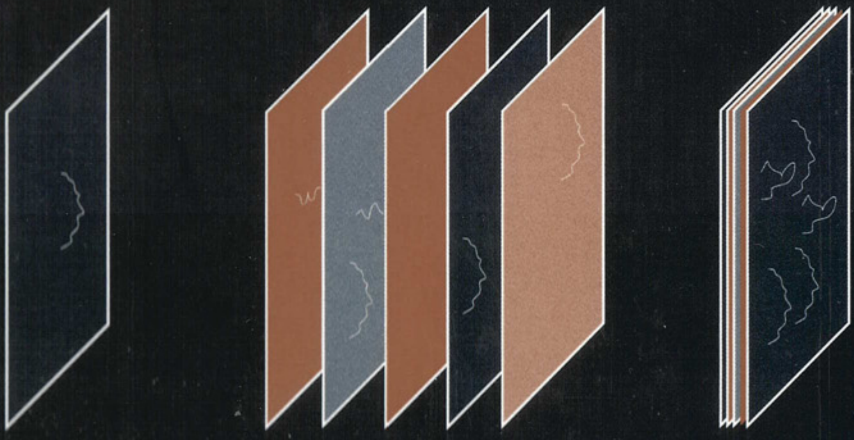
STRINGS, BRANES AND EXTRA DIMENSIONS

TASI 2001



Editors

Steven S. Gubser

Joseph D. Lykken

World Scientific

STRINGS, BRANES AND EXTRA DIMENSIONS

TASI 2001

This page intentionally left blank

STRINGS, BRANES AND EXTRA DIMENSIONS

TASI 2001

Boulder, Colorado, USA 4 – 29 June 2001

Editors

Steven S. Gubser

Princeton University, USA

Joseph D. Lykken

Fermi National Accelerator Laboratory, USA

World Scientific

NEW JERSEY • LONDON • SINGAPORE • SHANGHAI • HONG KONG • TAIPEI • CHENNAI

Published by

World Scientific Publishing Co. Pte. Ltd.
5 Toh Tuck Link, Singapore 596224
USA office: Suite 202, 1060 Main Street, River Edge, NJ 07661
UK office: 57 Shelton Street, Covent Garden, London WC2H 9HE

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library.

Brane Graphics on the cover: Courtesy of E. D'Hoker and D. Z. Freedman

STRINGS, BRANES AND EXTRA DIMENSIONS
TASI 2001

ISBN 981-238-788-9

Printed by Fulsland Offset Printing (S) Pte Ltd, Singapore

Foreword

Particle physics is undergoing a revolution. New ideas from string theory have stimulated a deep rethinking of diverse problems of physics involving black holes, quantum chromodynamics, and cosmology. At the same time, formal developments are deepening our understanding of string theory itself. Experience has shown that formal advances, such as string dualities and D-branes, often lead to new ideas for connecting string theory to testable extensions of the Standard Model. And indeed, the rich physics of branes has given birth to a whole new branch of particle phenomenology and model-building, related to direct and indirect experimental signatures of extra dimensions. These signatures may be observable at the Tevatron and LHC colliders, or at a future Linear Collider.

The ongoing revolution has been propelled by a number of surprising shocks to the orthodoxy of particle theory, an orthodoxy which accompanied the rise of the Standard Model three decades ago. String theory, developed as a theory of quantum gravity, has turned out to be a remarkable tool for gaining profound insights about gauge theories, especially supersymmetric gauge theories. It turns out that string theory in certain well-defined, negatively curved backgrounds is equivalent to a non-abelian gauge theory. The original concept of strings from the 1970s as idealizations of QCD flux tubes has reemerged in this exact duality.

String theory makes a robust prediction that there are extra dimensions of space, and it provides a number of efficient physical mechanisms for hiding their existence from our ordinary experience. String theory also provides a concrete realization of holography, a basic principle invoked to explain how black holes can exist in a quantum universe. String vacua generically contain branes, and we have only begun to map out the huge moduli space of possible string compactifications. This suggests that much of the complexity and broken symmetries of particle physics may simply reflect that we are living in a complex symmetry-breaking ground state of the higher dimensional string dynamics.

The lectures at the 2001 TASI school provided a fascinating snapshot of this revolution in the making. We are deeply grateful to all of the lecturers for their superb contributions both inside and outside the lecture hall. The nearly 60 students who participated in the 2001 TASI, with their energy, talent, and enthusiasm, also contributed greatly to the success of the school. A total of 23 student seminars were given during TASI; this activity was organized by the students themselves under the leadership of Andreas Birkedal-Hansen and Fred Leblond.

TASI 2001 took place on the beautiful Boulder campus of the University of Colorado. It ran from June 4 to June 29, 2001. Further information and background on the school can be found on the web at

`http://physics.colorado.edu/tasi01_annc.html`

`http://physics.colorado.edu/tasi01_info.html`

We would like to acknowledge the extraordinary efforts of general director K. T. Mahanthappa and of Kathy Oliver, who handled the thousand tasks that make a school run smoothly. We thank Mu-Chun Chen and Alex Flournoy for their important contributions to the everyday operations of the school and Tom DeGrand for assistance with computers. Tom DeGrand also continued his grand tradition of leading the student hikes and minimizing losses due to inexperience or over-enthusiasm.

The 2001 TASI school was made possible by the generous support of the University of Colorado, the US Department of Energy, and the National Science Foundation.

Steve Gubser and Joe Lykken
TASI 2001 Program Directors

CONTENTS

Lecturers

E. D'Hoker
S. P. deAlwis
T. DeGrand
K. Dienes
L. Dolan
G. Dvali
R. K. Ellis
D. Freedman
S. Gubser
J. Hewett
G. Kane
L. Krauss
A. Liddle
Z. Ligeti
J. Lykken
B. Ovrut
K. Rajagopal
M. Sher
M. Spiropulu
M. Strassler
W. Taylor
D. Waldram
B. Zwiebach

Local organizing committee

S. P. deAlwis
T. DeGrand
A. Hasenfratz
N. Irges
F. Knechtli
K. T. Mahanthappa

Director

S. Gubser

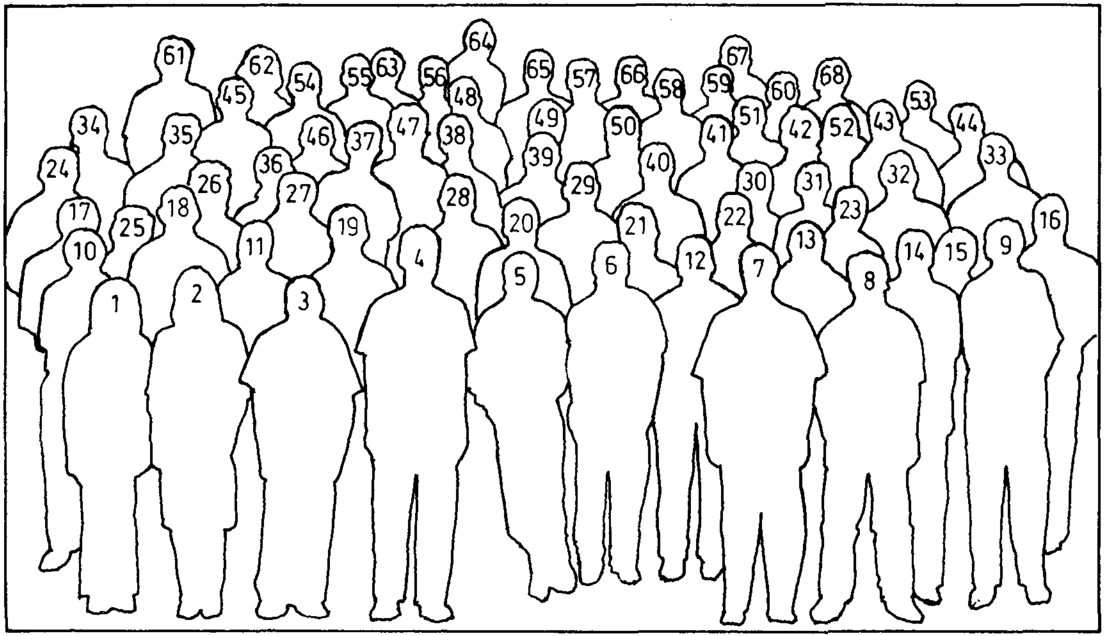
Student participants

Yashar Aghababaie
Anirban Basu
Alexey Batrachenko
Oren Ben-Bassat
Aaron Bergman
Andreas Birkedal-Hanson
Konstantin Bobkov
Kanokkuan Chaicherdsakul
Mu Chun Chen
Nihat Deger
Lisa Dyson
Theodore Erler
Jarah Evslin
Alex Flournoy
Patrick Fox
Shrihari Gopalakrishna
Elie Gorbatov
Chethan Gowdigere
Mariana Grana
Hector Hugo Hernandez
Sungho Hong
Cheien-Ian Hsueh
Jing Jiang
Li Jiang
Nicholas Jones
Joanna Karczmarek
Marko Kolanovic
Tibor Kucs
Mark Laidlaw
Peter Langfelder
Frederic Leblond
Matthew Lippert
Irit Maor
Andrei Micu
Octavian Micu
Alexander Mitov
Indrajit Mitra
Jorge Moreno
Filipe Moura
Brandon Murakami
Arkadas Ozakin
Hyukjae Park
F. Perez
Frank Petriello
Michele Redi
Vyacheslav Rychkov
Anton Ryzhov
Darius Sadri
Hisham Sati
Matsu Sato
Andrei Starinets
Florea Stoica
Liantao Wang
Brian Wecht
Brookie Williams
David Winters

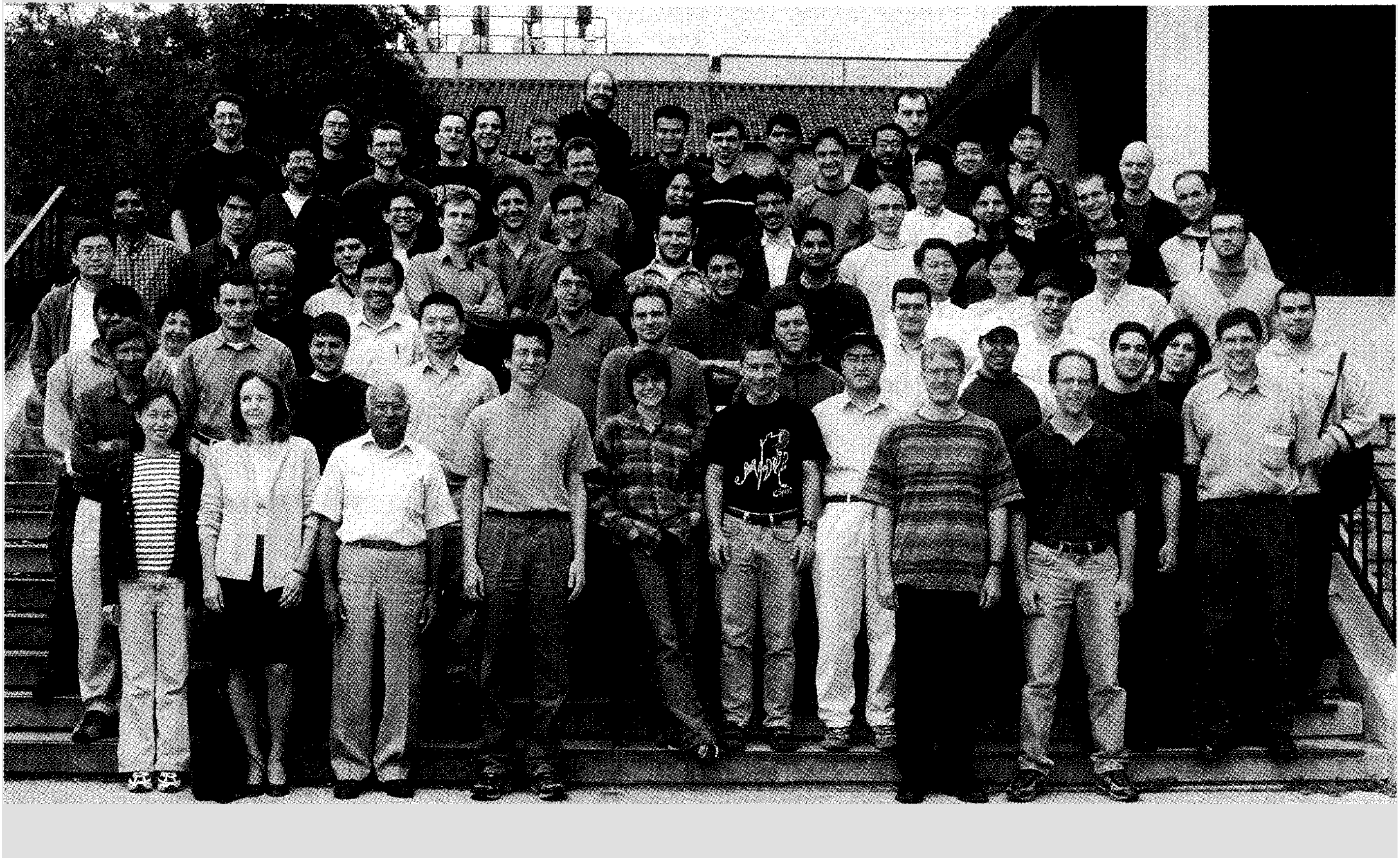
Student Seminars

1. A. Birkedal-Hansen, "On the Fate of the Neutralino Dark Matter," June 11.
2. F. Leblond, "Brane World Sum Rules and AdS Soliton," June 11.
3. B. Williams, "D-Branes and Cosmology," June 12.
4. S. Gopalakrishna, "Collider Signals of a Superlight Gravitino," June 12.
5. L.-T. Wang, "Intersecting D-Branes and Cosmology," June 14.
6. K. Kennaway, "Topological Sigma Models," June 14.
7. S. Rychkov, "Gauge Fields/Strings Duality and the Loop Equation," June 15.
8. A. Ryzhov, "1/4-BPS Operators in N=4 SYM," June 15.
9. M. Lippert, "AdS/CFT, Precursor and Holography," June 18.
10. I. Maor, "Measuring Equation of State of Dark Energy," June 18.
11. M. Grana, "AdS/CFT with Partial SUSY-Breaking," June 19.
12. A. Starinets, "Absorption by Non-Extremal Black Branes and AdS/CFT," June 19.
13. F. Moura, "Superspace Supersymmetrization of Higher Derivative Actions," June 20.
14. I. Mitra, "Stability of $AdS_p \times M_q$ Compactification without SUSY," June 20.
15. J. Moreno, "Non-Commutative Brane Worlds," June 21.
16. M. Sato, "BPS Bound States of D6-Branes and Lower Dimensional D-Branes," June 21.
17. J. Evslin, "Gauge Theory from M-Theory," June 25.
18. M.-C. Chen, "Fermion Masses and Mixing, CP-Violation in a SUSY $SO(10) \times U(2)_F$ Model," June 25.
19. F. Stoica, "Brane-World Perspective on the Hierarchy and Cosmological Constant Problem," June 27.
20. A. Flournoy, "One-Dimensional Quantum Mechanics (D-Branes as Non-Commutative Solitons)," June 27.
21. P. Fox, "Radion in the AdS and dS Brane Worlds," June 28.
22. B. Murakami, "Neutralino Dark Matter or Flavor Violating Z' Bosons," June 28.
23. H. Sati, "SUSY of the Brane Worlds."

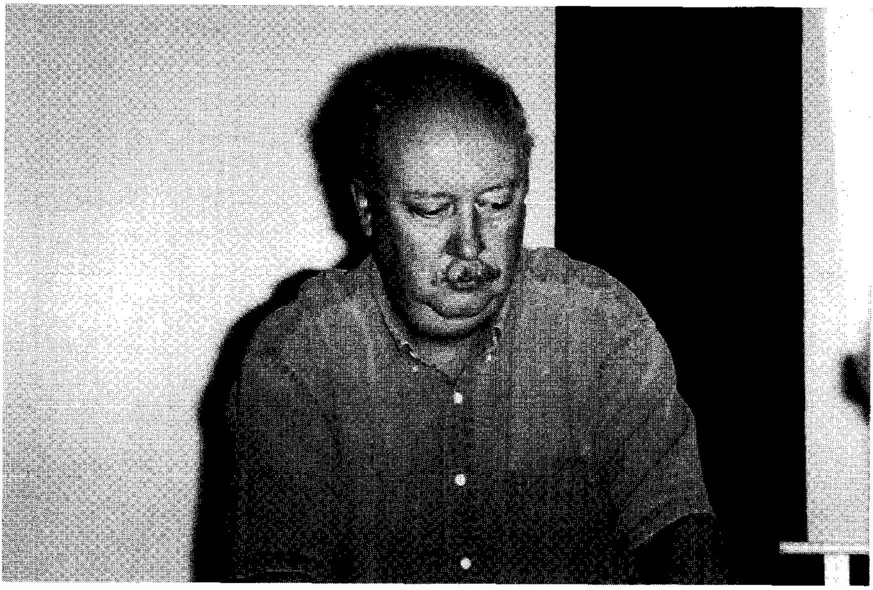
These seminars were organized by Andreas Birkedal-Hansen and Frederic Leblond.



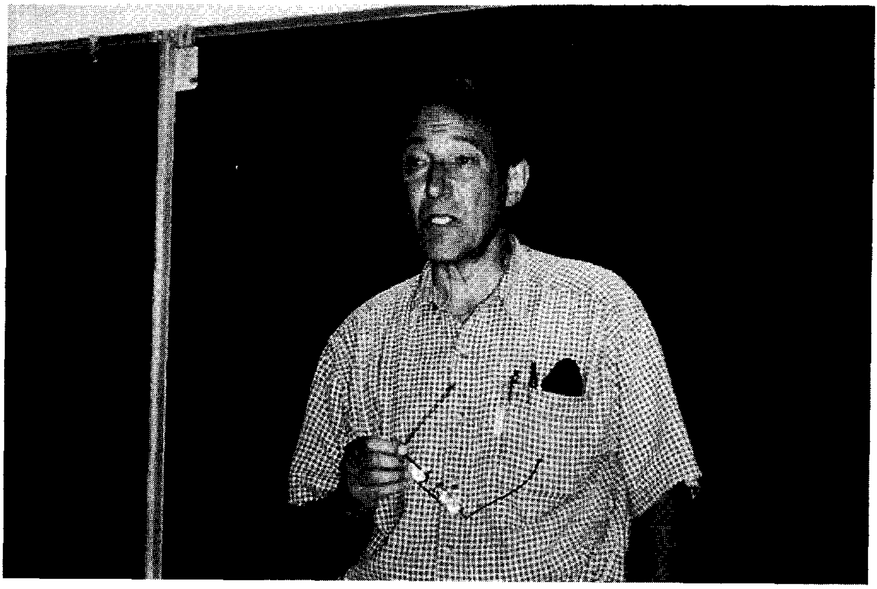
1. Kanokkuan Chaicherdsakul
2. Louise Dolan
3. KT Mahanthappa
4. Steve Gubser
5. Joanna Karczmarek
6. Peter Langfelder
7. Joseph Lykken
8. Washington Taylor
9. Barton Zwiebach
10. Shanta de Alwis
11. Frederic Leblond
12. Chien-Lan Hsueh
13. Jorge Moreno
14. Oren Ben-Bassat
15. Maria Spiropulu
16. Marko Kolanovic
17. Chethan Gowdigere
18. Konstantin Bobkov
19. Li Jiang
20. Tibor Kucs
21. Hisham Sati
22. Nihat Sadik Deger
23. Vyacheslav Rychkov
24. Hyukjae Park
25. Kathy Oliver
26. Lisa Dyson
27. Liantao Wang
28. Aaron Bergman
29. Arta Sadrzadeh
30. Brandon Murakami
31. Mu-Chun Chen
32. Andrei Starinets
33. Alex Flournoy
34. Indrajit Mitra
35. Ted Erler
36. Michele Redi
37. Alexey Batrachenko
38. Matthew Lippert
39. Octavian Micu
40. Shrihari Gopalakrishna
41. Darius Sadri
42. Jarah Evslin
43. Brook Williams
44. Elie Gorbatov
45. Arkadas Ozakin
46. Brian Wecht
47. Alexander Mitov
48. Andrei Micu
49. Mariana Grana
50. Filipe Moura
51. Burt Ovrut
52. Irit Maor
53. Kristian Kennaway
54. Anton Ryzhov
55. Andreas Birkedal-Hanson
56. Patrick Fox
57. Mark Laidlaw
58. Nicholas Jones
59. Anirban Basu
60. Sungho Hong
61. Frank Petriello
62. David Winters
63. Yashar Aghababaie
64. Thomas De Grand
65. Jing Jiang
66. Hector Hernandez
67. Florea Stoica
68. Matsuo Sato



This page intentionally left blank



ERIC D'HOKER



DANIEL Z. FREEDMAN

This page intentionally left blank

SUPERSYMMETRIC GAUGE THEORIES AND THE ADS/CFT CORRESPONDENCE

ERIC D'HOKER

Department of Physics and Astronomy
University of California, Los Angeles, CA 90095
E-mail: dhoker@physics.ucla.edu

DANIEL Z. FREEDMAN

Department of Mathematics and Center for Theoretical Physics
Massachusetts Institute of Technology, Cambridge, MA 02139
E-mail: dzf@math.mit.edu

In these lecture notes we first assemble the basic ingredients of supersymmetric gauge theories (particularly N=4 super-Yang-Mills theory), supergravity, and superstring theory. Brane solutions are surveyed. The geometry and symmetries of anti-de Sitter space are discussed. The AdS/CFT correspondence of Maldacena and its application to correlation functions in the the conformal phase of N=4 SYM are explained in considerable detail. A pedagogical treatment of holographic RG flows is given including a review of the conformal anomaly in four-dimensional quantum field theory and its calculation from five-dimensional gravity. Problem sets and exercises await the reader.

1. Introduction

These lecture notes describe one of the most exciting developments in theoretical physics of the past decade, namely Maldacena's bold conjecture concerning the equivalence between superstring theory on certain ten-dimensional backgrounds involving Anti-de Sitter space-time and four-dimensional supersymmetric Yang-Mills theories. This AdS/CFT conjecture was unexpected because it relates a theory of gravity, such as string theory, to a theory with no gravity at all. Additionally, the conjecture related highly non-perturbative problems in Yang-Mills theory to questions in classical superstring theory or supergravity. The promising advantage of the correspondence is that problems that appear to be intractable on one side may stand a chance of solution on the other side. We describe the initial conjecture, the development of evidence that it is correct, and some further applications.

The conjecture has given rise to a tremendous number of exciting directions of pursuit and to a wealth of promising results. In these lecture notes, we shall present a quick introduction to supersymmetric Yang-Mills theory (in particular of $\mathcal{N} = 4$ theory). Next, we give a concise description of just enough supergravity and superstring theory to allow for an accurate description of the conjecture and for discussions of correlation functions and holographic flows, namely the two topics that constitute the core subject of the lectures.

The notes are based on the loosely coordinated lectures of both authors at the 2001 TASI Summer School. It was decided to combine written versions in order to have a more complete treatment. The bridge between the two sets of lectures is Section 8 which presents a self-contained introduction to the subject and a more detailed treatment of some material from earlier sections.

The AdS/CFT correspondence is a broad principle and the present notes concern one of several pathways through the subject. An effort has been made to cite a reasonably complete set of references on the subjects we discuss in detail, but with less coverage of other aspects and of background material.

Serious readers will take the problem sets and exercises seriously!

1.1. *Statement of the Maldacena Conjecture*

The Anti-de Sitter/Conformal Field Theory (AdS/CFT) correspondence, as originally conjectured by Maldacena, advances a remarkable equivalence between two seemingly unrelated theories. On one side (the *AdS-side*) of the correspondence, we have 10-dimensional Type IIB string theory on the product space $\mathrm{AdS}_5 \times \mathrm{S}^5$, where the Type IIB 5-form flux through S^5 is an integer N and the equal radii L of AdS_5 and S^5 are given by $L^4 = 4\pi g_s N \alpha'^2$, where g_s is the string coupling. On the other side (the *SYM-side*) of the correspondence, we have 4-dimensional super-Yang-Mills (SYM) theory with maximal $\mathcal{N} = 4$ supersymmetry, gauge group $SU(N)$, Yang-Mills coupling $g_{YM}^2 = g_s$ in the conformal phase. The AdS/CFT conjecture states that these two theories, including operator observables, states, correlation functions and full dynamics, are equivalent to one another.[1,2,3]

Indications of the equivalence had appeared in earlier work.[4,5,6] For a general review of the subject, see Ref. 7.

In the strongest form of the conjecture, the correspondence is to hold for all values of N and all regimes of coupling $g_s = g_{YM}^2$. Certain limits of the conjecture are, however, also highly non-trivial. The 't Hooft limit on the SYM-side,[8] in which $\lambda \equiv g_{YM}^2 N$ is fixed as $N \to \infty$ corresponds to *classical string theory on* $\mathrm{AdS}_5 \times \mathrm{S}^5$ (no string loops) on the AdS-side. In this sense, classical string theory on $\mathrm{AdS}_5 \times \mathrm{S}^5$ provides with a classical Lagrangian formulation of the large N dynamics of $\mathcal{N} = 4$ SYM theory, often referred to as the *masterfield equations*. A further limit $\lambda \to \infty$ reduces classical string theory to classical Type IIB supergravity on $\mathrm{AdS}_5 \times \mathrm{S}^5$. Thus, strong coupling dynamics in SYM theory (at least in the large N limit) is mapped onto classical low energy dynamics in supergravity and string theory, a problem that offers a reasonable chance for solution.

The conjecture is remarkable because its correspondence is between a 10-dimensional theory of gravity and a 4-dimensional theory without gravity at all, in fact, with spin ≤ 1 particles only. The fact that all the 10-dimensional dynamical degrees of freedom can somehow be encoded in a 4-dimensional theory living at the boundary of AdS_5 suggests that the gravity bulk dynamics results from a *holographic image* generated by the dynamics of the boundary theory,[9] see also Ref. 10. Therefore, the correspondence is also often referred to as *holographic*.

1.2. *Applications of the Conjecture*

The original correspondence is between a $\mathcal{N} = 4$ SYM theory in its conformal phase and string theory on $\mathrm{AdS}_5 \times \mathrm{S}^5$. The power of the correspondence is further evidenced by the fact that the conjecture may be adapted to situations without conformal invariance and with less or no supersymmetry on the SYM side. The $\mathrm{AdS}_5 \times \mathrm{S}^5$ space-time is then replaced by other manifold or orbifold solutions to Type IIB theory, whose study is usually more involved than was the case for $\mathrm{AdS}_5 \times \mathrm{S}^5$ but still reveals useful information on SYM theory.

2. Supersymmetry and Gauge Theories

We begin by reviewing the particle and field contents and invariant Lagrangians in 4 dimensions, in preparation for a fuller discussion of $\mathcal{N} = 4$

super-Yang-Mills (SYM) theory in the next section. Standard references include Refs. 11, 12, 13; our conventions are those of Ref. 11.

2.1. *Supersymmetry Algebra in 3+1 Dimensions*

Poincaré symmetry of flat space-time $\mathbf{R}^4$ with metric $\eta_{\mu\nu} = \text{diag}(-+++)$, $\mu, \nu = 0,1,2,3$, is generated by the translations $\mathbf{R}^4$ and Lorentz transformations $SO(1,3)$, with generators P_μ and $L_{\mu\nu}$ respectively. The complexified Lorentz algebra is isomorphic to the complexified algebra of $SU(2)\times SU(2)$, and its finite-dimensional representations are usually labeled by two positive (or zero) half integers (s_+, s_-), $s_\pm \in \mathbf{Z}/2$. Scalar, 4-vector, and rank 2 symmetric tensors transform under $(0,0)$, $(\frac{1}{2},\frac{1}{2})$ and $(1,1)$ respectively, while left and right chirality fermions and self-dual and anti–self-dual rank 2 tensors transform under $(\frac{1}{2},0)$ and $(0,\frac{1}{2})$ and $(1,0)$ and $(0,1)$ respectively.

Supersymmetry (susy) enlarges the Poincaré algebra by including spinor *supercharges*,

$$a = 1,\cdots,\mathcal{N} \qquad \begin{cases} Q^a_\alpha & \alpha = 1,2 \quad \text{left Weyl spinor} \\ \\ \bar{Q}_{\dot{\alpha}a} = (Q^a_\alpha)^\dagger & \qquad\quad \text{right Weyl spinor} \end{cases} \tag{1}$$

Here, $\mathcal{N}$ is the number of independent supersymmetries of the algebra. Two-component spinor notation, $\alpha = 1,2$, is related to 4-component Dirac spinor notation by

$$\gamma^\mu = \begin{pmatrix} 0 & \sigma^\mu \\ \bar{\sigma}^\mu & 0 \end{pmatrix} \qquad\qquad Q^a = \begin{pmatrix} Q^a_\alpha \\ \bar{Q}^{\dot{\alpha}}_a \end{pmatrix} \tag{2}$$

The supercharges transform as Weyl spinors of $SO(1,3)$ and commute with translations. The remaining susy structure relations are

$$\{Q^a_\alpha, \bar{Q}_{\dot{\beta}b}\} = 2\sigma^\mu_{\alpha\dot{\beta}} P_\mu \delta^a_b \qquad\qquad \{Q^a_\alpha, Q^b_\beta\} = 2\epsilon_{\alpha\beta} Z^{ab} \tag{3}$$

By construction, the generators Z^{ab} are anti-symmetric in the indices I and J, and commute with all generators of the supersymmetry algebra. For the last reason, the Z^{ab} are usually referred to as *central charges*, and we have

$$Z^{ab} = -Z^{ba} \qquad\qquad [Z^{ab}, \text{anything}] = 0 \tag{4}$$

Note that for $\mathcal{N} = 1$, the anti-symmetry of Z implies that $Z = 0$.

The supersymmetry algebra is invariant under a global phase rotation of all supercharges Q^a_α, forming a group $U(1)_R$. In addition, when $\mathcal{N} > 1$, the different supercharges may be rotated into one another under the unitary group $SU(\mathcal{N})_R$. These (automorphism) symmetries of the supersymmetry algebra are called *R-symmetries.* In quantum field theories, part or all of these R-symmetries may be broken by anomaly effects.

2.2. *Massless Particle Representations*

To study massless representations, we choose a Lorentz frame in which the momentum takes the form $P^\mu = (E, 0, 0, E)$, $E > 0$. The susy algebra relation (3) then reduces to

$$\{Q^a_\alpha, (Q^b_\beta)^\dagger\} = 2(\sigma^\mu P_\mu)_{\alpha\dot\beta}\delta^a_b = \begin{pmatrix} 4E & 0 \\ 0 & 0 \end{pmatrix}_{\alpha\dot\beta} \delta^a_b \tag{5}$$

We consider only unitary particle representations, in which the operators Q^a_α act in a positive definite Hilbert space. The relation for $\alpha = \dot\beta = 2$ and $a = b$ implies

$$\{Q^a_2, (Q^a_2)^\dagger\} = 0 \qquad \Longrightarrow \qquad Q^a_2 = 0, \quad Z^{ab} = 0 \tag{6}$$

The relation $Q^a_2 = 0$ follows because the left hand side of (6) implies that the norm of $Q^a_2|\psi\rangle$ vanishes for any state $|\psi\rangle$ in the Hilbert space. The relation $Z^{ab} = 0$ then follows from (3) for $\alpha = 2$ and $\dot\beta = 1$. The remaining supercharge operators are

- Q^a_1 which lowers helicity by 1/2;
- $\bar Q^a_{\dot 1} = (Q^a_1)^\dagger$ which raises helicity by 1/2.

Together, Q^a_1 and $(Q^a_1)^\dagger$, with $a = 1, \cdots, \mathcal{N}$ form a representation of dimension $2^\mathcal{N}$ of the Clifford algebra associated with the Lie algebra $SO(2\mathcal{N})$. All the states in the representation may be obtained by starting from the highest helicity state $|h\rangle$ and applying products of Q^a_1 operators for all possible values of a.

We shall only be interested in CPT invariant theories, such as quantum field theories and string theories, for which the particle spectrum must be symmetric under a sign change in helicity. If the particle spectrum obtained as a Clifford representation in the above fashion is not already CPT self-conjugate, then we shall take instead the direct sum with its CPT conjugate. For helicity ≤ 1, the spectra are listed in table 1. The $\mathcal{N} = 3$ and $\mathcal{N} = 4$ particle spectra then coincide, and their quantum field theories are identical.

2.3. *Massive Particle Representations*

For massive particle representations, we choose the rest frame with $P^\mu = (M,0,0,0)$, so that the first set of susy algebra structure relations takes the form

$$\{Q^a_\alpha, (Q^b_\beta)^\dagger\} = 2M\delta^\beta_\alpha \delta^a_b \tag{7}$$

To deal with the full susy algebra, it is convenient to make use of the $SU(\mathcal{N})_R$ symmetry to diagonalize in blocks of 2×2 the anti-symmetric matrix $Z^{ab} = -Z^{ba}$. To do so, we split the label a into two labels : $a = (\hat{a}, \bar{a})$ where $\hat{a} = 1, 2$ and $\bar{a} = 1, \cdots, r$, where $\mathcal{N} = 2r$ for $\mathcal{N}$ even (and we append a further single label when $\mathcal{N}$ is odd). We then have

$$Z = \mathrm{diag}(\epsilon Z_1, \cdots, \epsilon Z_r, \#) \qquad \epsilon^{12} = -\epsilon^{21} = 1 \tag{8}$$

where $\#$ equals 0 for $\mathcal{N} = 2r+1$ and $\#$ is absent for $\mathcal{N} = 2r$. The $Z_{\bar{a}}$, $\bar{a} = 1, \cdots, r$ are real *central charges*. In terms of linear combinations $\mathcal{Q}^{\bar{a}}_{\alpha\pm} \equiv \frac{1}{2}\big(Q^{1\bar{a}}_\alpha \pm \sigma^0_{\alpha\dot{\beta}}(Q^{2\bar{a}}_\beta)^\dagger\big)$, the only non-vanishing susy structure relation left is (the $\pm$ signs below are correlated)

$$\{\mathcal{Q}^{\bar{a}}_{\alpha\pm}, (\mathcal{Q}^{\bar{b}}_{\beta\pm})^\dagger\} = \delta^{\bar{a}}_{\bar{b}}\delta^\beta_\alpha(M \pm Z_{\bar{a}}) \tag{9}$$

In any *unitary particle representation*, the operator on the left hand side of (9) must be positive, and thus we obtain the famous *BPS bound* (for Bogomolnyi-Prasad-Sommerfield, Ref. 14) giving a lower bound on the mass in terms of the central charges,

$$M \geq |Z_{\bar{a}}| \qquad \bar{a} = 1, \cdots, r = [\mathcal{N}/2] \tag{10}$$

Whenever one of the values $|Z_{\bar{a}}|$ equals M, the BPS bound is (partially) saturated and either the supercharge $\mathcal{Q}^{\bar{a}}_{\alpha+}$ or $\mathcal{Q}^{\bar{a}}_{\alpha-}$ must vanish. The supersymmetry representation then suffers *multiplet shortening*, and is usually referred to as BPS. More precisely, if we have $M = |Z_{\bar{a}}|$ for $\bar{a} = 1, \cdots, r_o$, and $M > |Z_{\bar{a}}|$ for all other values of $\bar{a}$, the susy algebra is effectively a

Table 1. Numbers of Massless States as a function of $\mathcal{N}$ and helicity

Helicity ≤ 1	$\mathcal{N}=1$ gauge	$\mathcal{N}=1$ chiral	$\mathcal{N}=2$ gauge	$\mathcal{N}=2$ hyper	$\mathcal{N}=3$ gauge	$\mathcal{N}=4$ gauge
1	1	0	1	0	1	1
1/2	1	1	2	2	3+1	4
0	0	1+1	1+1	4	3+3	6
−1/2	1	1	2	2	1+3	4
−1	1	0	1	0	1	1
Total #	2×2	2×2	2×4	8	2×8	16

Clifford algebra associated with $SO(4\mathcal{N} - 4r_o)$, the corresponding representation is said to be $1/2^{r_o}$ BPS, and has dimension $2^{2\mathcal{N}-2r_o}$.

2.4. *Field Contents and Lagrangians*

The analysis of the preceding two subsections has revealed that the supersymmetry particle representations for $1 \leq \mathcal{N} \leq 4$, with spin less or equal to 1, simply consist of customary spin 1 vector particles, spin 1/2 fermions and spin 0 scalars. Correspondingly, *the fields in supersymmetric theories with spin less or equal to 1 are customary spin 1 gauge fields, spin 1/2 Weyl fermion fields and spin 0 scalar fields, but these fields are restricted to enter in multiplets of the relevant supersymmetry algebras.*

Let $\mathcal{G}$ denote the gauge algebra, associated with a compact Lie group G. For any $1 \leq \mathcal{N} \leq 4$, we have a gauge multiplet, which transforms under the adjoint representation of $\mathcal{G}$. For $\mathcal{N} = 4$, this is the only possible multiplet. For $\mathcal{N} = 1$ and $\mathcal{N} = 2$, we also have *matter multiplets* : for $\mathcal{N} = 1$, this is the *chiral multiplet*, and for $\mathcal{N} = 2$ this is the *hypermultiplet*, both of which may transform under an arbitrary unitary, representation $\mathcal{R}$ of $\mathcal{G}$. Component fields consist of the customary gauge field A_μ, left Weyl fermions ψ_α and λ_α and scalar fields ϕ, H and X.

- $\mathcal{N} = 1$ *Gauge Multiplet* $(A_\mu\ \lambda_\alpha)$, where λ_α is the gaugino Majorana fermion;
- $\mathcal{N} = 1$ *Chiral Multiplet* $(\psi_\alpha\ \phi)$, where ψ_α is a left Weyl fermion and ϕ a complex scalar, in the representation $\mathcal{R}$ of $\mathcal{G}$.
- $\mathcal{N} = 2$ *Gauge Multiplet* $(A_\mu\ \lambda_{\alpha\pm}\ \phi)$, where $\lambda_{\alpha\pm}$ are left Weyl fermions, and ϕ is the complex *gauge scalar.* Under $SU(2)_R$ symmetry, A_μ and ϕ are singlets, while λ_+ and λ_- transform as a doublet.
- $\mathcal{N} = 2$ *Hypermultiplet* $(\psi_{\alpha\pm}\ H_\pm)$, where $\psi_{\alpha\pm}$ are left Weyl fermions and $H_\pm$ are two complex scalars, transforming under the

Table 2. Numbers of Massive States as a function of $\mathcal{N}$ and spin

Spin ≤ 1	$\mathcal{N}=1$ gauge	$\mathcal{N}=1$ chiral	$\mathcal{N}=2$ gauge	$\mathcal{N}=2$ BPS gauge	$\mathcal{N}=2$ BPS hyper	$\mathcal{N}=4$ BPS gauge
1	1	0	1	1	0	1
1/2	2	1	4	2	2	4
0	1	2	5	1	4	5
Total #	8	4	16	8	8	16

representation $\mathcal{R}$ of $\mathcal{G}$. Under $SU(2)_R$ symmetry, $\psi_\pm$ are singlets, while H_+ and H_- transform as a doublet.

- $\mathcal{N}=4$ *Gauge Multiplet* $(A_\mu\ \lambda^a_\alpha\ X^i)$, where λ^a_α, $a=1,\cdots,4$ are left Weyl fermions and X^i, $i=1,\cdots,6$ are real scalars. Under $SU(4)_R$ symmetry, A_μ is a singlet, λ^a_α is a $\mathbf{4}$ and the scalars X^i are a rank 2 anti-symmetric $\mathbf{6}$.

Lagrangians invariant under supersymmetry are customary Lagrangians of gauge, spin 1/2 fermion and scalar fields, (arranged in multiplets of the supersymmetry algebra) with certain special relations amongst the coupling constants and masses. We shall restrict attention to local Lagrangians in which each term has a total of no more than two derivatives on all boson fields and no more than one derivative on all fermion fields. All renormalizable Lagrangians are of this form, but all low energy effective Lagrangians are also of this type.

The case of the $\mathcal{N}=1$ gauge multiplet $(A_\mu\ \lambda_\alpha)$ by itself is particularly simple. We proceed by writing down all possible gauge invariant polynomial terms of dimension 4 using minimal gauge coupling,

$$\mathcal{L} = -\frac{1}{2g^2}\mathrm{tr}F_{\mu\nu}F^{\mu\nu} + \frac{\theta_I}{8\pi^2}\mathrm{tr}F_{\mu\nu}\tilde{F}^{\mu\nu} - \frac{i}{2}\mathrm{tr}\bar{\lambda}\bar{\sigma}^\mu D_\mu\lambda \tag{11}$$

where g is the gauge coupling, θ_I is the instanton angle, the field strength is $F_{\mu\nu}=\partial_\mu A_\nu-\partial_\nu A_\mu+i[A_\mu,A_\nu]$, $\tilde{F}_{\mu\nu}=\frac{1}{2}\epsilon_{\mu\nu\rho\sigma}F^{\rho\sigma}$ is the Poincaré dual of F, and $D_\mu=\partial_\mu\lambda+i[A_\mu,\lambda]$. Remarkably, $\mathcal{L}$ is automatically invariant under the $\mathcal{N}=1$ supersymmetry transformations

$$\begin{aligned}\delta_\xi A_\mu &= i\bar{\xi}\bar{\sigma}_\mu\lambda - i\bar{\lambda}\bar{\sigma}_\mu\xi\\ \delta_\xi\lambda &= \sigma^{\mu\nu}F_{\mu\nu}\xi\end{aligned} \tag{12}$$

where ξ is a spin 1/2 valued infinitesimal supersymmetry parameter. Note that the addition in (11) of a Majorana mass term $m\lambda\lambda$ would spoil supersymmetry.

As soon as scalar fields are to be included, such as is the case for any other multiplet, it is no longer so easy to guess supersymmetry invariant Lagrangians and one is led to the use of superfields. *Superfields* assemble all component fields of a given supermultiplet (together with auxiliary fields) into a supersymmetry multiplet field on which supersymmetry transformations act linearly. Superfield methods provide a powerful tool for producing supersymmetric field equations for any degree of supersymmetry. For $\mathcal{N}=1$ there is a standard off-shell superfield formulation as well (see Refs. 11, 12, 13 for standard treatments). Considerably more involved off-shell superfield formulations are also available for $\mathcal{N} = 2$ in terms of harmonic and analytic superspace,[15] see also the review of Ref. 16. For $\mathcal{N} = 4$ supersymmetry, no off-shell formulation is known at present; one is thus forced to work either in components or in the $\mathcal{N} = 1$ or $\mathcal{N} = 2$ superfield formulations. A survey of theories with extended supersymmetry may be found in Ref. 23.

2.5. *The $\mathcal{N}$=1 Superfield Formulation*

The construction of field multiplets containing all fields that transform linearly into one another under supersymmetry requires the introduction of anti-commuting spin 1/2 coordinates. For $\mathcal{N} = 1$ supersymmetry, we introduce a (constant) left Weyl spinor coordinate θ_α and its complex conjugate $\bar\theta^{\dot\alpha} = (\theta_\alpha)^\dagger$, satisfying $[x^\mu, \theta_\alpha] = \{\theta_\alpha, \theta_\beta\} = \{\theta_\alpha, \bar\theta^{\dot\beta}\} = \{\bar\theta^{\dot\alpha}, \bar\theta^{\dot\beta}\} = 0$. Superderivatives are defined by

$$D_\alpha \equiv \frac{\partial}{\partial\theta^\alpha} + i\sigma^\mu_{\alpha\dot\alpha}\bar\theta^{\dot\alpha}\partial_\mu \qquad\qquad \bar D_{\dot\alpha} \equiv -\frac{\partial}{\partial\bar\theta^{\dot\alpha}} - i\theta^\alpha\sigma^\mu_{\alpha\dot\alpha}\partial_\mu \tag{13}$$

where differentiation and integration of θ coordinates are defined by

$$\frac{\partial}{\partial\theta^\alpha}(1, \theta^\beta, \bar\theta^{\dot\beta}) \equiv \int d\theta^\alpha (1, \theta^\beta, \bar\theta^{\dot\beta}) \equiv (0, \delta_\alpha{}^\beta, 0) \tag{14}$$

For general notations and conventions for spinors and their contractions, see Ref. 11.

A *general superfield.* is defined as a general function of the superspace coordinates $x^\mu, \theta_\alpha, \bar\theta^{\dot\alpha}$ Since the square of each θ^α or of each $\bar\theta^{\dot\alpha}$ vanishes, superfields admit finite Taylor expansions in powers of θ and $\bar\theta$. A general superfield $S(x, \theta, \bar\theta)$ yields the following *component expansion*

$$\begin{aligned} S(x,\theta,\bar\theta) = \phi(x) &+ \theta\psi(x) + \bar\theta\bar\chi(x) + \bar\theta\bar\sigma^\mu\theta A_\mu(x) + \theta\theta f(x) + \bar\theta\bar\theta g^*(x) \\ &+ i\theta\theta\bar\theta\bar\lambda(x) - i\bar\theta\bar\theta\theta\rho(x) + \frac{1}{2}\theta\theta\bar\theta\bar\theta D(x) \end{aligned} \tag{15}$$

A bosonic superfield obeys $[S, \theta^\alpha] = [S, \bar\theta_{\dot\alpha}] = 0$, while a fermionic superfield obeys $\{S, \theta^\alpha\} = \{S, \bar\theta_{\dot\alpha}\} = 0$. If S is bosonic (resp. fermionic), the component fields ϕ, A_μ, f, g and D are bosonic (resp. fermionic) as well, while the fields ψ, χ, λ and ρ are fermionic (resp. bosonic). The superfields belong to a $\mathbf{Z}_2$ graded algebra of functions on superspace, with the even grading for bosonic odd grading for fermionic fields. We shall denote the grading by $g(S)$, or sometimes just S. Superderivatives on superfields satisfy the following graded differentiation rule

$$D_\alpha(S_1S_2) = (D_\alpha S_1)S_2 + (-)^{g(S_1)g(S_2)}S_1(D_\alpha S_2) \tag{16}$$

where $g(S_i)$ is the grading of the field S_i.

On superfields, supersymmetry transformations are realized in a linear way via super-differential operators. The infinitesimal supersymmetry parameter is still a constant left Weyl spinor ξ, as in (12) and we have

$$\delta_\xi S = (\xi Q + \bar\xi \bar Q)S \tag{17}$$

with the supercharges defined by

$$Q_\alpha = \frac{\partial}{\partial\theta^\alpha} - i\sigma^\mu_{\alpha\dot\alpha}\bar\theta^{\dot\alpha}\partial_\mu \qquad\qquad \bar Q_{\dot\alpha} = -\frac{\partial}{\partial\bar\theta^{\dot\alpha}} + i\theta^\alpha\sigma^\mu_{\alpha\dot\alpha}\partial_\mu \tag{18}$$

The super-differential operators D_α and Q_α differ only by a sign change. They generate left and right actions of supersymmetry respectively. Their relevant structure relations are

$$\{Q_\alpha, \bar Q_{\dot\beta}\} = 2\sigma^\mu_{\alpha\dot\beta}P_\mu \qquad\qquad \{D_\alpha, \bar D_{\dot\beta}\} = -2\sigma^\mu_{\alpha\dot\beta}P_\mu \tag{19}$$

where $P_\mu = i\partial_\mu$. Since left and right actions mutually commute, all components of D anti-commute with all components of Q : $\{Q_\alpha, D_\beta\} = \{Q_\alpha, \bar D^{\dot\beta}\} = 0$. Furthermore, the product of any three D's or any three Q's vanishes, $D_\alpha D_\beta D_\gamma = Q_\alpha Q_\beta Q_\gamma = 0$. The general superfield is reducible; the irreducible components are as follows.

(a) The *Chiral Superfield* Φ is obtained by imposing the condition

$$\bar D_{\dot\alpha}\Phi = 0 \tag{20}$$

This condition is invariant under the supersymmetry transformations of (17) since D and Q anti-commute. Equation (20) may be solved in terms of the composite coordinates $x^\mu_\pm = x^\mu \pm i\theta\sigma^\mu\bar\theta$ which satisfy $\bar D_{\dot\alpha}x^\mu_+ = 0$ and $D_\alpha x^\mu_- = 0$. We have

$$\Phi(x,\theta,\bar\theta) = \phi(x_+) + \sqrt{2}\theta\psi(x_+) + \theta\theta F(x_+) \tag{21}$$

The component fields ϕ and ψ are the scalar and left Weyl spinor fields of the chiral multiplet respectively, as discussed previously. The field equation for F is a non-dynamical or *auxiliary field* of the chiral multiplet.

(b) The *Vector Superfield* is obtained by imposing on a general superfield of the type (15) the condition

$$V = V^\dagger \tag{22}$$

thereby setting $\chi = \psi$, $g = f$ and $\rho = \lambda$ and requiring ϕ, A_μ and D to be real.

(c) The *Gauge Superfield* is a special case of a vector superfield, where V takes values in the gauge algebra $\mathcal{G}$. The reality condition $V = V^\dagger$ is preserved by the gauge transformation

$$e^V \longrightarrow e^{V'} = e^{-i\Lambda^\dagger} e^V e^{i\Lambda}\,. \tag{23}$$

where Λ is a chiral superfield taking values also in $\mathcal{G}$. Under the above gauge transformation law, the component fields ϕ, $\psi = \chi$, and $f = g$ of a general real superfield may be gauged away in an algebraic way. The gauge in which this is achieved is the *Wess-Zumino gauge*, where the gauge superfield is given by

$$V(x,\theta,\bar\theta) = \bar\theta\bar\sigma^\mu\theta A_\mu(x) + i\theta\theta\bar\theta\bar\lambda(x) - i\bar\theta\bar\theta\theta\lambda(x) + \frac{1}{2}\theta\theta\bar\theta\bar\theta D(x) \tag{24}$$

The component fields A_μ and λ are the gauge and gaugino fields of the gauge multiplet respectively, as discussed previously. The field D has not appeared previously and is an *auxiliary field*, just as F was an auxiliary field for the chiral multiplet.

2.6. *General $\mathcal{N}$=1 Susy Lagrangians via Superfields*

Working out the supersymmetry transformation (17) on chiral and vector superfields in terms of components, we see that the only contribution to the auxiliary fields is from the $\theta\partial$ term of Q and thus takes the form of a total derivative. However, because the form (24) was restricted to Wess-Zumino gauge, F and D transform by a total derivative only if F and D are themselves gauge singlets, in which case we have

$$\begin{aligned} \delta_\xi F &= i\sqrt{2}\partial_\mu(\bar\xi\bar\sigma^\mu\psi) \\ \delta_\xi D &= \partial_\mu(i\bar\xi\bar\sigma^\mu\lambda - i\bar\lambda\bar\sigma^\mu\xi) \end{aligned} \tag{25}$$

These transformation properties guarantee that the F and D auxiliary fields yield supersymmetric invariant Lagrangian terms,

$$\begin{aligned} &\mathrm{F-terms} \qquad \mathcal{L}_F = F = \int d^2\theta\ \Phi \\ &\mathrm{D-terms} \qquad \mathcal{L}_D = \frac{1}{2}D = \int d^4\theta\ V \end{aligned} \tag{26}$$

The F and D terms used to construct invariants need not be elementary fields, and may be gauge invariant composites of elementary fields instead. Allowing for this possibility, we may now derive the most general possible $\mathcal{N} = 1$ invariant Lagrangian in terms of superfields. To do so, we need the following ingredients $\mathcal{L}_U$, $\mathcal{L}_G$ and $\mathcal{L}_K$. Putting together contributions from these terms, we have the most general $\mathcal{N} = 1$ supersymmetric Lagrangian with the restrictions of above.

(1) Any complex analytic function U depending only on left chiral superfields Φ^i (but not on their complex conjugates) is itself a left chiral superfield, since $\bar{D}_{\dot{\alpha}}\Phi^i = 0$ implies that $\bar{D}_{\dot{\alpha}}U(\Phi^i) = 0$. Thus, for any complex analytic function U, called the *superpotential*, we may construct an invariant contribution to the Lagrangian by forming an F-term

$$\begin{aligned}\mathcal{L}_U &= \int d^2\theta\, U(\Phi^i) + \int d^2\bar{\theta}\overline{U(\Phi^i)} \\ &= \sum_i F^i \frac{\partial U}{\partial \phi^i} - \frac{1}{2}\sum_{i,j} \psi^i \psi^j \frac{\partial^2 U}{\partial \phi^i \partial \phi^j} + \text{cc} \end{aligned} \tag{27}$$

(2) The gauge field strength is a fermionic left chiral spinor superfield W_α, which is constructed out of the gauge superfield V by

$$4W_\alpha = -\bar{D}\bar{D}\big(e^{-V}D_\alpha e^{+V}\big) \tag{28}$$

The gauge field strength may be used as a chiral superfield along with elementary (scalar) chiral superfields to build up $\mathcal{N} = 1$ supersymmetric Lagrangians via F-terms. In view of our restriction to Lagrangians with no more than two derivatives on Bose fields, W can enter at most quadratically. Thus, the most general gauge kinetic and self-interaction term is from the F-term of the gauge field strength W_α and the elementary (scalar) chiral superfields Φ^i as follows,

$$\mathcal{L}_G = \int d^2\theta\, \tau_{cc'}(\Phi^i) W^c W^{c'} + \text{complex conjugate} \tag{29}$$

Here, c and c' stand for the gauge index running over the adjoint representation of $\mathcal{G}$, and are contracted in a gauge invariant way. The functions $\tau_{cc'}(\Phi^i)$ are complex analytic.

(3) The left and right chiral superfields Φ^i and $(\Phi^i)^\dagger$, as well as the gauge superfield V, may be combined into a gauge invariant vector superfield $K(e^V\Phi^i, (\Phi^i)^\dagger)$, provided the gauge algebra is realized linearly on the fields Φ^i. The function K is called the *Kähler potential*. Assuming that the gauge transformations Λ act on V by (23), the chiral superfields Φ transform as $\Phi \longrightarrow \Phi' = e^{-i\Lambda}\Phi$, so that $e^V\Phi$ transforms as Φ. An invariant

Lagrangian may be constructed via a D-term,

$$\mathcal{L}_K = \int d^4\theta K(e^V\Phi^i, (\Phi^i)^\dagger) \tag{30}$$

Upon expanding $\mathcal{L}_K$ in components, one sees immediately that the leading terms already generates an action with two derivatives on boson fields. As a result, K must be a function only of the superfields Φ^i and $(\Phi^i)^\dagger$ and V, but not of their derivatives.

2.7. $\mathcal{N} = 1$ *Non-Renormalization Theorems*

Non-renormalization theorems provide very strong results on the structure of the effective action at low energy as a function of the bare action. Until recently, their validity was restricted to perturbation theory and the proof of the theorems was based on supergraph methods.[17] Now, however, non-renormalization theorems have been extended to the non-perturbative regime, including the effects of instantons.[18] The assumptions underlying the theorems are that (1) a supersymmetric renormalization is carried out, (2) the effective action is constructed by Wilsonian renormalization (see Ref. 19 for a review). The last assumption allows one to circumvent any possible singularities resulting from the integration over massless states.

The non-renormalization theorems state that the superpotential $\mathcal{L}_U$ is *unrenormalized*, or more precisely that it receives no quantum corrections, infinite or finite. Furthermore, the gauge field term $\mathcal{L}_G$ is renormalized only through the gauge coupling $\tau_{cc'}$, such that its complex analytic dependence on the chiral superfields is preserved. In perturbation theory, $\tau_{cc'}$ receives quantum contributions only through 1-loop graphs, essentially because the $U(1)_R$ axial anomaly is a 1-loop effect in view of the Adler-Bardeen theorem. Non-perturbatively, instanton corrections also enter, but in a very restricted way. The special renormalization properties of these two F-terms are closely related to their holomorphicity.[18] The Kähler potential term $\mathcal{L}_K$ on the other hand does receive renormalizations both at the perturbative and non-perturbative levels.

3. $\mathcal{N} = 4$ Super Yang-Mills

The Lagrangian for the $\mathcal{N} = 4$ super-Yang Mills theory is unique and given by Ref. 20 as

$$\mathcal{L} = \mathrm{tr}\Big\{ -\frac{1}{2g^2}F_{\mu\nu}F^{\mu\nu} + \frac{\theta_I}{8\pi^2}F_{\mu\nu}\tilde{F}^{\mu\nu} - \sum_a i\bar{\lambda}^a\bar{\sigma}^\mu D_\mu\lambda_a - \sum_i D_\mu X^i D^\mu X^i$$

$$+\sum_{a,b,i} gC_i^{ab}\lambda_a[X^i,\lambda_b]+\sum_{a,b,i} g\bar{C}_{iab}\bar{\lambda}^a[X^i,\bar{\lambda}^b]+\frac{g^2}{2}\sum_{i,j}[X^i,X^j]^2\Big\} \tag{31}$$

The constants C_i^{ab} and C_{iab} are related to the Clifford Dirac matrices for $SO(6)_R \sim SU(4)_R$. This is evident when considering $\mathcal{N}=4$ SYM in $D=4$ as the dimensional reduction on T^6 of $D=10$ super-Yang-Mills theory (see problem set 4.1). By construction, the Lagrangian is invariant under $\mathcal{N}=4$ Poincaré supersymmetry, whose transformation laws are given by

$$\begin{aligned}
\delta X^i &= [Q^a_\alpha, X^i] = C^{iab}\lambda_{\alpha b}\\
\delta\lambda_b &= \{Q^a_\alpha,\lambda_{\beta b}\} = F^+_{\mu\nu}(\sigma^{\mu\nu})^\alpha{}_\beta\delta^a_b + [X^i,X^j]\epsilon_{\alpha\beta}(C_{ij})^a{}_b\\
\delta\bar{\lambda}^b_{\dot\beta} &= \{Q^a_\alpha,\bar{\lambda}^b_{\dot\beta}\} = C_i^{ab}\bar{\sigma}^\mu_{\alpha\dot\beta}D_\mu X^i\\
\delta A_\mu &= [Q^a_\alpha, A_\mu] = (\sigma_\mu)_\alpha{}^{\dot\beta}\bar{\lambda}^a_{\dot\beta}
\end{aligned} \tag{32}$$

The constants $(C_{ij})^a{}_b$ are related to bilinears in Clifford Dirac matrices of $SO(6)_R$.

Classically, $\mathcal{L}$ is *scale invariant*. This may be seen by assigning the standard mass-dimensions to the fields and couplings

$$[A_\mu]=[X^i]=1 \qquad [\lambda_a]=\frac{3}{2} \qquad [g]=[\theta_I]=0 \tag{33}$$

All terms in the Lagrangian are of dimension 4, from which scale invariance follows. Actually, in a relativistic field theory, scale invariance and Poincaré invariance combine into a larger *conformal symmetry*, forming the group $SO(2,4)\sim SU(2,2)$. Furthermore, the combination of $\mathcal{N}=4$ Poincaré supersymmetry and conformal invariance produces an even larger *superconformal symmetry* given by the supergroup $SU(2,2|4)$.

Remarkably, upon perturbative quantization, $\mathcal{N}=4$ SYM theory exhibits no ultraviolet divergences in the correlation functions of its canonical fields. Instanton corrections also lead to finite contributions and is believed that the theory is UV finite. As a result, the renormalization group β-function of the theory vanishes identically (since no dependence on any renormalization scale is introduced during the renormalization process). The theory is exactly scale invariant at the quantum level, and the superconformal group $SU(2,2|4)$ is a fully quantum mechanical symmetry.

The *Montonen-Olive or S-duality conjecture* in addition posits a discrete global symmetry of the theory.[21] To state this invariance, it is standard to combine the real coupling g and the real instanton angle θ_I into a single complex coupling

$$\tau \equiv \frac{\theta_I}{2\pi} + \frac{4\pi i}{g^2} \tag{34}$$

The quantum theory is invariant under $\theta_I \to \theta_I + 2\pi$, or $\tau \to \tau + 1$. The *Montonen-Olive conjecture* states that the quantum theory is also invariant under the $\tau \to -1/\tau$. The combination of both symmetries yields the S-duality group SL(2,**Z**), generated by

$$\tau \to \frac{a\tau + b}{c\tau + d} \qquad\qquad ad - bc = 1, \quad a, b, c, d \in \mathbf{Z} \tag{35}$$

Note that when $\theta_I = 0$, the S-duality transformation amounts to $g \to 1/g$, thereby exchanging strong and weak coupling.

3.1. *Dynamical Phases*

To analyze the dynamical behavior of $\mathcal{N} = 4$ theory, we look at the potential energy term,

$$-g^2 \sum_{i,j} \int \mathrm{tr}[X^i, X^j]^2$$

In view of the positive definite behavior of the Cartan-Killing form on the compact gauge algebra $\mathcal{G}$, each term in the sum is positive or zero. When the full potential is zero, a minimum is thus automatically attained corresponding to a $\mathcal{N} = 4$ supersymmetric ground state. In turn, any $\mathcal{N} = 4$ supersymmetric ground state is of this form,

$$[X^i, X^j] = 0, \qquad i, j = 1, \cdots, 6 \tag{36}$$

There are two classes of solutions to this equation,

- The *superconformal phase*, for which $\langle X^i \rangle = 0$ for all $i = 1, \cdots, 6$. The gauge algebra $\mathcal{G}$ is unbroken. The superconformal symmetry $SU(2,2|4)$ is also unbroken. The physical states and operators are gauge invariant (i.e. $\mathcal{G}$-singlets) and transform under unitary representations of $SU(2,2|4)$.
- The *spontaneously broken or Coulomb phase*, where $\langle X^i \rangle \neq 0$ for at least one i. The detailed dynamics will depend upon the degree of residual symmetry. Generically, $\mathcal{G} \to U(1)^r$ where $r = \mathrm{rank}\ \mathcal{G}$, in which case the low energy behavior is that of r copies of $\mathcal{N} = 4$ $U(1)$ theory. Superconformal symmetry is spontaneously broken since the non-zero vacuum expectation value $\langle X^i \rangle$ sets a scale.

3.2. *Isometries and Conformal Transformations*

In the first part of these lectures, we shall consider the SYM theory in the conformal phase and therefore make heavy use of superconformal symmetry. In the present subsection, we begin by reviewing conformal symmetry first. Let M be a Riemannian (or pseudo-Riemannian) manifold of dimension D with metric $G_{\mu\nu}$, $\mu, \nu = 0, 1, \cdots, D-1$. We shall now review the notions of diffeomorphisms, isometries and conformal transformations.

- A *diffeomorphism* of M is a differentiable map of local coordinates x^μ, $\mu = 1, \cdots, D$, of M given either globally by $x^\mu \to x'^\mu(x)$ or infinitesimally by a vector field $v^\mu(x)$ so that $\delta_v x^\mu = -v^\mu(x)$. Under a general diffeomorphism, the metric on M transforms as

$$\begin{aligned} G'_{\mu\nu}(x')dx'^\mu dx'^\nu &= G_{\mu\nu}dx^\mu dx^\nu \\ \delta_v G_{\mu\nu} &= \nabla_\mu v_\nu + \nabla_\nu v_\mu \qquad\qquad \nabla_\mu v_\nu \equiv \partial_\mu v_\nu - \Gamma^\rho_{\mu\nu} v_\rho \end{aligned} \tag{37}$$

- An *isometry* is a diffeomorphism under which the metric is invariant,

$$G'_{\mu\nu}(x) = G_{\mu\nu}(x) \qquad \text{or} \qquad \delta_v G_{\mu\nu} = \nabla_\mu v_\nu + \nabla_\nu v_\mu = 0 \tag{38}$$

- A *conformal transformation* is a diffeomorphism that preserves the metric up to an overall (in general x-dependent) scale factor, thereby preserving all angles,

$$G'_{\mu\nu}(x) = w(x)G_{\mu\nu}(x) \quad \text{or} \quad \delta_v G_{\mu\nu} = \nabla_\mu v_\nu + \nabla_\nu v_\mu = \tfrac{2}{D} G_{\mu\nu} \nabla_\rho v^\rho \tag{39}$$

The case of $M = \mathbf{R}^D$, $D \geq 3$, flat Minkowski space-time with flat metric $\eta_{\mu\nu} = \mathrm{diag}(- + \cdots +)$ is an illuminating example. (When $D = 2$, the conformal algebra is isomorphic to the infinite-dimensional Virasoro algebra.) Since now $\nabla_\mu = \partial_\mu$, the equations for isometries reduce to $\partial_\mu v_\nu + \partial_\nu v_\mu = 0$, while those for conformal transformations become $\partial_\mu v_\nu + \partial_\nu v_\mu - 2/D \eta_{\mu\nu} \partial_\rho v^\rho = 0$. The solutions are

$$\begin{array}{lll} \text{isometries} & (1) & v^\mu \text{ constant} \;:\; \text{translations} \\ & (2) & v_\mu = \omega_{\mu\nu} x^\nu \;:\; \text{Lorentz} \\ \text{conformal} & (3) & v^\mu = \lambda x^\mu \;:\; \text{dilations} \\ & (4) & v_\mu = 2c_\rho x^\rho x_\mu - x_\rho x^\rho c_\mu \;:\; \text{special conformal} \end{array} \tag{40}$$

In a local field theory, continuous symmetries produce conserved currents, according to Noether's Theorem. Currents associated with isometries and conformal transformations may be expressed in terms of the stress tensor

$T_{\mu\nu}$. This is because the stress tensor for any local field theory encodes the response of the theory to a change in metric; as a result, it is automatically symmetric $T^{\mu\nu} = T^{\nu\mu}$. We have

$$j_v^\mu \equiv T^{\mu\nu} v_\nu \tag{41}$$

Covariant conservation of this current requires that $\nabla_\mu j_v^\mu = (\nabla_\mu T^{\mu\nu}) v_\nu + T^{\mu\nu}\nabla_\mu v_\nu = 0$. For an *isometry*, conservation thus requires that $\nabla_\mu T^{\mu\nu} = 0$. For a *conformal transformation*, conservation in addition requires that $T_\mu{}^\mu = 0$. Starting out with Poincaré and scale invariance, all of the above conditions will have to be met so that special conformal invariance will be automatic.

3.3. *(Super) Conformal $\mathcal{N}$=4 Super Yang-Mills*

In this subsection, we show that the global continuous symmetry group of $\mathcal{N} = 4$ SYM is given by the supergroup $SU(2,2|4)$, see Ref. 22. The ingredients are as follows.

- *Conformal Symmetry*, forming the group $SO(2,4) \sim SU(2,2)$ is generated by translations P^μ, Lorentz transformations $L_{\mu\nu}$, dilations D and special conformal transformations K^μ;
- *R-symmetry*, forming the group $SO(6)_R \sim SU(4)_R$, generated by T^A, $A = 1, \cdots, 15$;
- *Poincaré supersymmetries* generated by the supercharges Q^a_α and their complex conjugates $\bar{Q}_{\dot{\alpha}a}$, $a = 1, \cdots, 4$. The presence of these charges results immediately from $\mathcal{N} = 4$ Poincaré supersymmetry;
- *Conformal supersymmetries* generated by the supercharges $S_{\alpha a}$ and their complex conjugates $\bar{S}^a_{\dot{\alpha}}$. The presence of these symmetries results from the fact that the Poincaré supersymmetries and the special conformal transformations K_μ do not commute. Since both are symmetries, their commutator must also be a symmetry, and these are the S generators.

The two bosonic subalgebras $SO(2,4)$ and $SU(4)_R$ commute. The supercharges Q^a_α and $\bar{S}^a_{\dot{\alpha}}$ transform under the **4** of $SU(4)_R$, while $\bar{Q}_{\dot{\alpha}a}$ and $S_{\alpha a}$ transform under the **4***. From these data, it is not hard to see how the various generators fit into a superalgebra,

$$\begin{pmatrix} P_\mu \; K_\mu \; L_{\mu\nu} \; D & Q^a_\alpha \; \bar{S}^a_{\dot{\alpha}} \\ & \\ \bar{Q}_{\dot{\alpha}a} \; S_{\alpha a} & T^A \end{pmatrix} \tag{42}$$

All structure relations are rather straightforward, except the relations between the supercharges, which we now spell out. To organize the structure relations, it is helpful to make use of a natural grading of the algebra given by the dimension of the generators,

$$\begin{aligned} [D] = [L_{\mu\nu}] = [T^A] = 0 \qquad & [P^\mu] = +1 \qquad [K_\mu] = -1 \\ & [Q] = +1/2 \qquad [S] = -1/2 \end{aligned} \tag{43}$$

Thus, we have

$$\begin{aligned} \{Q^a_\alpha, Q^b_\beta\} &= \{S_{\alpha a}, S_{\beta b}\} = \{Q^a_\alpha, \bar S^b_{\dot\beta}\} = 0 \\ \{Q^a_\alpha, \bar Q_{\dot\beta b}\} &= 2\sigma^\mu_{\alpha\dot\beta} P_\mu \delta_b{}^a \\ \{S_{\alpha a}, \bar S^b_{\dot\beta}\} &= 2\sigma^\mu_{\alpha\dot\beta} K_\mu \delta_a{}^b \\ \{Q^a_\alpha, S_{\beta b}\} &= \epsilon_{\alpha\beta}\big(\delta_b{}^a D + T^a{}_b\big) + \frac{1}{2}\delta_b{}^a L_{\mu\nu}\sigma^{\mu\nu}_{\alpha\beta} \end{aligned} \tag{44}$$

3.4. *Superconformal Multiplets of Local Operators*

We shall be interested in constructing and classifying all local, gauge invariant operators in the theory that are polynomial in the canonical fields. The restriction to polynomial operators stems from the fact that it is those operators that will have definite dimension.

The canonical fields X^i, λ_a and A_μ have unrenormalized dimensions, given by 1, 3/2 and 1 respectively. Gauge invariant operators will be constructed rather from the gauge covariant objects X^i, λ_a, $F^\pm_{\mu\nu}$ and the covariant derivative D_μ, whose dimensions are

$$[X^i] = [D_\mu] = 1 \qquad [F^\pm_{\mu\nu}] = 2 \qquad [\lambda_a] = \frac{3}{2} \tag{45}$$

Here, $F^\pm_{\mu\nu}$ stands for the (anti) self-dual gauge field strength. Thus, if we temporarily ignore the renormalization effects of composite operators, we see that all operator dimensions will be positive and that the number of operators whose dimension is less than a given number is finite. The only operator with dimension 0 will be the unit operator.

Next, we introduce the notion of *superconformal primary operator.* Since the conformal supercharges S have dimension $-1/2$, successive application of S to any operator of definite dimension must at some point yield 0 since otherwise we would start generating operators of negative dimension, which is impossible in a unitary representation. Therefore one defines a *superconformal primary operator* $\mathcal{O}$ to be a non-vanishing operator such that

$$[S,\mathcal{O}]_{\pm}=0 \qquad \mathcal{O}\neq 0 \tag{46}$$

An equivalent way of defining a superconformal primary operator is as the lowest dimension operator in a given superconformal multiplet or representation. It is important to distinguish a superconformal primary operator from a *conformal primary operator*, which is instead annihilated by the special conformal generators K^{μ}, and is thus defined by a weaker condition. Therefore, every superconformal primary is also a conformal primary operator, but the converse is not, in general, true.

Finally, an operator $\mathcal{O}$ is called a *superconformal descendant operator* of an operator $\mathcal{O}'$ when it is of the form,

$$\mathcal{O}=[Q,\mathcal{O}']_{\pm} \tag{47}$$

for some well-defined local polynomial gauge invariant operator $\mathcal{O}'$. If $\mathcal{O}$ is a descendant of $\mathcal{O}'$, then these two operators belong to the same superconformal multiplet. Since the dimensions are related by $\Delta_{\mathcal{O}}=\Delta_{\mathcal{O}'}+1/2$, the operator $\mathcal{O}$ can never be a conformal primary operator, because there is in the same multiplet at least one operator $\mathcal{O}'$ of dimension lower than $\mathcal{O}$. As a result, in a given irreducible superconformal multiplet, there is a single superconformal primary operator (the one of lowest dimension) and all others are superconformal descendants of this primary.

It is instructive to have explicit forms for the superconformal primary operators in $\mathcal{N}=4$ SYM. The construction is most easily carried out by using the fact that a superconformal primary operator is NOT the Q-commutator of another operator. Thus, a key ingredient in the construction is the Q transforms of the canonical fields. We shall need these here only schematically,

$$\begin{aligned} \{Q,\lambda\} &= F^{+}+[X,X] \qquad & [Q,X] &= \lambda \\ \{Q,\bar\lambda\} &= DX & [Q,F] &= D\lambda \end{aligned} \tag{48}$$

A local polynomial operator containing any of the elements on the rhs of the above structure relations cannot be primary. As a result, chiral primary operators can involve neither the gauginos λ nor the gauge field strengths $F^{\pm}$. Being thus only functions of the scalars X, they can involve neither derivatives nor commutators of X. As a result, superconformal primary operators are gauge invariant scalars involving only X in a symmetrized way.

The simplest are the *single trace operators*, which are of the form

$$\mathrm{str}\Big(X^{i_1}X^{i_2}\cdots X^{i_n}\Big) \tag{49}$$

where i_j, $j = 1, \cdots, n$ stand for the $SO(6)_R$ fundamental representation indices. Here, "str" denotes the symmetrized trace over the gauge algebra and as a result of this operation, the above operator is totally symmetric in the $SO(6)_R$-indices i_j. In general, the above operators transform under a reducible representation (namely the symmetrized product of n fundamentals) and irreducible operators may be obtained by isolating the traces over $SO(6)_R$ indices. Since $\mathrm{tr}X^i = 0$, the simplest operators are

$$\begin{aligned} \sum_i \mathrm{tr}X^i X^i &\sim \text{Konishi multiplet} \\ \mathrm{tr}X^{\{i}X^{j\}} &\sim \text{supergravity multiplet} \end{aligned} \tag{50}$$

where $\{ij\}$ stands for the traceless part only. The reasons for these nomenclatures will become clear once we deal with the AdS/CFT correspondence.

More complicated are the *multiple trace operators*, which are obtained as products of single trace operators. Upon taking the tensor product of the individual totally symmetric representations, we may now also encounter (partially) anti-symmetrized representations of $SO(6)_R$. There is a one-to-one correspondence between chiral primary operators and unitary superconformal multiplets, and so all state and operator multiplets may be labeled in terms of the superconformal chiral primary operators.

3.5. $\mathcal{N} = 4$ *Chiral or BPS Multiplets of Operators*

The unitary representations of the superconformal algebra $SU(2,2|4)$ may be labeled by the quantum numbers of the bosonic subgroup, listed below,

$$\begin{array}{ccc} SO(1,3) \times & SO(1,1) & \times\, SU(4)_R \\ (s_+, s_-) & \Delta & [r_1, r_2, r_3] \end{array} \tag{51}$$

where $s_\pm$ are positive or zero half integers, Δ is the positive or zero dimension and $[r_1, r_2, r_3]$ are the Dynkin labels of the representations of $SU(4)_R$. It is sometimes preferable to refer to $SU(4)_R$ representations by their dimensions, given in terms of $\bar{r}_i \equiv r_i + 1$ by

$$\dim[r_1, r_2, r_3] = \frac{1}{12}\bar{r}_1\bar{r}_2\bar{r}_3(\bar{r}_1 + \bar{r}_2)(\bar{r}_2 + \bar{r}_3)(\bar{r}_1 + \bar{r}_2 + \bar{r}_3) \tag{52}$$

instead of their Dynkin labels. The complex conjugation relation is $[r_1, r_2, r_3]^* = [r_3, r_2, r_1]$.

In unitary representations, the dimensions Δ of the operators are bounded from below by the spin and $SU(4)_R$ quantum numbers. To see this, it suffices to restrict to primary operators since they have the lowest dimension in a given irreducible multiplet. As shown previously, such operators are scalars, so that the spin quantum numbers vanish, and the dimension is bounded from below by the $SU(4)_R$ quantum numbers. A systematic analysis of Ref. 24, (see also Refs. 25, 26) for this case reveals the existence of 4 distinct series,

(1) $\Delta = r_1 + r_2 + r_3$;
(2) $\Delta = \frac{3}{2}r_1 + r_2 + \frac{1}{2}r_3 \geq 2 + \frac{1}{2}r_1 + r_2 + \frac{3}{2}r_3$ this requires $r_1 \geq r_3 + 2$;
(3) $\Delta = \frac{1}{2}r_1 + r_2 + \frac{3}{2}r_3 \geq 2 + \frac{3}{2}r_1 + r_2 + \frac{1}{2}r_3$ this requires $r_3 \geq r_1 + 2$;
(4) $\Delta \geq \mathrm{Max}\left[2 + \frac{3}{2}r_1 + r_2 + \frac{1}{2}r_3 ; 2 + \frac{1}{2}r_1 + r_2 + \frac{3}{2}r_3\right]$

Clearly, cases 2 and 3 are complex conjugates of one another.

Cases 1 2 and 3 correspond to *discrete series of representations*, for which at least one supercharge Q commutes with the primary operator. Such representations are shortened and usually referred to as *chiral multiplets* or *BPS multiplets*. The term BPS multiplet arises from the analogy with the BPS multiplets of Poincaré supersymmetry discussed in subsections §2.3. Since these representations are shortened, their dimension is *unrenormalized* or protected from receiving quantum corrections.

Case 4 corresponds to *continuous series of representations*, for which no supercharges Q commute with the primary operator. Such representations are referred to as *non-BPS*. Notice that the dimensions of the operators in the continuous series is separated from the dimensions in the discrete series by a gap of at least 2 units of dimension.

The BPS multiplets play a special role in the AdS/CFT correspondence. In Table 3 below, we give a summary of properties of various BPS and non-BPS multiplets. In the column labeled by $\#Q$ is listed the number of Poincaré supercharges that leave the primary invariant.

Half-BPS operators

It is possible to give an explicit description of all 1/2 BPS operators. The simplest series is given by single-trace operators of the form

$$\mathcal{O}_k(x) \equiv \frac{1}{n_k} \mathrm{str}\Big(X^{\{i_1}(x) \cdots X^{i_k\}}(x) \Big) \tag{53}$$

where "str" stands for the symmetrized trace introduced previously,

Table 3. Characteristics of BPS and Non-BPS multiplets

Operator type	$\#Q$	spin range	$SU(4)_R$ primary	dimension Δ
identity	16	0	$[0,0,0]$	0
1/2 BPS	8	2	$[0,k,0],\ \ k \geq 2$	k
1/4 BPS	4	3	$[\ell,k,\ell],\ \ \ell \geq 1$	$k+2\ell$
1/8 BPS	2	7/2	$[\ell,k,\ell+2m]$	$k+2\ell+3m,\ \ m \geq 1$
non-BPS	0	4	any	unprotected

$\{i_1 \cdots i_k\}$ stands for the $SO(6)_R$ traceless part of the tensor, and n_k stands for an overall normalization of the operator which will be fixed by normalizing its 2-point function. The dimension of these operators is unrenormalized, and thus equal to k.

However, it is also possible to have multiple trace 1/2 BPS operators. They are built as follows. The tensor product of n representations $[0,k_1,0]\otimes \cdots \otimes [0,k_n,0]$, always contains the representation $[0,k,0]$, $k = k_1 + \cdots + k_n$, with multiplicity 1. (The highest weight of the representation $[0,k,0]$ is then the sum of the highest weights of the component representations.) The most general 1/2 BPS gauge invariant operators are given by the projection onto the representation $[0,k,0]$ of the corresponding product of operators,

$$\mathcal{O}_{(k_1,\cdots,k_n)}(x) \equiv \Big[\mathcal{O}_{k_1}(x)\cdots\mathcal{O}_{k_n}(x)\Big]_{[0,k,0]} \qquad k = k_1 + \cdots + k_n \quad (54)$$

Here the brackets [ ] stand for the operators product of the operators inside. This product is in general singular and thus ambiguous, but the projection onto the representation $[0,k,0]$ is singularity free and thus unique.

1/4 and 1/8 BPS Operators

There are no single-trace 1/4 BPS operators. The simplest construction is in terms of double trace operators. It is easiest to list all possibilities in a single expression, using the notations familiar already from the 1/2 BPS case. The operators are of the form

$$\Big[\mathcal{O}_{k_1}(x)\cdots\mathcal{O}_{k_n}(x)\Big]_{[\ell,k,\ell]} \qquad k+2\ell = k_1 + \cdots + k_n \quad (55)$$

In the free theory, the above operators will be genuinely 1/4 BPS, but in the interacting theory, the operators will also contain an admixture of descendants of non-BPS operators.[27] The series of 1/8 BPS operators starts with triple trace operators, and are generally of the form

$$\Big[\mathcal{O}_{k_1}(x)\cdots\mathcal{O}_{k_n}(x)\Big]_{[\ell,k,\ell+2m]} \qquad k+2\ell+3m = k_1 + \cdots + k_n \quad (56)$$

In the interacting theory, admixtures with descendants again have to be included.

3.6. *Problem Sets*

(3.1) Show that the 1-loop renormalization group β-function for $\mathcal{N} = 4$ SYM vanishes.

(3.2) Express the $\mathcal{N} = 4$ SYM Lagrangian in terms of $\mathcal{N} = 1$ superfields.

(3.3) Work out the full conformal $SO(2,4) \sim SU(2,2)$ and superconformal $SU(2,2|4)$ structure relations (commutators and anti-commutators of the generators).

(3.4) Derive the Noether currents associated with the Poincaré Q^a_α and conformal $\bar{S}_{\dot{\alpha}a}$ supercharges (and complex conjugates) in terms of the canonical fields of $\mathcal{N} = 4$ SYM.

(3.5) In the Abelian Coulomb phase of $\mathcal{N} = 4$ SYM, where the gauge algebra $\mathcal{G}$ is spontaneously broken to $U(1)^r$, $r = \text{rank}\ \mathcal{G}$, the global superconformal algebra $SU(2,2|4)$ is also spontaneously broken. To simplify matters, you may take $\mathcal{G} = SU(2)$. (a) Identify the generators of $SU(2,2|4)$ which are preserved and (b) those which are spontaneously broken, thus producing Goldstone bosons and fermions. (c) Express the Goldstone boson and fermion fields in terms of the canonical fields of $\mathcal{N} = 4$ SYM.

4. Supergravity and Superstring Theory

In this section, we shall review the necessary supergravity and superstring theory to develop the theory of D-branes and D3-branes in particular.

4.1. *Spinors in General Dimensions*

Consider D-dimensional Minkowski space-time M_D with flat metric $\eta_{\mu\nu} = \text{diag}(- + \cdots +)$, $\mu, \nu = 0, 1, \cdots, D-1$. The Lorentz group is $SO(1, D-1)$ and the generators of the Lorentz algebra $J_{\mu\nu}$ obey the standard structure relations

$$[J_{\mu\nu}, J_{\rho\sigma}] = -i\eta_{\mu\rho}J_{\nu\sigma} + i\eta_{\nu\rho}J_{\mu\sigma} - i\eta_{\nu\sigma}J_{\mu\rho} + i\eta_{\mu\sigma}J_{\nu\rho} \tag{57}$$

The *Dirac spinor* representation, denoted S_D, is defined in terms of the standard Clifford-Dirac matrices Γ_μ,

$$J_{\mu\nu} = \frac{i}{4}[\Gamma_\mu, \Gamma_\nu] \qquad \{\Gamma_\mu, \Gamma_\nu\} = 2\eta_{\mu\nu} \tag{58}$$

Its (complex) dimension is given by $\dim_{\mathbf{C}} S_D = 2^{[D/2]}$.

For D even, the Dirac spinor representation is always reducible because in that case there exists a chirality matrix $\bar\Gamma$, with square $\bar\Gamma^2 = I$, which anti-commutes with all Γ_μ and therefore commutes with $J_{\mu\nu}$,

$$\bar\Gamma \equiv i^{\frac{1}{2}D(D-1)+1}\Gamma_0\Gamma_1\cdots\Gamma_{d-1} \qquad \{\bar\Gamma, \Gamma_\mu\} = 0 \Rightarrow [\bar\Gamma, J_{\mu\nu}] = 0 \tag{59}$$

As a result, the Dirac spinor is the direct sum of two *Weyl spinors* $S_D = S_+ \oplus S_-$. The reality properties of the Weyl spinors depends on D (mod 8), and is given as follows,

$$\begin{array}{lll} D \equiv 0,4 \pmod 8 & S_- = S_+^* & \text{both complex} \\ D \equiv 2,6 \pmod 8 & S_+ \quad S_- & \text{self} - \text{conjugate} \end{array} \tag{60}$$

For both even and odd D, the *charge conjugate* ψ^c of a Dirac spinor ψ is defined by

$$\psi^c \equiv C\Gamma_0\psi^* \qquad C\Gamma_\mu C^{-1} = -(\Gamma_\mu)^T \tag{61}$$

Requiring that a spinor be real is a basis dependent condition and thus not properly Lorentz covariant. The proper Lorentz invariant condition for reality is that a spinor be its own charge conjugate $\psi^c = \psi$; such a spinor is called a *Majorana spinor.* The Majorana condition requires that $(\psi^c)^c = \psi$, or $C\Gamma_0(C\Gamma_0)^* = I$, which is possible only in dimensions $D \equiv 0,1,2,3,4 \pmod 8$. In dimensions $D \equiv 0,4 \pmod 8$, a Majorana spinor is equivalent to a Weyl spinor, while in dimension $D \equiv 2 \pmod 8$ it is possible to impose the Majorana and Weyl conditions at the same time, resulting in *Majorana-Weyl spinors.* In dimensions $D \equiv 5,6,7 \pmod 8$, one may group spinors into doublets $\Psi_\pm$ and it is possible to impose a *symplectic Majorana condition* given by $\Psi^c_\pm = \mp\Psi_\mp$. Useful reviews are in Refs. 12, 28.

4.2. *Supersymmetry in General Dimensions*

The basic Poincaré supersymmetry algebra in M_D is obtained by supplementing the Poincaré algebra with $\mathcal{N}$ supercharges Q^I_α, $I = 1, \cdots, \mathcal{N}$. Here Q transforms in the spinor representation S, which could be a Dirac spinor, a Weyl spinor, a Majorana spinor or a Majorana-Weyl spinor, depending on D. Thus, α runs over the spinor indices $\alpha = 1, \cdots, \mathrm{dim}.S$. Whatever the spinor is, we shall always write it as a Dirac spinor. The fundamental supersymmetry algebra could include central charges just as was the case for $D = 4$. However, we shall here be interested mostly in a restricted class of supersymmetry representations in which we have a massless graviton, such as we have in supergravity and in superstring theory. Therefore, we may ignore the central charges.

A general result, valid in dimension $D \geq 4$, states that *interacting massless fields of spin > 2 cannot be causal, and are excluded on physical grounds.* Considering theories with a massless graviton, and assuming that supersymmetry is realized linearly, the massless graviton must be part of a massless supermultiplet of states and fields. By the above general result, this multiplet cannot contain fields and states of spin > 2. This fact puts severe restrictions on which supersymmetry algebras can be realized in various dimensions.

The existence of massless unitary representations of the supersymmetry algebra requires vanishing central charges, just as was the case in $d = 4$. Thus, we shall consider the Poincaré supersymmetry algebras of the form (useful reviews are in Refs. 12, 28, see also Refs. 29 and 30),

$$\{Q^I_\alpha, (Q^J_\beta)^\dagger\} = 2\delta^I{}_J (\Gamma_\mu)_\alpha^\beta P^\mu \qquad\qquad \{Q^I_\alpha, Q^J_\beta\} = 0 \tag{62}$$

To analyze massless representations, choose $P^\mu = (E, 0, \cdots, 0, E)$, $E > 0$, so that the supersymmetry algebra in this representation simplifies and becomes

$$\{Q^I_\alpha, (Q^J_\beta)^\dagger\} = 2\delta^I{}_J \begin{pmatrix} 4E & 0 \\ 0 & 0 \end{pmatrix}_\alpha^\beta \tag{63}$$

On this unitary massless representation, half of the supercharges effectively vanish $Q^I_\alpha = 0$, $\alpha = \frac{1}{2}\dim S + 1, \cdots, \dim S$. Half of the remaining supercharges may be viewed as lowering operators for the Clifford algebra, while the other half may be viewed as raising operators. Thus, the total number of raising operators is $1/4 \cdot \mathcal{N} \cdot \dim_{\mathbf{R}} S$. Each operator raising helicity by 1/2, and total helicity ranging at most from -2 to $+2$, we should have at most 8 raising operators and this produces an important bound,

$$\mathcal{N} \dim_{\mathbf{R}} S \leq 32 \tag{64}$$

In other words, the maximum number of Poincaré supercharges is always equal to 32.

The largest dimension D for which the bound may be satisfied is $D = 11$ and $\mathcal{N} = 1$, for which there are precisely 32 Majorana supercharges. In $D = 10$, the bound is saturated for $\mathcal{N} = 2$ and 16-dimensional Majorana-Weyl spinors. There is indeed a unique $D = 11$ supergravity theory discovered by Cremmer, Julia and Scherk.[31] Many of the lower dimensional theories may be constructed by Kaluza-Klein compactification on a circle or on a torus of the $D = 11$ theory and we shall therefore treat this method first.[32]

4.3. *Kaluza-Klein Compactification on a Circle*

We wish to compactify one space dimension on a circle S^1_R of radius R. Accordingly, we decompose the coordinates x^μ of $\mathbf{R}^D$ into a coordinate y on the circle and the remaining coordinates $x^{\bar\mu}$. The wave operator with flat metric in D dimensions $\Box_D$ then becomes

$$\Box_D = \Box_{D-1} + \frac{\partial^2}{\partial y^2} \tag{65}$$

We shall be interested in finding out how various fields behave, in particular in the limit $R \to 0$, referred to as *dimensional reduction.*

We begin with a scalar field $\phi(x^\mu)$ obeying periodic boundary conditions on S^1_R, which has the following Fourier decomposition,

$$\phi(x^{\bar\mu}, y) = \sum_{n \in \mathbf{Z}} \phi_n(x^{\bar\mu}) e^{2\pi i n y/R} \tag{66}$$

The d-dimensional kinetic term of a scalar field with mass m then decomposes as follows,

$$\int d^d x \phi(-\Box_d + m^2)\phi = \sum_{n\in\mathbf{Z}} 2\pi R \int d^{d-1}x \phi_n \big(-\Box_{d-1} + m^2 + \tfrac{4\pi^2 n^2}{R^2}\big)\phi_n \tag{67}$$

As $R \to 0$, all modes except $n = 0$ acquire an infinitely heavy mass and decouple. The zero mode $n = 0$ is the unique mode invariant under translations on S^1_R. Thus, *the dimensional reduction on a circle of a scalar field with periodic boundary conditions is again a scalar field.* Under dimensional reduction with any other boundary condition, there will be no zero mode left and thus the scalar field will completely decouple.

Next, consider a bosonic field with periodic boundary conditions transforming under an arbitrary tensor representation of the Lorentz group $SO(1, D-1)$ on M_D. Let us begin with a vector field $A_\mu(x^\nu)$ in the fundamental of $SO(1, D-1)$. The index μ must now also be split into a component along the direction y and the remaining $D-1$ directions $\bar\mu$. The first results in a scalar $A_y(x^{\bar{\nu}})$, while the second results in a vector $A_{\bar{\mu}}(x^{\bar{\nu}})$ of the $D-1$ dimensional Lorentz group $SO(1, D-2)$. We notice that this decomposition is nothing but the branching rule for the fundamental representation of $SO(1, D-1)$ decomposing under the subgroup $SO(1, D-2)$. For a field A obeying period boundary conditions and transforming under a general tensor representation T of $SO(1, D-1)$, dimensional reduction on a circle will produce a direct sum of representations T_i of $SO(1, D-2)$, which is the restriction of T to the subgroup $SO(1, D-2)$.

For a spinor field obeying periodic boundary conditions and transforming under a general spinor representation S of $SO(1, D-1)$, dimensional reduction will produce a direct sum of representations S_i of $SO(1, D-2)$ which is the restriction of S to the subgroup $SO(1, D-2)$. Finally, assembling bosons and fermions with periodic boundary conditions in a supersymmetry multiplet, we see that dimensional reduction will preserve all Poincaré supersymmetries, and that the supercharges will behave as the spinor fields described above under this reduction.

An important example is the rank 2 symmetric tensor, i.e. the metric $G_{\mu\nu}$,

$$G_{\mu\nu} \to \begin{cases} G_{yy} & \text{scalar mixing with dilaton} \\ G_{\bar{\mu}y} & \text{graviphoton} \\ G_{\bar{\mu}\bar{\nu}} & \text{metric} \end{cases} \tag{68}$$

Again, fields obeying boundary conditions other than periodic will completely decouple.

4.4. *D=11 and D=10 Supergravity Particle and Field Contents*

In this subsection, we begin by listing the field contents and the number of physical degrees of freedom of the $\mathcal{N}=1$, $D=11$ supergravity theory. By dimensional reduction on a circle, we find the $\mathcal{N}=2$, $D=10$ Type IIA theory, which is parity conserving and has two Majorana-Weyl gravitini of opposite chiralities. Finally, we list the field and particle contents for the $\mathcal{N}=2$, $D=10$ Type IIB theory, which is chiral and has two Majorana-Weyl gravitini of the same chirality.

The $\mathcal{N}=1$, $D=11$ supergravity theory has the following field and particle contents,

$$D = 11 \begin{cases} G_{\mu\nu} & SO(9) & 44_B & \text{metric–graviton} \\ A_{\mu\nu\rho} & & 84_B & \text{antisymmetric rank 3} \\ \psi_{\mu\alpha} & & 128_F & \text{Majorana gravitino} \end{cases} \tag{69}$$

Here and below, the numbers following the little group (for the massless representations) $SO(9)$ represent the number of physical degrees of freedom in the multiplet. For example, the graviton in $D = 11$ is given by the rank 2 symmetric traceless representation of $SO(9)$, of dimension $9 \times 10/2 - 1 = 44$. The Majorana spinor $\psi_{\mu\alpha}$ as a vector has 9 physical components, but it also satisfies the Γ-tracelessness condition $(\Gamma^\mu)^{\beta\alpha}\psi_{\mu\alpha} = 0$, which cuts the number down to 8. The 32 component spinor satisfies a Dirac equation, which cuts its number of physical components down to 16, yielding a total of $8 \times 16 = 128$. The subscripts B and F refer to the bosonic or fermionic nature of the state.

The $\mathcal{N} = 2$, $D = 10$ Type IIA theory is obtained by dimensional reduction on a circle,

$$\text{Type IIA} \begin{cases} G_{\mu\nu} & SO(8) & 35_B & \text{metric–graviton} \\ \Phi & & 1_B & \text{dilaton} \\ B_{\mu\nu} & & 28_B & \text{NS–NS rank 2 antisymmetric} \\ A_{3\mu\nu\rho} & & 56_B & \text{antisymmetric rank 3} \\ A_{1\mu} & & 8_B & \text{graviphoton} \\ \psi^{\pm}_{\mu\alpha} & & 112_F & \text{Majorana–Weyl gravitinos} \\ \lambda^{\pm}_{\alpha} & & 16_F & \text{Majorana–Weyl dilatinos} \end{cases} \tag{70}$$

Here, the gravitinos are again Γ-traceless. The two gravitinos $\psi^{\pm}_{\mu\alpha}$ as well as the two dilatinos $\lambda^{\pm}_{\alpha}$ have opposite chiralities and the theory is parity conserving.

The $\mathcal{N} = 2$, $D = 10$ Type IIB theory has the following field and particle contents,

$$\text{Type IIB} \begin{cases} G_{\mu\nu} & SO(8) & 35_B & \text{metric–graviton} \\ C + i\Phi & & 2_B & \text{axion–dilaton} \\ B_{\mu\nu} + iA_{2\mu\nu} & & 56_B & \text{rank 2 antisymmetric} \\ A^{+}_{4\mu\nu\rho\sigma} & & 35_B & \text{antisymmetric rank 4} \\ \psi^{I}_{\mu\alpha}\ {}_{I=1,2} & & 112_F & \text{Majorana–Weyl gravitinos} \\ \lambda^{I}_{\alpha}\ {}_{I=1,2} & & 16_F & \text{Majorana–Weyl dilatinos} \end{cases} \tag{71}$$

The rank 4 antisymmetric tensor $A^{+}_{\mu\nu\rho\sigma}$ has self-dual field strength, a fact that is indicated with the + superscript. The gravitinos are again Γ-traceless. The two gravitinos $\psi^{I}_{\mu\alpha}$ have the same chirality, while the two dilatinos λ^I_α also have the same chirality but opposite to that of the gravitinos. The theory is chiral or parity violating.

4.5. *D=11 and D=10 Supergravity Actions*

Remarkably, the $D = 11$ supergravity theory has a relatively simple action. It is convenient to use exterior differential notation for all anti-symmetric tensor fields, such as the rank 3 tensor $A_3 \equiv 1/3! A_{3\mu\nu\rho} dx^\mu dx^\nu dx^\rho$, with field strength $F_4 \equiv dA_3$,

$$S_{11} = \frac{1}{2\kappa_{11}^2} \int \left[\sqrt{G}(R_G - \frac{1}{2}|F_4|^2) - \frac{1}{6} A_3 \wedge F_4 \wedge F_4 \right] + \text{fermions} \tag{72}$$

where κ_{11}^2 is the 11-dimensional Newton constant. The action for the Type IIA theory may be deduced from this action by dimensional reduction, but we shall not need it here. There are also $D = 10$ supergravities with only $\mathcal{N} = 1$ supersymmetry, which in particular may couple to $D = 10$ super-Yang-Mills theory.

There exists no completely satisfactory action for the Type IIB theory, since it involves an antisymmetric field A_4^+ with self-dual field strength. However, one may write an action involving both dualities of A_4 and then impose the self-duality as a supplementary field equation. Doing so, one obtains[i] (see for example Refs. 33, 28)

$$\begin{aligned} S_{IIB} = &+\frac{1}{4\kappa_B^2} \int \sqrt{G} e^{-2\Phi} (2R_G + 8\partial_\mu \Phi \partial^\mu \Phi - |H_3|^2) \\ &-\frac{1}{4\kappa_B^2} \int \left[\sqrt{G}(|F_1|^2 + |\tilde{F}_3|^2 + \frac{1}{2}|\tilde{F}_5|^2) + A_4^+ \wedge H_3 \wedge F_3 \right] + \text{fermions} \end{aligned} \tag{73}$$

where the field strengths are defined by

$$\begin{cases} F_1 = dC \\ H_3 = dB \\ F_3 = dA_2 \\ F_5 = dA_4^+ \end{cases} \qquad \begin{cases} \tilde{F}_3 = F_3 - CH_3 \\ \tilde{F}_5 = F_5 - \frac{1}{2} A_2 \wedge H_3 + \frac{1}{2} B \wedge F_3 \end{cases} \tag{74}$$

and we have the supplementary self-duality condition $*\tilde{F}_5 = \tilde{F}_5$.

---

[i]We use the notation

$$G \equiv -\det G_{\mu\nu}$$

and

$$\int \sqrt{G} |F_p|^2 \equiv \frac{1}{p!} \int \sqrt{G} G^{\mu_1 \nu_1} \cdots G^{\mu_p \nu_p} \bar{F}_{\mu_1 \cdots \mu_p} F_{\nu_1 \cdots \nu_p}$$

where $\bar{F}$ denotes the complex conjugate of F. For real fields, this definition coincides with that of Ref. 28

The above form of the action naturally arises from the string low energy approximation. The first line in (73) originates from the NS-NS sector while the second line (except for the fermions) originates from the RR sector, as we shall see shortly. Type IIB supergravity is invariant under the non-compact symmetry group $SU(1,1) \sim SL(2,\mathbf{R})$, but this symmetry is not manifest in (73). To render the symmetry manifest, we redefine fields from the *string metric* $G_{\mu\nu}$ used in (73) to the *Einstein metric* $G_{E\mu\nu}$, along with expressing the tensor fields in terms of complex fields,[ii]

$$G_{E\mu\nu} \equiv e^{-\Phi/2} G_{\mu\nu} \qquad \tau \equiv C + ie^{-\Phi}$$
$$G_3 \equiv (F_3 - \tau H_3)/\sqrt{\mathrm{Im}\tau} \tag{75}$$

The action may then be written simply as,

$$\begin{aligned} S_{\mathrm{IIB}} = & \frac{1}{4\kappa_B^2} \int \sqrt{G_E} \Big(2R_{G_E} - \frac{\partial_\mu \bar{\tau} \partial^\mu \tau}{(\mathrm{Im}\tau)^2} - \frac{1}{2}|F_1|^2 - |G_3|^2 - \frac{1}{2}|\tilde{F}_5|^2 \Big) \\ & - \frac{1}{4i\kappa_B^2} \int A_4 \wedge \bar{G}_3 \wedge G_3 \end{aligned} \tag{76}$$

Under the $SU(1,1) \sim SL(2,\mathbf{R})$ symmetry of Type IIB supergravity, the metric and A_4^+ fields are left invariant. The dilaton-axion field τ changes under a Möbius transformation,

$$\tau \to \tau' = \frac{a\tau + b}{c\tau + d} \qquad ad - bc = 1, \;\; a, b, c, d \in \mathbf{R} \tag{77}$$

Finally, the $B_{\mu\nu}$ and $A_{2\mu\nu}$ fields rotate into one another under the linear transformation associated with the above Möbius transformation, and this may most easily be re-expressed in terms of the complex 3-form field G_3,

$$G_3 \to G_3' = \frac{c\bar{\tau} + d}{|c\tau + d|} G_3 \tag{78}$$

The susy transformation laws of Type IIB supergravity[33,34] on the fermion fields – the dilatino λ and the gravitino ψ_M – are of the form,

---

[ii]The detailed relation with the $SU(1,1)$ formulation of Type IIB supergravity is given as follow : the $SU(1,1)$ frame $V_\pm^\alpha$, $\alpha = 1,2$ is given by $V_+^1 = \tau/\sqrt{\mathrm{Im}\tau}$, $V_-^1 = \bar{\tau}/\sqrt{\mathrm{Im}\tau}$, and $V_\pm^2 = 1/\sqrt{\mathrm{Im}\tau}$. The frame transforms as a $SU(1,1)$ doublet and satisfied $V_-^1 V_+^2 - V_-^2 V_+^1 = 1$. The complex 3-form is defined by $G_3 = V_+^2 F_3 - V_+^1 H_3$ and is a $SU(1,1)$ singlet. The complex variable τ parametrizes the coset $SU(1,1)/U(1)$; under this local $U(1)$ group, $V_\pm$ have charge ± 1 while G_3 has charge $+1$.

(we shall not need the transformation laws on bosons),

$$
\begin{aligned}
\delta\lambda &= \frac{i}{\kappa_B}\Gamma^\mu\eta^*\frac{\partial_\mu\tau}{\mathrm{Im}\tau} - \frac{i}{24}\Gamma^{\mu\nu\rho}\eta G_{3\mu\nu\rho} + (\mathrm{Fermi})^2 \\
\delta\psi_\mu &= \frac{1}{\kappa_B}D_\mu\eta + \frac{i}{480}\Gamma^{\mu_1\cdots\mu_5}\Gamma_\mu\eta F_{5\mu_1\cdots\mu_5} \\
&\quad + \frac{1}{96}(\Gamma_\mu{}^{\rho\sigma\tau}G_{3\rho\sigma\tau} - 9\Gamma^{\nu\rho}G_{3\mu\nu\rho})\eta^* + (\mathrm{Fermi})^2
\end{aligned}
\tag{79}
$$

Note that in the $SU(1,1)$ formulation, the supersymmetry transformation parameter η has $U(1)$ charge $1/2$, so that λ has charge $3/2$ and ψ_μ has charge $1/2$.

4.6. *Superstrings in $D = 10$*

The geometrical data of superstring theory in the Ramond-Neveu-Schwarz (RNS) formulation are the bosonic worldsheet field x^μ and the fermionic worldsheet fields $\psi^\mu_\pm$, which may both be viewed as functions of local worldsheet coordinates ξ^1, ξ^2. The subscript $\pm$ indicates the two worldsheet chiralities. Both x^μ and $\psi^\mu_\pm$ transform under the vector representation of the space-time Lorentz group. The theory has two sectors, the Neveu-Schwarz (NS) and Ramond (R) sectors. The NS ground state is a space-time boson, while the R ground state is a space-time fermion. The full space-time bosonic (resp. fermionic) spectrum of the theory is obtained by applying x^μ and ψ^μ fields to the NS (resp. R) ground states. Space-time supersymmetry is achieved by imposing a suitable Gliozzi-Scherk-Olive (GSO) projection.[29] For simplicity, we shall only consider theories with orientable strings; the Type II and heterotic string theories fit in this category. Interactions arise from the joining and splitting of the worldsheets, so that the number of handles (which equals the genus for orientable worldsheets) corresponds to the number of loops in a field theory reinterpretation of the string diagram. (Standard references on superstring theory include Refs. 34, 28, lecture notes[35] and a review on perturbation theory.[36])

One aspect of string theory that we shall make use of in these lectures is the fact that (1) the low energy limit of string theory is supergravity and that (2) string theory produces definite and calculable higher derivative corrections to the supergravity action and field equations. To explain these facts, it is easiest to concentrate on the space-time bosonic fields, since space-time fermionic fields require the use of the more complicated fermion vertex operator. For Type II theories, the space-time bosons arise from two sectors in turn; the NS-NS sector and the R-R sector. Fields in the R-R sector again couple to the string worldsheet through the use of the fermion vertex operator, and for simplicity we shall ignore also these fields here (even

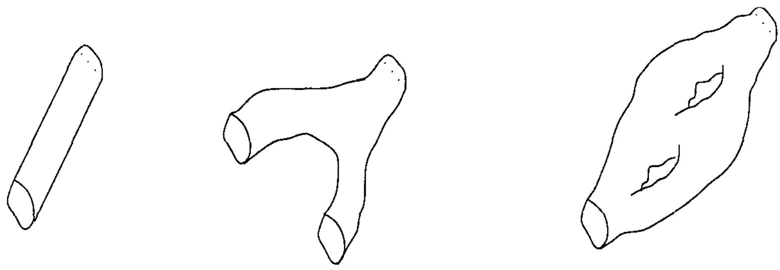


Figure 1. Propagating closed strings (a) free, (b) interaction, (c) two-loop

though they will of course be very important for the AdS/CFT conjecture). The remaining fields are now the same for all four closed orientable string theories, Type IIA, Type IIB and the two heterotic strings, namely the metric $G_{\mu\nu}$, the NS-NS antisymmetric rank 2 tensor $B_{\mu\nu}$ and the dilaton Φ. The full worldsheet action for the coupling of these fields is still very complicated on a worldsheet with general worldsheet metric and worldsheet gravitino fields χ_m. The contribution from the worldsheet bosonic field x^μ gives rise to a generalized non-linear sigma model,

$$S_x = \frac{1}{4\pi\alpha'} \int_\Sigma \sqrt{\gamma} \big[\{ \gamma^{mn} G_{\mu\nu}(x) + \epsilon^{mn} B_{\mu\nu}(x) \} \partial_m x^\mu \partial_n x^\nu + \alpha' R_\gamma^{(2)} \Phi(x) \big] \tag{80}$$

where α' is the square of the Planck length, γ_{mn} is the worldsheet metric, γ^{mn} its inverse and $R_\gamma^{(2)}$ its associated Gaussian curvature. The contribution from the worldsheet fermionic field $\psi_\pm^\mu$ gives rise to a worldsheet supersymmetric completion of the above non-linear sigma model. Here, we quote its form only for a flat worldsheet metric and vanishing worldsheet gravitino field,

$$S_\psi = \frac{1}{4\pi\alpha'} \int_\Sigma d^2\xi \big[G_{\mu\nu}(x) (\psi_+^\mu D_{\bar z} \psi_+^\nu + \psi_-^\mu D_z \psi_-^\nu) + \tfrac{1}{2} R_{\mu\nu\rho\sigma} \psi_+^\mu \psi_+^\nu \psi_-^\rho \psi_-^\sigma \big] \tag{81}$$

where $R_{\mu\nu\rho\sigma}$ is the Riemann tensor for the metric $G_{\mu\nu}$ and the covariant derivatives are given by

$$\begin{aligned} D_{\bar z} \psi_+^\mu &= \partial_{\bar z} \psi_+^\mu + \left(\Gamma^\mu_{\rho\sigma}(x) + \frac{1}{2} H_3{}^\mu{}_{\rho\sigma}(x) \right) \partial_{\bar z} x^\rho \psi_+^\sigma \\ D_z \psi_-^\mu &= \partial_z \psi_-^\mu + \left(\Gamma^\mu_{\rho\sigma}(x) - \frac{1}{2} H_3{}^\mu{}_{\rho\sigma}(x) \right) \partial_z x^\rho \psi_-^\sigma \end{aligned} \tag{82}$$

where $H_{3\mu\rho\sigma}$ is the field strength of $B_{\mu\nu}$ and $\Gamma^\mu_{\rho\sigma}$ is the Levi-Civita connections for G.

The non-chiral scattering amplitudes are given by the functional integral over all x^μ and $\psi_\pm$ as well as over all worldsheet metrics γ_{mn} and all worldsheet gravitini fields χ_m by

$$\text{amplitude} = \sum_{\text{topologies}} \int D\gamma_{mn} D\chi_m \int Dx^\mu D\psi \; e^{-S_x + S_\psi} \tag{83}$$

The full amplitudes must then be obtained by first chirally splitting[36,37] the non-chiral amplitudes in terms of the conformal blocks of the corresponding conformal field theories of the left and right movers and imposing the GSO projection.

The quantization prescription given by the above formula for the amplitude is in the first quantized formulation of string theory. There, a given string configuration (a given worldsheet topology) is quantized in the presence of external background fields, such as the metric $G_{\mu\nu}$, the rank 2 anti-symmetric tensor field $B_{\mu\nu}$ and the dilaton Φ. The quantization of the string produces excitations of these very fields as well as of all the other string modes. In comparison with the first quantized formulation of particles is field theory, the background fields may be interpreted as vacuum expectation values of the corresponding field operators.

If the vacuum expectation value of the dilaton field is $\phi = \langle \Phi \rangle$, then the contribution of the vacuum expectation value to the string amplitude is governed by the Euler number $\chi(\Sigma)$ of the worldsheet Σ,

$$\frac{1}{2\pi} \int_\Sigma \sqrt{\gamma} R_\gamma^{(2)} = \chi(\Sigma) = 2 - 2h - b \tag{84}$$

where h is the genus or number of handles and b is the number of boundaries or punctures. Therefore, a genus h worldsheet (without boundary) will receive a multiplicative contribution of $e^{-(2-2h)\phi} = g_s^{2h-2}$ which gives reason to identify $g_s = e^\phi$ with the (closed) string coupling constant. For open string theories, the expansion is rather in integer powers of the open string coupling constant $g_o = e^{\phi/2}$.

4.7. *Conformal Invariance and Supergravity Field Equations*

As a two-dimensional quantum field theory, the generalized non-linear sigma model makes sense for any background field assignment. However, when the non-linear sigma model is to define a consistent string theory, further physical conditions need to be satisfied. The most crucial one is that the single string spectrum be free of negative norm states. Such states always appear because Poincaré invariance of the theory forces the string map x^μ to obey the following canonical relations $[x^\mu, \dot{x}^\nu] \sim G^{\mu\nu}$, so that x^0 creates negative norm states.

The decoupling of negative norm states out of the Fock space construction occurs via *worldsheet conformal invariance* of the non-linear sigma model. In particular, conformal invariance requires worldsheet scale invariance of the full quantum mechanical non-linear sigma model. Transformations of the worldsheet scale Λ are broken by quantum mechanical anomalies whose form is encoded by the β-functions of the renormalization group (RG). As will be explained in the next paragraph, each background field has a β-function, and worldsheet scale and conformal invariance thus require the vanishing of these β-functions.

The background fields $G_{\mu\nu}(x)$, $B_{\mu\nu}(x)$ and $\Phi(x)$ may be viewed as generating functions for an infinite series of coupling constants. For example, for the metric we have,

$$G_{\mu\nu}(x) = \sum_{n=0}^{\infty} \frac{1}{n!}(x-x_0)^{\mu_1} \cdots (x-x_0)^{\mu_n} \partial_{\mu_1} \cdots \partial_{\mu_n} G_{\mu\nu}(x_0) \qquad (85)$$

where each of the Taylor expansion coefficients $\partial_{\mu_1} \cdots \partial_{mu_n} G_{\mu\nu}(x_0)$ may be viewed as an independent set of couplings. Under renormalization, and thus under RG flow, this infinite number of couplings flows into itself, and the corresponding flows may again be described by generating functions $\beta^G_{\mu\nu}(x)$, $\beta^B_{\mu\nu}(x)$ and $\beta^\Phi(x)$ defined, for example, for the metric by

$$\beta^G_{\mu\nu}(x) = \frac{\partial G_{\mu\nu}}{\partial \ln \Lambda} \equiv \sum_{n=0}^{\infty} \frac{1}{n!}(x-x_0)^{\mu_1} \cdots (x-x_0)^{\mu_n} \frac{\partial \partial_{\mu_1} \cdots \partial_{\mu_n} G_{\mu\nu}(x_0)}{\partial \ln \Lambda} \qquad (86)$$

Customarily, when an infinite number of couplings occur in a quantum field theory, it is termed *non-renormalizable*, because the prediction of any physical observable would require an infinite number of input data to be specified at the renormalization point. In string theory, however, this infinite number of couplings is exactly what is required to describe the dynamics of a string in a consistent background. We now explain how this comes about.

First, we assume that the whole renormalization process of the non-linear sigma model will preserve space-time diffeomorphism invariance. The number of terms that can appear in the RG flow is then finite, order by order in the α' expansion.[38] Second, the presence of an infinite number of couplings makes it possible to have the string propagate in an infinite family of space-times. The leading order β-functions are given by Ref. 39 as

$$\begin{aligned}
\beta^G_{\mu\nu} &= \tfrac{1}{2}R_{\mu\nu} - \tfrac{1}{8}H_{\mu\rho\sigma}H_\nu{}^{\rho\sigma} + \partial_\mu\Phi\partial_\nu\Phi + \mathcal{O}(\alpha') \qquad (87)\\
\beta^B_{\mu\nu} &= -\tfrac{1}{2}D_\rho H^\rho{}_{\mu\nu} + \partial_\rho H^\rho{}_{\mu\nu} + \mathcal{O}(\alpha')\\
\beta^\Phi &= \tfrac{1}{6}(D-10) + \alpha'\left[2\partial_\mu\Phi\partial^\mu\Phi - 2\nabla^\mu\partial_\mu\Phi + \tfrac{1}{2}R_G - \tfrac{1}{24}H_{\mu\nu\rho}H^{\mu\nu\rho}\right] + \mathcal{O}(\alpha')^2
\end{aligned}$$

To leading order in α', the requirement of scale invariance reduces precisely to the supergravity field equations for the Type II theory where all RR A-fields have been (consistently) set to 0. String theory provides higher α' corrections to the supergravity field equations, which by dimensional analysis must be also terms with higher derivatives in x^μ.

4.8. *Branes in Supergravity*

A rank $p+1$ antisymmetric tensor field $A_{\mu_1\cdots\mu_{p+1}}$ may be identified with a $(p+1)$-form,

$$A_{p+1} \equiv \frac{1}{(p+1)!}A_{\mu_1\cdots\mu_{p+1}}dx^{\mu_1}\wedge\cdots\wedge dx^{\mu_{p+1}} \qquad (88)$$

A $(p+1)$-form naturally couples to geometrical objects Σ_{p+1} of space-time dimension $p+1$, because a diffeomorphism invariant action may be constructed as follows

$$S_{p+1} = T_{p+1}\int_{\Sigma_{p+1}} A_{p+1} \qquad (89)$$

The action is invariant under Abelian gauge transformations $\rho_p(x)$ of rank p

$$A_{p+1} \to A_{p+1} + d\rho_p \qquad (90)$$

because S_{p+1} transforms with a total derivative. The field A_{p+1} has a gauge invariant field strength F_{p+2}, which is a $p+2$ form whose flux is conserved. *Solutions to supergravity with non-trivial A_{p+1} charge are referred to as p-branes, after the space-dimension of their geometry.*

Each A_{p+1} gauge field has a magnetic dual A^{magn}_{D-3-p} which is a differential form field of rank $D-3-p$, whose field strength is related to that of A_{p+1} by Poincaré duality

$$dA^{\text{magn}}_{D-3-p} \equiv *dA_{p+1} \qquad (91)$$

Accordingly, each p-brane also has a magnetic dual, which is a $(D-4-p)$ brane and which now couples to the field A^{magn}_{D-3-p}.

The possible branes in $D = 11$ supergravity are very restricted because the only antisymmetric tensor field in the theory is $A_{\mu\nu\rho}$ of rank 3, so that we have a 2-brane, denoted $M2$ and its magnetic dual $M5$. The branes in Type IIA/B theory are further distinguished as follows. When the antisymmetric field whose charge they carry is in the R-R sector, the brane is referred to as a *D-brane*. D-branes were introduced first in string theory in Ref. 40. On the other hand, the 1-brane that couples to the NS-NS field $B_{\mu\nu}$ is nothing but the fundamental string, denoted F1, whose magnetic dual is NS5.[41] Below we present a Table of the branes occurring for various p in the $D = 11$ supergravity and in the Type IIA/B supergravities in $D = 10$.

4.9. *Brane Solutions in Supergravity*

Each brane is realized as a 1/2 BPS solution in supergravity. The geometry of these solutions will be important, and we describe it now. A p-brane has a $(p + 1)$-dimensional flat hypersurface, with Poincaré invariance group $\mathbf{R}^{p+1} \times SO(1,p)$. The transverse space is then of dimension $D - p - 1$ and solutions may always be found with maximal rotational symmetry $SO(D - p - 1)$ in this transverse space. Thus, p-branes in supergravity may be thought of as solutions with symmetry groups

$$\begin{cases} D = 11 & \mathbf{R}^{p+1} \times SO(1,p) \times SO(10-p) \\ D = 10 & \mathbf{R}^{p+1} \times SO(1,p) \times SO(9-p) \end{cases} \tag{92}$$

For example the $M2$ brane has symmetry group $\mathbf{R}^3 \times SO(1,2) \times SO(8)$ while the $D3$ brane has instead $\mathbf{R}^4 \times SO(1,3) \times SO(6)$. We shall denote the coordinates as follows

$$\begin{array}{llll} \text{Coordinates // to brane} & x^\mu & & \mu = 0, 1, \cdots, p \\ \text{Coordinates} \perp \text{to brane} & y^u = x^{p+u} & & u = 1, 2, \cdots, D-p-1 \end{array}$$

Table 4. Branes in various theories

name	$D = 11$	Type IIA	Type IIB	Magnetic Dual
D(-1) instanton	—	—	$A_0 = C + ie^{-\Phi}$	D7
D0 particle	—	$A_{1\mu}$	—	D6
F1 string	—	$B_{\mu\nu}$	$B_{\mu\nu}$	NS5
D1 string	—	—	$A_{2\mu\nu}$	D5
M2 membrane	$A_{\mu\nu\rho}$	—	—	M5
D2 brane	—	$A_{3\mu\nu\rho}$	—	D4
D3 brane	—	—	$A^+_{4\mu\nu\rho\sigma}$	D3

Poincaré invariance in $p+1$ dimensions forces the metric in those directions to be a rescaling of the Minkowski flat metric, while rotation invariance in the transverse directions forces the metric in those directions to be a rescaling of the Euclidean metric in those dimensions. Furthermore, the metric rescaling functions should be independent of x^μ, $\mu = 0, 1, \cdots, p$. Substituting an Ansatz with the above restrictions into the field equations, one finds that the solution may be expressed in terms of a single function H as follows,[42]

$$\begin{array}{lll} \text{Dp} & ds^2 = H(\vec{y})^{-1/2} dx^\mu dx_\mu + H(\vec{y})^{1/2} d\vec{y}^2 & e^\Phi = H(\vec{y})^{(3-p)/4} \\ \text{NS5} & ds^2 = dx^\mu dx_\mu + H(\vec{y}) d\vec{y}^2 & e^{2\Phi} = H(\vec{y}) \\ \text{M2} & ds^2 = H(\vec{y})^{-2/3} dx^\mu dx_\mu + H(\vec{y})^{1/3} d\vec{y}^2 & \\ \text{M5} & ds^2 = H(\vec{y})^{-1/3} dx^\mu dx_\mu + H(\vec{y})^{2/3} d\vec{y}^2 & \end{array} \tag{93}$$

Here, the Dp metric is expressed in the string frame. The single function H must be harmonic with respect to $\vec{y}$.

Assuming maximal rotational symmetry by $SO(D-p-1)$ in the transversal dimensions, and using the fact that the metric should tend to flat space-time as $y \to \infty$, the most general solution is parametrized by a single scale factor L and is given by

$$H(y) = 1 + \frac{L^{D-p-3}}{y^{D-p-3}} \tag{94}$$

Since α' is the only dimensionful parameter of the theory, L must be a numerical constant (possibly dependent on the dimensionless string couplings) times the above α' dependence. Of particular interest will be the solution of N coincident branes, for which we have $L^{D-p-3} = N\rho_p$. For Dp branes, we have $\rho_p = g_s\ (4\pi)^{(5-p)/2}\Gamma((7-p)/2)(\alpha')^{(D-p-3)/2}$.

It is easy to see that one still has a solution when H is harmonic without insisting on rotation invariance in the transverse space, so that the general solution is of the form,

$$H(\vec{y}) = 1 + \sum_{I=1}^{N} \frac{C_I}{|\vec{y} - \vec{y}_I|^{D-3-p}} \qquad C_I = N_I \rho_p\,,\ N_I \in \mathbf{N} \tag{95}$$

for any array of N points $\vec{y}$.

It is very important in the theory of branes in Type IIA/B string theory to understand the dependence of the string coupling g_s of the various brane solutions, in particular of c_p. To do so, we return to the supergravity field equations, (omitting derivative terms in the dilaton and axion fields for simplicity),

$$\text{IIA} \quad R_{\mu\nu} = \frac{1}{4} H_{\mu\rho\sigma} H_\nu{}^{\rho\sigma} + e^{2\Phi} \left[F_{2\mu\rho} F_{2\nu}{}^\rho + \frac{1}{6} F_{4\mu\sigma\rho\tau} F_{4\nu}{}^{\rho\sigma\tau} \right] \tag{96}$$

$$\text{IIB} \quad R_{\mu\nu} = \frac{1}{4} H_{\mu\rho\sigma} H_\nu{}^{\rho\sigma} + e^{2\Phi} \left[F_{1\mu} F_{1\nu} + \frac{1}{4} \tilde F_{3\mu\sigma\rho} \tilde F_{3\nu}{}^{\rho\sigma} + \frac{1}{24} \tilde F^+_{5\mu\rho\sigma\tau\upsilon} \tilde F^{+\rho\sigma\tau\upsilon}_{5\nu} \right]$$

Recall that the string coupling is given by $g_s = e^\phi$ where $\phi = \langle \Phi \rangle$. In both Type IIA and Type IIB, the fundamental string F1 and the NS5 brane have non-vanishing $H_{\mu\rho\sigma}$ fields, but *vanishing RR fields* F_i. Therefore, these brane solutions do not involve the string coupling constant g_s and ρ_p is independent of g_s. D-brane solutions on the other hand will have $H_{\mu\rho\sigma} = 0$, but have at least one of the R-R antisymmetric fields $F_i \neq 0$. Such solutions will involve the string coupling explicitly and therefore $\rho_p \sim g_s$. This leads for example to the expression given for ρ_p above. Each brane solution breaks precisely half of the supersymmetries of the corresponding theory, as is shown in Problem Set (4.1).

4.10. *Branes in Superstring Theory*

While originally found as solutions to supergravity field equations, the p-branes of Type IIA/B supergravity are expected to extend to solutions of the full Type IIA/B string equations. These solutions will then break precisely half of the supersymmetries of the string theory. As compared to the supergravity solutions, the full string solutions may, of course, be subject to α' corrections of their metric and other fields. Often, it is useful to compare these semi-classical solutions of string theory with solitons in quantum field theory, such as the familiar 't Hooft–Polyakov magnetic monopole. The fundamental string F1 and the NS5 brane indeed very much behave as large size semi-classical solitons, whose energy depends on the string coupling via $1/g_s^2$, as is familiar from solitons in quantum field theory.

Besides its supergravity low energy limit, the only other well-understood limit of string theory is that of weak coupling where $g_s \to 0$. It is in this approximation that string theory may be defined in terms of a genus expansion in string worldsheets. Remarkably, D-branes (but not the F1 string or NS5 branes) admit a special limit as well. As may be seen from (95), in the limit where $g_s \to 0$, the metric becomes flat everywhere, except on the $(p+1)$-dimensional hyperplane characterized by $\vec y = 0$, where the metric appears to be singular. Thus, in the weak-coupling limit, the D-brane solution of supergravity reduces to a localized defect in flat space-time. Strings propagating in this background are moving in flat space-time, except when the string reaches the D-brane. The interaction of the string with the D-brane is summarized by a boundary condition on the string dynamics. The correct conditions turn out to be Dirichlet boundary conditions in the directions perpendicular to the brane and Neumann conditions parallel to the brane. The Dp-brane may alternatively be described in string perturbation theory as a $(p+1)$-dimensional hypersurface in flat 10-dimensional space-time on which open strings end with the above boundary conditions. The open string end points are thus tied to be on the brane, but can move freely along the brane. This was indeed the original formulation;[40] see also Ref. 43.

4.11. *The Special Case of D3 branes*

The D3-brane solution is of special interest for a variety of reasons : (1) its worldbrane has 4-dimensional Poincaré invariance; (2) it has constant axion and dilaton fields; (3) it is regular at $y = 0$; (4) it is self-dual. Given its special importance, we shall present here a more complete description of the D3-brane. The solution is characterized by

$$\begin{cases} g_s = e^{\phi}, \ C \text{ constant} \\ B_{\mu\nu} = A_{2\mu\nu} = 0 \\ ds^2 = H(y)^{-1/2} dx^\mu dx_\mu + H(y)^{1/2}(dy^2 + y^2 d\Omega_5^2) \\ F^+_{5\mu\nu\rho\sigma\tau} = \epsilon_{\mu\nu\rho\sigma\tau\upsilon}\partial^\upsilon H \end{cases} \tag{97}$$

Here, $\epsilon_{\mu\nu\rho\sigma\tau\upsilon}$ is the volume element transverse to the 4-dimensional Minkowski D3-brane in $D = 10$. The N-brane solution with general locations of N_I parallel D3-branes located at transverse position $\vec{y}_i$ is given by

$$H(\vec{y}) = 1 + \sum_{I=1}^{N} \frac{4\pi g_s N_I(\alpha')^2}{|\vec{y} - \vec{y}_I|^4} \tag{98}$$

where the total number of D3-branes is $N = \sum_I N_I$. The fact that the geometry is regular as $\vec{y} \to \vec{y}_I$ despite the apparent singularity in the metric will be shown in the next section.

It is useful to compare the scales involved in the D3 brane solution and their relations with the coupling constant.[iii] The radius L of the D3

[iii]The discussion given here may be extended to Dp branes to some extent. However,

brane solution to string theory is a scale that is not necessarily of the same order of magnitude as the Planck length ℓ_P, which is defined by $\ell_P^2 = \alpha'$. Their ratio is given instead by $L^4 = 4\pi g_s N \ell_P^4$. For $g_s N \ll 1$, the radius L is much smaller than the string length ℓ_P, and thus the supergravity approximation is not expected to be a reliable approximation to the full string solution. In this regime we have $g_s \ll 1$, so that string perturbation theory is expected to be reliable and the D3 brane may be treated using conformal field theory techniques. For $g_s N \gg 1$, the radius L is much larger than the string length ℓ_P, and thus the supergravity approximation is expected to be a good approximation to the full string solution. It is possible to have at the same time $g_s \ll 1$ provided N is very large, so string perturbation theory may be *simultaneously* a good approximation.

The D3 brane solution is more properly a two-parameter family of solutions, labeled by the string coupling g_s and the instanton angle $\theta_I = 2\pi C$, or the single complex parameter $\tau = C + ie^{-\phi}$. The $SU(1,1) \sim SL(2,\mathbf{R})$ symmetry of Type IIB supergravity acts transitively on τ, so all solutions lie in a single orbit of this group. In superstring theory, however, the range of θ_I is quantized so that the identification $\theta_I \sim \theta_I + 2\pi$ may be made, and as a result also $\tau \sim \tau + 1$. Therefore, the allowed Möbius transformations must be elements of the $SL(2,\mathbf{Z})$ subgroup of $SL(2,\mathbf{R})$, for which $a, b, c, d \in \mathbf{Z}$. These transformations map between equivalent solutions in string theory. Thus, the string theories defined on D3 backgrounds which are related by an $SL(2,\mathbf{Z})$ duality will be equivalent to one another. This property will be of crucial importance in the AdS/CFT correspondence where it will emerge as the reflection of Montonen-Olive duality in $\mathcal{N} = 4$ SYM theory.

4.12. *Problem Sets*

(4.1) The Lagrangian for $D = 10$ super-Yang-Mills theory (which is constructed to be invariant under $\mathcal{N} = 1$ supersymmetry) is given by

$$\mathcal{L} = -\frac{1}{2g^2}\mathrm{tr}(F_{\mu\nu}F^{\mu\nu} - 2i\bar{\lambda}\Gamma^\mu D_\mu \lambda) \tag{99}$$

The supersymmetry transformations are given by ($\Gamma^{\mu\nu} \equiv \frac{1}{2}[\Gamma^\mu, \Gamma^\nu]$)

$$\delta A_\mu = -i\bar{\zeta}\Gamma_\mu \lambda \qquad\qquad \delta\lambda = \frac{1}{2}F_{\mu\nu}\Gamma^{\mu\nu}\zeta \tag{100}$$

when $p \neq 3$, the dilaton is not constant and the strength of the coupling will depend upon the distance to the brane.

for a Majorana-Weyl spinor gaugino λ. Show that under dimensional reduction on a flat 6-dimensional torus, (with periodic boundary conditions on all fields), the theory reduces to $D = 4$, $\mathcal{N} = 4$ super-Yang-Mills. Use this reduction to relate the matrices C_i in the Lagrangian for the $D = 4$ theory to the Clifford Dirac matrices of $SO(6)$, and to derive the supersymmetry transformations of the theory.

(4.2) Assume the following Ansatz for a D3 brane solution to the Type IIB sugra field equations : constant dilaton ϕ, vanishing axion $C = 0$, vanishing two-forms $A_{2\mu\nu} = B_{\mu\nu} = 0$, $F_{5\mu\nu\rho\sigma\tau} \sim \epsilon_{\mu\nu\rho\sigma\tau\upsilon}\partial^\upsilon H$ and metric of the form

$$ds^2 = H^{-\frac{1}{2}}(\vec{y})dx^\mu dx_\mu + H^{\frac{1}{2}}(\vec{y})d\vec{y}^2$$

Here, x^μ, $\mu = 0, \cdots, 3$ are the coordinates along the brane, while $\vec{y} \in \mathbf{R}^6$ are the coordinates perpendicular to the brane. Show that the sugra equations hold provided H is harmonic in the transverse directions (i.e. satisfies $\Box_y H = 0$, except at the position of the brane, where a pole will occur).

(4.3) Continuing with the set-up of (4.2), show that regularity of the solution requires the poles of H to have integer strength.

(4.4) Show that the D3 brane solution preserves 16 supersymmetries (i.e. half of the total number).

5. The Maldacena AdS/CFT Correspondence

In the preceding sections, we have provided descriptions of $D = 4$, $\mathcal{N} = 4$ super-Yang-Mills theory on the one hand and of D3 branes in supergravity and superstring theory on the other hand. We are now ready to exhibit the *Maldacena or near-horizon limit* close to the D3 branes and formulate precisely the Maldacena or AdS/CFT correspondence which conjectures the identity or duality between $\mathcal{N} = 4$ SYM and Type IIB superstring theory on $AdS_5 \times S^5$. We shall also present the three different forms of the conjecture, the first being a correspondence with the full quantum string theory, the second being with classical string theory and finally the weakest form being with classical supergravity on $AdS_5 \times S^5$. In this section, the precise mapping between both sides of the conjecture will be made for the global symmetries as well as for the fields and operators. The mapping between the correlation functions will be presented in the next section. For a general review see Ref. 7; see also Ref. 44.

5.1. *Non-Abelian Gauge Symmetry on D3 branes*

Open strings whose both end points are attached to a single brane can have arbitrarily short length and must therefore be massless. This excitation mode induces a massless $U(1)$ gauge theory on the worldbrane which is effectively 4-dimensional flat space-time.[45] Since the brane breaks half of the total number of supersymmetries (it is 1/2 BPS), the $U(1)$ gauge theory must have $\mathcal{N}=4$ Poincaré supersymmetry. In the low energy approximation (which has at most two derivatives on bosons and one derivative on fermions in this case), the $\mathcal{N}=4$ supersymmetric $U(1)$ gauge theory is free.

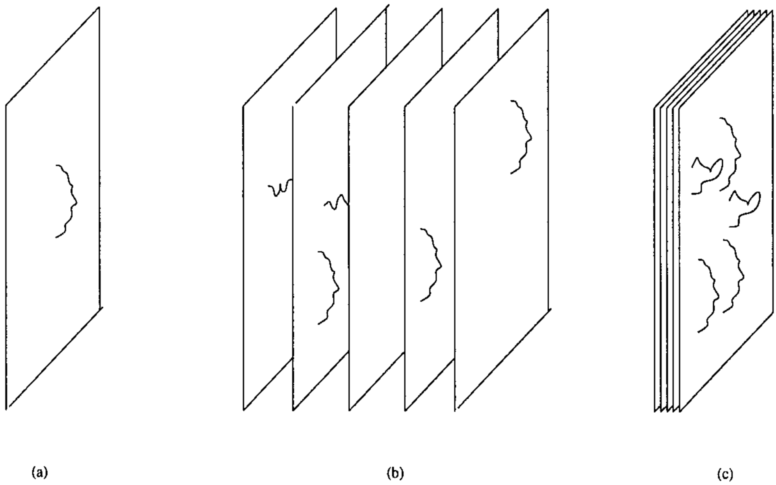


Figure 2. D-branes : (a) single, (b) well-separated, (c) (almost) coincident

With a number $N>1$ of parallel separated D3-branes, the end points of an open string may be attached to the same brane. For each brane, these strings can have arbitrarily small length and must therefore be massless. These excitation modes induce a massless $U(1)^N$ gauge theory with $\mathcal{N}=4$ supersymmetry in the low energy limit. An open string can also, however, have one of its ends attached to one brane while the other end is attached to a different brane. The mass of such a string cannot get arbitrarily small since the length of the string is bounded from below by the separation distance between the branes (see however problem set (5.4)). There are N^2-N such possible strings. In the limit where the N branes all tend to be coincident, all string states would be massless and the $U(1)^N$ gauge symmetry is enhanced to a full $U(N)$ gauge symmetry. Separating the branes should then be interpreted as Higgsing the gauge theory to the Coulomb branch where the gauge symmetry is spontaneously broken (generically to $U(1)^N$). The overall $U(1) = U(N)/SU(N)$ factor actually corresponds to the overall position of the branes and may be ignored when considering dynamics on the branes, thereby leaving only a $SU(N)$ gauge symmetry.[46] These various configurations are depicted in Fig. 2.

In the low energy limit, N coincident branes support an $\mathcal{N} = 4$ super-Yang-Mills theory in 4-dimensions with gauge group $SU(N)$.

5.2. *The Maldacena Limit*

The space-time metric of N coincident D3-branes may be recast in the following form,[iv]

$$ds^2 = \left(1 + \frac{L^4}{y^4}\right)^{-\frac{1}{2}} \eta_{ij} dx^i dx^j + \left(1 + \frac{L^4}{y^4}\right)^{\frac{1}{2}} (dy^2 + y^2 d\Omega_5^2) \tag{101}$$

where the *radius* L *of the D3-brane* is given by

$$L^4 = 4\pi g_s N (\alpha')^2 \tag{102}$$

To study this geometry more closely, we consider its limit in two regimes.

As $y \gg L$, we recover flat space-time $\mathbf{R}^{10}$. When $y < L$, the geometry is often referred to as the *throat* and would at first appear to be singular as $y \ll L$. A redefinition of the coordinate

$$u \equiv L^2/y \tag{103}$$

and the large u limit, however, transform the metric into the following asymptotic form

$$ds^2 = L^2 \left[\frac{1}{u^2} \eta_{ij} dx^i dx^j + \frac{du^2}{u^2} + d\Omega_5^2 \right] \tag{104}$$

which corresponds to a product geometry. One component is the five-sphere S^5 with metric $L^2 d\Omega_5^2$. The remaining component is the hyperbolic space AdS_5 with constant negative curvature metric $L^2 u^{-2}(du^2 + \eta_{ij} dx^i dx^j)$. In conclusion, the geometry close to the brane ($y \sim 0$ or $u \sim \infty$) is regular

[iv] In this section, we shall denote 10-dimensional indices by $M, N, \cdots$, 5-dimensional indices by $\mu, \nu, \cdots$ and 4-dimensional Minkowski indices by $i, j, \cdots$, and the Minkowski metric by $\eta_{ij} = \text{diag}(- + ++)$.

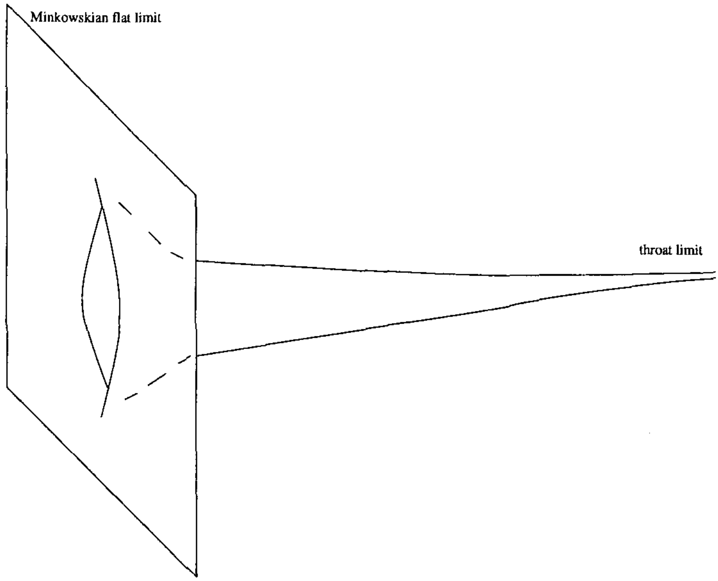


Figure 3. Minkowski region of AdS (a), and throat region of AdS (b)

and highly symmetrical, and may be summarized as $AdS_5 \times S^5$ where both components have identical radius L.

The Maldacena limit[1] corresponds to keeping fixed g_s and N as well as all physical length scales, while letting $\alpha' \to 0$. Remarkably, this limit of string theory exists and is (very !) interesting. In the Maldacena limit, only the $AdS_5 \times S^5$ region of the D3-brane geometry survives the limit and contributes to the string dynamics of physical processes, while the dynamics in the asymptotically flat region decouples from the theory.

To see this decoupling in an elementary way, consider a physical quantity, such as the effective action $\mathcal{L}$ and carry out its α' expansion in an arbitrary background with Riemann tensor, symbolically denoted by R. The expansion takes on the schematic form

$$\mathcal{L} = a_1 \alpha' R + a_2 (\alpha')^2 R^2 + a_3 (\alpha')^3 R^3 + \cdots \tag{105}$$

Now physical objects and length scales in the asymptotically flat region are characterized by a scale $y \gg L$, so that by simple scaling arguments we have $R \sim 1/y^2$. Substitution this behavior into (105) yields the following expansion of the effective action,

$$\mathcal{L} = a_1 \alpha' \frac{1}{y^2} + a_2 (\alpha')^2 \frac{1}{y^4} + a_3 (\alpha')^3 \frac{1}{y^6} + \cdots \tag{106}$$

Keeping the physical size y fixed, the entire contribution to the effective action from the limit $\alpha' \to 0$ is then seen to vanish.

A more precise way of establishing this decoupling is by taking the Maldacena limit directly on the string theory non-linear sigma model in the D3 brane background. We shall concentrate here on the metric part, thereby ignoring the contributions from the tensor field F_5^+. We denote the $D = 10$ coordinates by x^M, $M = 0, 1, \cdots, 9$, and the metric by $G_{MN}(x)$. The first 4 coordinates coincide with x^μ of the Poincaré invariant D3 worldvolume, while the coordinates on the 5-sphere are x^M for $M = 5, \cdots, 9$ and $x^4 = u$. The full D3 brane metric of (101) takes the form $ds^2 = G_{MN} dx^M dx^N = L^2 \bar{G}_{MN}(x; L) dx^M dx^N$, where the rescaled metric $\bar{G}_{MN}$ is given by

$$\bar{G}_{MN}(x; L) dx^M dx^N = \left[1 + \tfrac{L^4}{u^4}\right]^{\frac{1}{2}} \left(\tfrac{du^2}{u^2} + d\Omega_5^2\right) + \left[1 + \tfrac{L^4}{u^4}\right]^{-\frac{1}{2}} \tfrac{1}{u^2} \eta_{ij} dx^i dx^j \tag{107}$$

Inserting this metric into the non-linear sigma model, we obtain

$$\begin{aligned} S_G &= \tfrac{1}{4\pi\alpha'} \int_\Sigma \sqrt{\gamma} \gamma^{mn} G_{MN}(x) \partial_m x^M \partial_n x^N \\ &= \tfrac{L^2}{4\pi\alpha'} \int_\Sigma \sqrt{\gamma} \gamma^{mn} \bar{G}_{MN}(x; L) \partial_m x^M \partial_n x^N \end{aligned} \tag{108}$$

The overall coupling constant for the sigma model dynamics is given by

$$\frac{L^2}{4\pi\alpha'} = \sqrt{\frac{\lambda}{4\pi}} \qquad \lambda \equiv g_s N \tag{109}$$

Keeping g_s and N fixed but letting $\alpha' \to 0$ implies that $L \to 0$. Under this limit the sigma model action admits a smooth limit, given by

$$S_G = \sqrt{\frac{\lambda}{4\pi}} \int_\Sigma \sqrt{\gamma} \gamma^{mn} \bar{G}_{MN}(x; 0) \partial_m x^M \partial_n x^N \tag{110}$$

where the metric $\bar{G}_{MN}(x; 0)$ is the metric on $\mathrm{AdS}_5 \times \mathrm{S}^5$,

$$\bar{G}_{MN}(x; L) dx^M dx^N = \frac{1}{u^2} \eta_{ij} dx^i dx^j + \frac{du^2}{u^2} + d\Omega_5^2 \tag{111}$$

rescaled to unit radius. Manifestly, the coupling $1/\sqrt{\lambda}$ has taken over the role of α' as the non-linear sigma model coupling constant and the radius L has cancelled out.

5.3. *Geometry of Minkowskian and Euclidean AdS*

Before moving on to the actual Maldacena conjecture, we clarify the geometry of AdS space-time, both with Minkowskian and Euclidean signatures. Minkowskian AdS_{d+1} (of unit radius) may be defined in $\mathbf{R}^{d+1}$ with coordinates $(Y_{-1}, Y_0, Y_1, \cdots, Y_d)$ as the $d+1$ dimensional connected hyperboloid with isometry $SO(2,d)$ given by the equation

$$-Y_{-1}^2 - Y_0^2 + Y_1^2 + \cdots + Y_d^2 = -1 \tag{112}$$

with induced metric $ds^2 = -dY_{-1}^2 - dY_0^2 + dY_1^2 + \cdots + dY_d^2$. The topology of the manifold is that of the cylinder $S^1 \times \mathbf{R}$ times the sphere S^{d-1}, and is therefore not simply connected. The topology of the boundary is consequently given by $\partial\mathrm{AdS}_{d+1} = S^1 \times S^{d-1}$. The manifold may be represented by the coset $SO(2,d)/SO(1,d)$. A schematic rendition of the manifold is given in Fig. 4 (a), with $r^2 = Y_1^2 + \cdots + Y_d^2$.

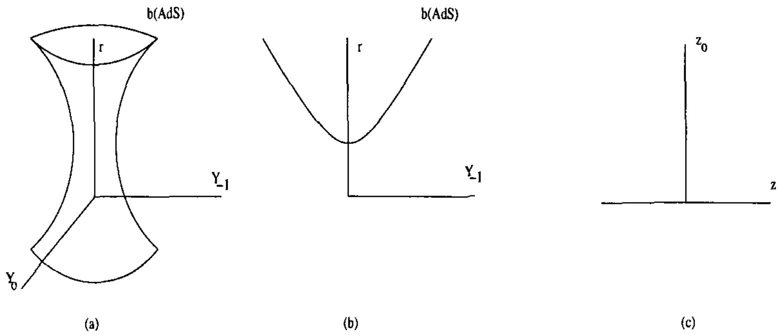


Figure 4. Anti-de Sitter Space (a) Euclidean, (b) Minkowskian, (c) upper half space

Euclidean AdS_{d+1} (of unit radius) may be defined in Minkowski flat space $\mathbf{R}^{d+1}$ with coordinates $(Y_{-1}, Y_0, Y_1, \cdots, Y_d)$ as the $d+1$ dimensional disconnected hyperboloid with isometry $SO(1,d)$ given by the equation

$$-Y_{-1}^2 + Y_0^2 + Y_1^2 + \cdots + Y_d^2 = -1 \tag{113}$$

with induced metric $ds^2 = -dY_{-1}^2 + dY_0^2 + dY_1^2 + \cdots + dY_d^2$. The topology of the manifold is that of $\mathbf{R}^{d+1}$. The topology of the boundary is that of the d-sphere, $\partial\mathrm{AdS}_{d+1} = S^d$. The manifold may be represented by the coset $SO(1,d+1)/SO(d+1)$. A schematic rendition of the manifold is given in Fig. 4 (b), with $r^2 = Y_0^2 + Y_1^2 + \cdots + Y_d^2$. Introducing the coordinates $Y_{-1} + Y_0 = \frac{1}{z_0}$ and $z_i = z_0 Y_i$ for $i = 1, \cdots, d$, we may map Euclidean

AdS_{d+1} onto the upper half space H_{d+1} with Poincaré metric ds^2, defined by

$$H_{d+1} = \{(z_0, \vec{z}),\ z_0 \in \mathbf{R}^+, \vec{z} \in \mathbf{R}^d\} \qquad ds^2 = \frac{1}{z_0^2}(dz_0^2 + d\vec{z}^2) \quad (114)$$

A schematic rendition is given in Fig. 4 (c). A standard stereographic transformation may be used to map H_{d+1} onto the unit ball.

5.4. *The AdS/CFT Conjecture*

The AdS/CFT or Maldacena conjecture states the *equivalence* (also referred to as *duality*) between the following theories[1]

- Type IIB superstring theory on $\mathrm{AdS}_5 \times \mathrm{S}^5$ where both AdS_5 and S^5 have the same radius L, where the 5-form F_5^+ has integer flux $N = \int_{S^5} F_5^+$ and where the string coupling is g_s;
- $\mathcal{N} = 4$ super-Yang-Mills theory in 4-dimensions, with gauge group $SU(N)$ and Yang-Mills coupling g_{YM} in its (super)conformal phase;

with the following identifications between the parameters of both theories,

$$g_s = g_{YM}^2 \qquad\qquad L^4 = 4\pi g_s N(\alpha')^2 \qquad (115)$$

and the axion expectation value equals the SYM instanton angle $\langle C \rangle = \theta_I$. Precisely what is meant by *equivalence* or *duality* will be the subject of the remainder of this section, as well as of the next one. In brief, *equivalence* includes a precise map between the states (and fields) on the superstring side and the local gauge invariant operators on the $\mathcal{N} = 4$ SYM side, as well as a correspondence between the correlators in both theories.

The above statement of the conjecture is referred to as the *strong form*, as it is to hold for all values of N and of $g_s = g_{YM}^2$. String theory quantization on a general curved manifold (including $\mathrm{AdS}_5 \times \mathrm{S}^5$), however, appears to be very difficult and is at present out of reach. Therefore, it is natural to seek limits in which the Maldacena conjecture becomes more tractable but still remains non-trivial.

5.4.1. *The 't Hooft Limit*

The 't Hooft limit consists in keeping the 't Hooft coupling $\lambda \equiv g_{YM}^2 N = g_s N$ fixed and letting $N \to \infty$. In Yang-Mills theory, this limit is well-defined, at least in perturbation theory, and corresponds to a topological expansion of the field theory's Feynman diagrams. On the AdS side, one

Table 5. The three forms of the AdS/CFT conjecture in order of decreasing strength

• $\mathcal{N}$=4 conformal SYM all N, g_{YM} • $g_s = g_{YM}^2$	$\Leftrightarrow$	• Full Quantum Type IIB string theory on $AdS_5 \times S^5$ • $L^4 = 4\pi g_s N \alpha'^2$
• 't Hooft limit of $\mathcal{N}$=4 SYM $\lambda = g_{YM}^2 N$ fixed, $N \to \infty$ • $1/N$ expansion	$\Leftrightarrow$	• Classical Type IIB string theory on $AdS_5 \times S^5$ • g_s string loop expansion
• Large λ limit of $\mathcal{N}$=4 SYM (for $N \to \infty$) • $\lambda^{-1/2}$ expansion	$\Leftrightarrow$	• Classical Type IIB supergravity on $AdS_5 \times S^5$ • α' expansion

may interpret the 't Hooft limit as follows. The string coupling may be re-expressed in terms of the 't Hooft coupling as $g_s = \lambda/N$. Since λ is being kept fixed, the 't Hooft limit corresponds to *weak coupling string perturbation theory.*

This form of the conjecture, though weaker than the original version is still a very powerful correspondence between classical string theory and the large N limit of gauge theories. The problem of finding an action built out of classical fields to which the large N limit of gauge theories are classical solutions is a challenge that had been outstanding since 't Hooft's original paper.[8] The above correspondence gives a concrete, though still ill-understood, realization of this "large N master-equation".

5.4.2. *The Large λ Limit*

In taking the 't Hooft limit, $\lambda = g_s N$ is kept fixed while $N \to \infty$. Once this limit has been taken, the only parameter left is λ. Quantum field theory perturbation theory corresponds to $\lambda \ll 1$. On the AdS side of the correspondence, it is actually natural to take $\lambda \gg 1$ instead. It is very instructive to establish the meaning of an expansion around λ large. To do, we expand in powers of α' a physical quantity such as the effective action, as we already did in (105),

$$\mathcal{L} = a_1 \alpha' R + a_2 (\alpha')^2 R^2 + a_3 (\alpha')^3 R^3 + \cdots \tag{116}$$

The distance scales in which we are now interested are those typical of the throat, whose scale is set by the AdS radius L. Thus, the scale of the Riemann tensor is set by

$$R \sim 1/L^2 = (g_s N)^{-\frac{1}{2}}/\alpha' = \lambda^{-\frac{1}{2}}/\alpha' \tag{117}$$

and therefore, the expansion of the effective action in powers of α' effectively becomes an expansion in powers of $\lambda^{-\frac{1}{2}}$,

$$\mathcal{L} = a_1\lambda^{-\frac{1}{2}} + a_2\lambda^{-1} + a_3(\alpha')^3\lambda^{-\frac{3}{2}} + \cdots \tag{118}$$

The interchange of the roles of α' and $\lambda^{-1/2}$ may also be seen directly from the worldsheet non-linear sigma model action of (110). Clearly, any α' dependence has disappeared from the string theory problem and the role of α' as a scale has been replaced by the parameter $\lambda^{-1/2}$.

5.5. *Mapping Global Symmetries*

A key necessary ingredient for the AdS/CFT correspondence to hold is that the global unbroken symmetries of the two theories be identical. The continuous global symmetry of $\mathcal{N} = 4$ super-Yang-Mills theory in its conformal phase was previously shown to be the superconformal group $SU(2,2|4)$, whose maximal bosonic subgroup is $SU(2,2) \times SU(4)_R \sim SO(2,4) \times SO(6)_R$. Recall that the bosonic subgroup arises as the product of the conformal group $SO(2,4)$ in 4-dimensions by the $SU(4)_R$ automorphism group of the $\mathcal{N} = 4$ Poincaré supersymmetry algebra. This bosonic group is immediately recognized on the AdS side as the isometry group of the $\mathrm{AdS}_5 \times \mathrm{S}^5$ background. The completion into the full supergroup $SU(2,2|4)$ was discussed for the SYM theory in subsection §3.3, and arises on the AdS side because 16 of the 32 Poincaré supersymmetries are preserved by the array of N parallel D3-branes, and in the AdS limit, are supplemented by another 16 conformal supersymmetries (which are broken in the full D3-brane geometry). Thus, the global symmetry $SU(2,2|4)$ matches on both sides of the AdS/CFT correspondence.

$\mathcal{N} = 4$ super-Yang-Mills theory also has Montonen-Olive or S-duality symmetry, realized on the complex coupling constant τ by Möbius transformations in $SL(2,\mathbf{Z})$. On the AdS side, this symmetry is a global discrete symmetry of Type IIB string theory, which is unbroken by the D3-brane solution, in the sense that it maps non-trivially only the dilaton and axion expectation values, as was shown earlier. Thus, S-duality is also a symmetry of the AdS side of the AdS/CFT correspondence. It must be noted, however, that S-duality is a useful symmetry only in the strongest form of the AdS/CFT conjecture. As soon as one takes the 't Hooft limit $N \to \infty$ while keeping $\lambda = g_{YM}^2 N$ fixed, S-duality no longer has a consistent action. This may be seen for $\theta_I = 0$, where it maps $g_{YM} \to 1/g_{YM}$ and thus $\lambda \to N^2/\lambda$.

5.6. *Mapping Type IIB Fields and CFT Operators*

Given that we have established that the global symmetry groups on both sides of the AdS/CFT correspondence coincide, it remains to show that the actual representations of the supergroup $SU(2,2|4)$ also coincide on both sides. The spectrum of operators on the SYM side was explained already in subsection §3.5. Suffice it to recall here the special significance of the short multiplet representations, namely 1/2 BPS representations with a span of spin 2, 1/4 BPS representations with a span of spin 3 and 1/8 BPS representations with a span of spin 7/2. Non-BPS representations in general have a span of spin 4.

A special role is played by the *single color trace operators* because out of them, all higher trace operators may be constructed using the OPE. Thus one should expect single trace operators on the SYM side to correspond to single particle states (or canonical fields) on the AdS side;[1] see also Ref. 47. Multiple trace states should then be interpreted as bound states of these one particle states. Multiple trace BPS operators have the property that their dimension on the AdS side is simply the sum of the dimensions of the BPS constituents. Such bound states occur in the spectrum at the lower edge of the continuum threshold and are therefore called *threshold bound states*. A good example to keep in mind when thinking of threshold bound states in ordinary quantum field theory is another case of BPS objects : magnetic monopoles[48] in the Prasad-Sommerfield limit[14] (or exactly in the Coulomb phase of $\mathcal{N}=4$ SYM). A collection of N magnetic monopoles with like charges forms a static solution of the BPS equations and therefore form a threshold bound state. Very recently, a direct coupling of double-trace operators to AdS supergravity has been studied in Ref. 49.

To identify the contents of irreducible representations of $SU(2,2|4)$ on the AdS side, we describe all Type IIB massless supergravity and massive string degrees of freedom by fields φ living on $\mathrm{AdS}_5 \times \mathrm{S}^5$. We introduce coordinates z^μ, $\mu=0,1,\cdots,4$ for AdS_5 and y^u, $u=1,\cdots,5$ for S^5, and decompose the metric as

$$ds^2 = g_{\mu\nu}{}^{AdS} dz^\mu dz^\nu + g^S_{uv} dy^u dy^v \tag{119}$$

The fields then become functions $\varphi(z,y)$ associated with the various $D=10$ degrees of freedom. It is convenient to decompose $\varphi(z,y)$ in a series on S^5,

$$\varphi(z,y) = \sum_{\Delta=0}^{\infty} \varphi_\Delta(z) Y_\Delta(y) \tag{120}$$

Table 6. Mapping of String and Sugra states onto SYM Operators

Type IIB string theory	$\mathcal{N}=4$ conformal super-Yang-Mills
Supergravity Excitations 1/2 BPS, spin ≤ 2	Chiral primary + descendants $\mathcal{O}_2 = \mathrm{tr}X^{\{i}X^{j\}}$ + desc.
Supergravity Kaluza-Klein 1/2 BPS, spin ≤ 2	Chiral primary + Descendants $\mathcal{O}_\Delta = \mathrm{tr}X^{\{i_1}\cdots X^{i_\Delta\}}$ + desc.
Type IIB massive string modes non-chiral, long multiplets	Non-Chiral operators, dims $\sim \lambda^{1/4}$ e.g. Konishi $\mathrm{tr}X^iX^i$
Multiparticle states	products of ops at distinct points $\mathcal{O}_{\Delta_1}(x_1)\cdots\mathcal{O}_{\Delta_n}(x_n)$
Bound states	product of operators at same point $\mathcal{O}_{\Delta_1}(x)\cdots\mathcal{O}_{\Delta_n}(x)$

where Y_Δ stands for a basis of spherical harmonics on S^5. For scalars for example, Y_Δ are labelled by the rank Δ of the totally symmetric traceless representations of $SO(6)$. Just as fields on a circle received a mass contribution from the momentum mode on the circle, so also do fields compactified on S^5 receive a contribution to the mass. From the eigenvalues of the Laplacian on S^5, for various spins, we find the following relations between mass and scaling dimensions,

$$\begin{aligned} &\text{scalars} && m^2 = \Delta(\Delta-4) \\ &\text{spin } 1/2,\ 3/2 && |m| = \Delta - 2 \\ &p-\text{form} && m^2 = (\Delta-p)(\Delta+p-4) \\ &\text{spin } 2 && m^2 = \Delta(\Delta-4) \end{aligned} \tag{121}$$

The complete correspondence between the representations of $SU(2,2|4)$ on both sides of the correspondence is given in Table 6. The mapping of the descendant states is also very interesting. For the $D=10$ supergravity multiplet, this was worked out in Ref. 50, and is given in Table 7. Generalizations to $\mathrm{AdS}_4 \times S^7$ were discussed in Refs. 51, 52, 53 while those to $\mathrm{AdS}_7 \times S^4$ were discussed in Refs. 54, 55, with recent work on AdS/CFT for M-theory on these spaces in Refs. 56, 57, 58, 59. General reviews may be found in Refs. 61, 60. Recently, conjectures involving also de Sitter spacetimes have been put forward in Ref. 62 and references therein. Finally, we point out that the existence of singleton and doubleton representations of the conformal group SO(2,4) is closely related with the AdS/CFT correspondence; for recent accounts, see Refs. 65 63, 64 and 66, and references therein. Additional references on the (super)symmetries of AdS are in Refs. 68 67, and 69.

Table 7. Super-Yang-Mills Operators, Supergravity Fields and their Quantum Numbers under $SO(2,4) \times U(1)_Y \times SU(4)_R$. The range of k is $k \geq 0$, unless otherwise specified.

SYM Operator	desc	SUGRA	dim	spin	Y	$SU(4)_R$	lowest reps
$\mathcal{O}_k \sim \mathrm{tr} X^k,\ k \geq 2$	–	$h^\alpha_\alpha \quad a_{\alpha\beta\gamma\delta}$	k	$(0,0)$	0	$(0,k,0)$	**20',50,105**
$\mathcal{O}_k^{(1)} \sim \mathrm{tr}\lambda X^k,\ k \geq 1$	Q	$\psi_{(\alpha)}$	$k+\frac{3}{2}$	$(\frac{1}{2},0)$	$\frac{1}{2}$	$(1,k,0)$	**20,60,140'**
$\mathcal{O}_k^{(2)} \sim \mathrm{tr}\lambda\lambda X^k$	Q^2	$A_{\alpha\beta}$	$k+3$	$(0,0)$	1	$(2,k,0)$	$\mathbf{10}_c,\mathbf{45}_c,\mathbf{126}_c$
$\mathcal{O}_k^{(3)} \sim \mathrm{tr}\lambda\bar\lambda X^k$	$Q\bar Q$	$h_{\mu\alpha} \quad a_{\mu\alpha\beta\gamma}$	$k+3$	$(\frac{1}{2},\frac{1}{2})$	0	$(1,k,1)$	**15,64,175**
$\mathcal{O}_k^{(4)} \sim \mathrm{tr} F_+ X^k,\ k \geq 1$	Q^2	$A_{\mu\nu}$	$k+2$	$(1,0)$	1	$(0,k,0)$	$\mathbf{6}_c,\mathbf{20}_c,\mathbf{50}_c$
$\mathcal{O}_k^{(5)} \sim \mathrm{tr} F_+\bar\lambda X^k$	$Q^2\bar Q$	ψ_μ	$k+\frac{7}{2}$	$(1,\frac{1}{2})$	$\frac{1}{2}$	$(0,k,1)$	**4*, 20*, 60***
$\mathcal{O}_k^{(6)} \sim \mathrm{tr} F_+\lambda X^k$	Q^3	"λ"	$k+\frac{7}{2}$	$(\frac{1}{2},0)$	$\frac{3}{2}$	$(1,k,0)$	**4,20,60**
$\mathcal{O}_k^{(7)} \sim \mathrm{tr}\lambda\lambda\bar\lambda X^k$	$Q^2\bar Q$	$\psi_{(\alpha)}$	$k+\frac{9}{2}$	$(0,\frac{1}{2})$	$\frac{1}{2}$	$(2,k,1)$	**36,140,360**
$\mathcal{O}_k^{(8)} \sim \mathrm{tr} F_+^2 X^k$	Q^4	B	$k+4$	$(0,0)$	2	$(0,k,0)$	$\mathbf{1}_c,\mathbf{6}_c,\mathbf{20'}_c$
$\mathcal{O}_k^{(9)} \sim \mathrm{tr} F_+ F_- X^k$	$Q^2\bar Q^2$	$h'_{\mu\nu}$	$k+4$	$(1,1)$	0	$(0,k,0)$	**1,6,20'**
$\mathcal{O}_k^{(10)} \sim \mathrm{tr} F_+\lambda\bar\lambda X^k$	$Q^3\bar Q$	$A_{\mu\alpha}$	$k+5$	$(\frac{1}{2},\frac{1}{2})$	1	$(1,k,1)$	**15,64,175**
$\mathcal{O}_k^{(11)} \sim \mathrm{tr} F_+\bar\lambda\bar\lambda X^k$	$Q^2\bar Q^2$	$a_{\mu\nu\alpha\beta}$	$k+5$	$(1,0)$	0	$(0,k,2)$	$\mathbf{10}_c,\mathbf{45}_c,\mathbf{126}_c$
$\mathcal{O}_k^{(12)} \sim \mathrm{tr}\lambda\lambda\bar\lambda\bar\lambda X^k$	$Q^2\bar Q^2$	$h_{(\alpha\beta)}$	$k+6$	$(0,0)$	0	$(2,k,2)$	**84,300,2187**
$\mathcal{O}_k^{(13)} \sim \mathrm{tr} F_+^2\bar\lambda X^k$	$Q^4\bar Q$	"λ"	$k+\frac{11}{2}$	$(0,\frac{1}{2})$	$\frac{3}{2}$	$(0,k,1)$	**4*, 20*, 60***
$\mathcal{O}_k^{(14)} \sim \mathrm{tr} F_+\lambda\bar\lambda\bar\lambda X^k$	$Q^3\bar Q^2$	$\psi_{(\alpha)}$	$k+\frac{13}{2}$	$(\frac{1}{2},0)$	$\frac{1}{2}$	$(1,k,2)$	**36*,140*,360***
$\mathcal{O}_k^{(15)} \sim \mathrm{tr} F_+ F_-\lambda X^k$	$Q^3\bar Q^2$	ψ_μ	$k+\frac{11}{2}$	$(\frac{1}{2},1)$	$\frac{1}{2}$	$(1,k,0)$	**4,20,60**
$\mathcal{O}_k^{(16)} \sim \mathrm{tr} F_+ F_-^2 X^k$	$Q^4\bar Q^2$	$A_{\mu\nu}$	$k+6$	$(1,0)$	1	$(0,k,0)$	$\mathbf{1}_c,\mathbf{6}_c,\mathbf{20'}_c$
$\mathcal{O}_k^{(17)} \sim \mathrm{tr} F_+ F_-\lambda\bar\lambda X^k$	$Q^3\bar Q^3$	$h_{\mu\alpha} \quad a_{\mu\alpha\beta\gamma}$	$k+7$	$(\frac{1}{2},\frac{1}{2})$	0	$(1,k,1)$	**15,64,175**
$\mathcal{O}_k^{(18)} \sim \mathrm{tr} F_+^2\bar\lambda\bar\lambda X^k$	$Q^4\bar Q^2$	$A_{\alpha\beta}$	$k+7$	$(0,0)$	1	$(0,k,2)$	$\mathbf{10}_c,\mathbf{45}_c,\mathbf{126}_c$
$\mathcal{O}_k^{(19)} \sim \mathrm{tr} F_+^2 F_-\bar\lambda X^k$	$Q^4\bar Q^3$	$\psi_{(\alpha)}$	$k+\frac{15}{2}$	$(0,\frac{1}{2})$	$\frac{1}{2}$	$(0,k,1)$	**4*, 20*, 60***
$\mathcal{O}_k^{(20)} \sim \mathrm{tr} F_+^2 F_-^2 X^k$	$Q^4\bar Q^4$	$h^\alpha_\alpha \quad a_{\alpha\beta\gamma\delta}$	$k+8$	$(0,0)$	0	$(0,k,0)$	**1,6,20'**

5.7. *Problem Sets*

(5.1) The Poincaré upper half space is defined by $H_{d+1} = \{(z_0, \vec z) \in \mathbf{R}^{d+1}, z_0 > 0\}$ with metric $ds^2 = (dz_0^2 + d\vec z^{\,2})/z_0^2$. (a) Show – by solving the geodesic equations – that the geodesics of H_{d+1} are the half-circles of arbitrary radius R, centered at an arbitrary point $(0, \vec c)$ on the boundary of H_{d+1}. Compute the geodesic distance between any two arbitrary points.

(5.2) We now represent Euclidean AdS$_{d+1}$ as the manifold in $\mathbf{R}^{d+2}$ given by the equation $-Y_{-1}^2 + Y_0^2 + \vec Y^2 = -1$, with induced metric $ds^2 = -dY_{-1}^2 + dY_0^2 + d\vec Y^2$. Show that the geodesics found in problem (5.1) above are simply the sections by planes through the origin, given by the equation

$$Y_{-1} - Y_0 = (R^2 - \vec c^{\,2})(Y_{-1} + Y_0) + 2\vec c \cdot \vec Y$$

(You may wish to explore the analogy with the geometry and geodesics of the sphere S^{d+1}.)

(5.3) The geodesic distance between two separate D3 branes is actually infinite, as may be seen by integrating the infinitesimal distance ds of the D3 metric. Using the worldsheet action of a string suspended between the two D3 branes, explain why this string still has a finite mass spectrum.

(5.4) Consider a classical bosonic string in AdS_{d+1} space-time, with its dynamics governed by the Polyakov action, namely in the presence of the AdS_{d+1} metric $G_{\mu\nu}(x)$. (We ignore the anti-symmetric tensor fields for simplicity.)

$$S[x] = \int_\Sigma d^2\xi \sqrt{\gamma}\gamma^{mn}\partial_m x^\mu \partial_n x^\nu G_{\mu\nu}(x)$$

Solve the string equations assuming a special Ansatz that the solution be spherically symmetric, i.e. invariant under the $SO(d)$ subgroup of $SO(2,d)$.

6. AdS/CFT Correlation Functions

In the preceding section, evidence was presented for the Maldacena correspondence between $\mathcal{N}=4$ super-conformal Yang-Mills theory with $SU(N)$ gauge group and Type IIB superstring theory on $\mathrm{AdS}_5 \times \mathrm{S}^5$. The evidence was based on the precise matching of the global symmetry group $SU(2,2|4)$, as well as of the specific representations of this group. In particular, the single trace 1/2 BPS operators in the SYM theory matched in a one-to-one way with the canonical fields of supergravity, compactified on $\mathrm{AdS}_5 \times \mathrm{S}^5$. In the present section, we present a more detailed version of the AdS/CFT correspondence by mapping the correlators on both sides of the correspondence.

6.1. *Mapping Super Yang-Mills and AdS Correlators*

We work with Euclidean AdS_5, or $H = \{(z_0, \vec{z}), z_0 > 0, \vec{z} \in \mathbf{R}^4\}$ with Poincaré metric $ds^2 = z_0^{-2}(dz_0^2 + d\vec{z}^2)$, and boundary $\partial H = \mathbf{R}^4$. (Often, this space will be graphically represented as a disc, whose boundary is a circle; see Fig. 5.) The metric diverges at the boundary $z_0 = 0$, because the overall scale factor blows up there. This scale factor may be removed by a Weyl rescaling of the metric, but such rescaling is not unique. *A unique well-defined limit to the boundary of AdS_5 can only exist if the boundary theory is scale invariant.*[3] For finite values of $z_0 > 0$, the geometry will still have 4-dimensional Poincaré invariance but need not be scale invariant.

Superconformal $\mathcal{N}=4$ Yang-Mills theory is scale invariant and may thus consistently live at the boundary ∂H. The dynamical observables of $\mathcal{N}=4$ SYM are the local gauge invariant polynomial operators described in section 3; they naturally live on the boundary ∂H, and are characterized by their dimension, Lorentz group $SO(1,3)$ and $SU(4)_R$ quantum numbers.[3]

On the AdS side, we shall decompose all 10-dimensional fields onto Kaluza-Klein towers on S^5, so that effectively all fields $\varphi_\Delta(z)$ are on AdS_5, and labeled by their dimension Δ (other quantum number are implicit). Away from the bulk interaction region, it is assumed that the bulk fields are free asymptotically (just as this is assumed in the derivation of the LSZ formalism in flat space-time quantum field theory). The free field then satisfies $(\Box + m_\Delta^2)\varphi_\Delta^0 = 0$ with $m_\Delta^2 = \Delta(\Delta - 4)$ for scalars. The two independent solutions are characterized by the following asymptotics as $z_0 \to 0$,

$$\varphi_\Delta^0(z_0, \vec{z}) = \begin{cases} z_0^\Delta & \text{normalizable} \\ \\ z_0^{4-\Delta} & \text{non-normalizable} \end{cases} \tag{122}$$

Returning to the interacting fields in the fully interacting theory, solutions will have the same asymptotic behaviors as in the free case. It was argued in Ref. 70 that the normalizable modes determine the vacuum expectation values of operators of associated dimensions and quantum numbers. The non-normalizable solutions on the other hand do not correspond to bulk excitations because they are not properly square normalizable. Instead, they represent the coupling of external sources to the supergravity or string theory. The precise correspondence is as follows.[3] The non-normalizable solutions φ_Δ define *associated boundary fields* $\bar{\varphi}_\Delta$ by the following relation

$$\bar{\varphi}_\Delta(\vec{z}) \equiv \lim_{z_0 \to 0} \varphi_\Delta(z_0, \vec{z}) z_0^{4-\Delta} \tag{123}$$

Given a set of boundary fields $\bar{\varphi}_\Delta(\vec{z})$, it is assumed that a complete and unique bulk solution to string theory exists. We denote the fields of the associated solution φ_Δ.

The mapping between the correlators in the SYM theory and the dynamics of string theory is given as follows.[3,2] First, we introduce a generating functional $\Gamma[\bar{\varphi}_\Delta]$ for all the correlators of single trace operators $\mathcal{O}_\Delta$ on the SYM side in terms of the source fields $\bar{\varphi}_\Delta$,

$$\exp\{-\Gamma[\bar{\varphi}_\Delta]\} \equiv \langle \exp\left\{ \int_{\partial H} \bar{\varphi}_\Delta \mathcal{O}_\Delta \right\} \rangle \tag{124}$$

This expression is understood to hold order by order in a perturbative expansion in the number of fields $\bar{\varphi}_\Delta$. On the AdS side, we assume that we have an action $S[\varphi_\Delta]$ that summarizes the dynamics of Type IIB string theory on $\mathrm{AdS}_5 \times \mathrm{S}^5$. In the supergravity approximation, $S[\varphi_\Delta]$ is just the Type IIB supergravity action on $\mathrm{AdS}_5 \times \mathrm{S}^5$. Beyond the supergravity approximation, $S[\varphi_\Delta]$ will also include α' corrections due to massive string effects. The mapping between the correlators is given by

$$\Gamma[\bar\varphi_\Delta] = \text{extr}\ S[\varphi_\Delta] \tag{125}$$

where the extremum on the rhs is taken over all fields φ_Δ that satisfy the asymptotic behavior (123) for the boundary fields $\bar\varphi_\Delta$ that are the sources to the SYM operators $\mathcal{O}_\Delta$ on the lhs. Additional references on the field-state-operator mapping may be found in Refs. 74, 71, 72, 73, 76, and 75.

6.2. *Quantum Expansion in 1/N – Witten Diagrams*

The actions of interest to us will have an overall coupling constant factor. For example, the part of the Type IIB supergravity action for the dilaton Φ and the axion C in the presence of a metric $G_{\mu\nu}$ in the Einstein frame, is given by

$$S[G,\Phi,C] = \frac{1}{2\kappa_5^2}\int_H \sqrt{G}\Big[-R_G + \Lambda + \tfrac{1}{2}\partial_\mu\Phi\partial^\mu\Phi + \tfrac{1}{2}e^{2\Phi}\partial_\mu C\partial^\mu C\Big] \tag{126}$$

and the 5-dimensional Newton constant κ_5^2 is given by $\kappa_5^2 = 4\pi^2/N^2$, a relation that will be explained and justified in (239). For large N, κ_5 will be small and one may perform a small κ_5, i.e. a semi-classical expansion of the correlators generated by this action. The result is a set of rules, analogous to Feynman rules, which may be summarized by *Witten diagrams*. The Witten diagram is represented by a disc, whose interior corresponds to the interior of AdS while the boundary circle corresponds to the boundary of AdS.[3] The graphical rules are as follows,

- Each external source to $\bar\varphi_\Delta(\vec x_I)$ is located at the boundary circle of the Witten diagram at a point $\vec x_I$.
- From the external source at $\vec x_I$ departs a propagator to either another boundary point, or to an interior interaction point via a *boundary-to-bulk propagator.*
- The structure of the interior interaction points is governed by the interaction vertices of the action S, just as in Feynman diagrams.
- Two interior interaction points may be connected by *bulk-to-bulk propagators*, again following the rules of ordinary Feynman diagrams.

Tree-level 2-, 3- and 4-point function contributions are given as an example in Fig. 5. The approach that will be taken here is based on the component formulation of sugra. It is possible however to make progress directly in superspace,[77] but we shall not discuss this here.

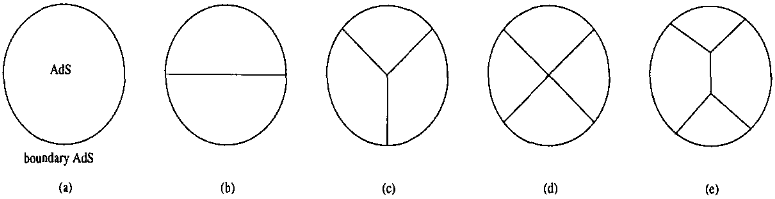


Figure 5. Witten diagrams (a) empty, (b) 2-pt, (c) 3-pt, (d) 4-pt contact, (e) exchange

6.3. *AdS Propagators*

We shall define and list the solution for the propagators of general scalar fields, of massless gauge fields and massless gravitons. The propagators are considered in Euclidean AdS_{d+1}, a space that we shall denote by H. Recall that the Poincaré metric is given by

$$ds^2 = g_{\mu\nu}dz^\mu dz^\nu = z_0^{-2}(dz_0^2 + d\vec{z}^2) \tag{127}$$

Here, we have set the AdS_{d+1} radius to unity. By $SO(1, d+1)$ isometry of H, the Green functions essentially depend upon the $SO(1, d+1)$-invariant distance between two points in H. The *geodesic distance* is given by (see problem 5.1)

$$d(z, w) = \int_w^z ds = \ln\left(\frac{1 + \sqrt{1 - \xi^2}}{\xi}\right) \qquad \xi \equiv \frac{2z_0 w_0}{z_0^2 + w_0^2 + (\vec{z} - \vec{w})^2} \tag{128}$$

Given its algebraic dependence on the coordinates, it is more convenient to work with the object ξ than with the geodesic distance. The *chordal distance* is given by $u = \xi^{-1} - 1$. The distance relation may be inverted to give $u = \xi^{-1} - 1 = \cosh d - 1$.

The massive scalar bulk-to-bulk propagator

Let $\varphi_\Delta(z)$ be a scalar field of conformal weight Δ and mass^2 $m^2 = \Delta(\Delta - d)$ whose linearized dynamics is given by a coupling to a scalar source J via the action

$$S_{\varphi_\Delta} = \int_H d^{d+1}z\sqrt{g}\left[\frac{1}{2}g^{\mu\nu}\partial_\mu\varphi_\Delta\partial_\nu\varphi_\Delta + \frac{1}{2}m^2\varphi_\Delta^2 - \varphi_\Delta J\right] \tag{129}$$

The field is then given in response to the source by

$$\varphi_\Delta(z) = \int_H d^{d+1}z'\sqrt{g}G_\Delta(z, z')J(z') \tag{130}$$

where the scalar Green function satisfies the differential equation

$$(\Box_g + m^2)G_\Delta(z,z') = \delta(z,z') \qquad \delta(z,z') \equiv \frac{1}{\sqrt{g}}\delta(z-z') \tag{131}$$

The (positive) scalar Laplacian is given by

$$\Box_g = -\frac{1}{\sqrt{g}}\partial_\mu \sqrt{g} g^{\mu\nu}\partial_\nu = -z_0^2\partial_0^2 + (d-1)z_0\partial_0 - z_0^2\sum_{i=1}^{d}\partial_i^2 \tag{132}$$

The scalar Green function is the solution to a hypergeometric equation, given by Ref. 84[v]

$$G_\Delta(z,w) = G_\Delta(\xi) = \frac{2^{-\Delta}C_\Delta}{2\Delta - d}\xi^\Delta F\left(\frac{\Delta}{2}, \frac{\Delta}{2}+\frac{1}{2}; \Delta - \frac{d}{2}+1; \xi^2\right) \tag{133}$$

where the overall normalization constant is defined by

$$C_\Delta = \frac{\Gamma(\Delta)}{\pi^{d/2}\Gamma(\Delta - \frac{d}{2})} \tag{134}$$

Since $0 \leq \xi \leq 1$, the hypergeometric function is defined by its convergent Taylor series for all ξ except at the coincident point $\xi = 1$ where $z = w$.

The massive scalar boundary-to-bulk propagator

An important limiting case of the scalar bulk-to-bulk propagator is when the source is on the boundary of H. The action to linearized order is given by

$$S_{\varphi_\Delta} = \int_H d^{d+1}z\sqrt{g}\left[\tfrac{1}{2}g^{\mu\nu}\partial_\mu\varphi_\Delta\partial_\nu\varphi_\Delta + \tfrac{1}{2}m^2\varphi_\Delta^2\right] - \int_{\partial H} d^d\vec{z}\;\bar{\varphi}_\Delta(\vec{z})\bar{J}(\vec{z}) \tag{135}$$

where the bulk field φ_Δ is related to the boundary field $\bar{\varphi}_\Delta$ by the limiting relation,

$$\bar{\varphi}_\Delta(\vec{z}) = \lim_{z_0\to\infty} z_0^{\Delta-d}\varphi_\Delta(z_0,\vec{z}) \tag{136}$$

The corresponding *boundary-to-bulk propagator* is the Poisson kernel, [3]

$$K_\Delta(z,\vec{x}) = C_\Delta\left(\frac{z_0}{z_0^2 + (\vec{z}-\vec{x})^2}\right)^\Delta \tag{137}$$

---

[v] The study of quantum Liouville theory with a SO(2,1) invariant vacuum[85] is closely related to the study of AdS_2, as was shown in Ref. 86. Propagators and amplitudes were studied there long ago[85] and the $\mathcal{N}=1$ supersymmetric generalization is also known.[87]

The bulk field generated in response to the boundary source $\bar{J}$ is given by

$$\varphi_\Delta(z) = \int_{\partial H} d^d\vec{z}\; K_\Delta(z,\vec{x})\bar{J}(\vec{x}) \tag{138}$$

This propagator will be especially important in the AdS/CFT correspondence.

The gauge propagator

Let $A_\mu(z)$ be a massless or massive gauge field, whose linearized dynamics is given by a coupling to a covariantly conserved bulk current j^μ via the action

$$S_A = \int_H d^{d+1}z\; \sqrt{g}\left[\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \frac{1}{2}m^2 A_\mu A^\mu - A_\mu j^\mu\right] \tag{139}$$

It would be customary to introduce a gauge fixing term, such as Feynman gauge, to render the second order differential operator acting on A_μ invertible when $m = 0$. A more convenient way to proceed is to remark that the differential operator needs to be inverted only on the subspace of all j^μ that are covariantly conserved. The gauge propagator is a bivector $G_{\mu\nu'}(z,z')$ which satisfies

$$\begin{aligned} -\tfrac{1}{\sqrt{g}}\partial_\sigma\left[\sqrt{g}g^{\sigma\rho}\partial_{[\rho}G_{\mu]\nu'}(z,z')\right] + m^2 G_{\mu\nu'}(z,z') \\ = g_{\mu\nu}\delta(z,z') + \partial_\mu\partial_{\nu'}\Lambda(u) \end{aligned} \tag{140}$$

The term in Λ is immaterial when integrated against a covariantly conserved current. For the massless case, the gauge propagator is given by Refs. 78, 79, see also Ref. 88

$$G_{\mu\nu'},(z,z') = -(\partial_\mu\partial_{\nu'}u)F(u) + \partial_\mu\partial_{\nu'}S(u) \tag{141}$$

where S is a gauge transformation function, while the physical part of the propagator takes the form,

$$F(u) = \frac{\Gamma((d-1)/2)}{4\pi^{(d+1)/2}}\frac{1}{[u(u+2)]^{(d-1)/2}} \tag{142}$$

The massless graviton propagator

The action for matter coupled to gravity in an AdS background is given by

$$S_g = \frac{1}{2}\int_H d^{d+1}z\; \sqrt{g}(-R_g + \Lambda) + S_m \tag{143}$$

where R_g is the Ricci scalar for the metric g and Λ is the "cosmological constant". S_m is the matter action, whose variation with respect to the metric is, by definition, the stress tensor $T_{\mu\nu}$. The stress tensor is covariantly conserved $\nabla_\mu T^{\mu\nu} = 0$. Einstein's equations read

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu}(R_g - \Lambda) = T_{\mu\nu} \qquad\qquad T_{\mu\nu} = \frac{\delta S_m}{\sqrt{g}\delta g^{\mu\nu}} \tag{144}$$

We take $\Lambda = -d(d-1)$, so that in the absence of matter sources, we obtain Euclidean AdS$= H$ with $R_g = -d(d+1)$ as the maximally symmetric solution. To obtain the equation for the graviton propagator $G_{\mu\nu;\mu'\nu'}(z,w)$, it suffices to linearize Einstein's equations around the AdS metric in terms of small deviations $h_{\mu\nu} = \delta g_{\mu\nu}$ of the metric. One finds

$$h_{\mu\nu}(z) = \int_H d^{d+1}w \ \sqrt{g} G_{\mu\nu;\mu'\nu'}(z,w) T^{\mu'\nu'}(w) \tag{145}$$

where the graviton propagator satisfies

$$\begin{aligned} &W_{\mu\nu}{}^{\kappa\lambda} G_{\kappa\lambda\mu'\nu'} \\ &= \left(g_{\mu\mu'} g_{\nu\nu'} + g_{\mu\nu'} g_{\nu\mu'} - \tfrac{2 g_{\mu\nu} g_{\mu'\nu'}}{d-1} \right) \delta(z,w) + \nabla_{\mu'} \Lambda_{\mu\nu;\nu'} + \nabla_{\nu'} \Lambda_{\mu\nu;\mu'} \end{aligned}$$

and the differential operator W is defined by

$$\begin{aligned} W_{\mu\nu}{}^{\kappa\lambda} G_{\kappa\lambda\mu'\nu'} \equiv & -\nabla^\sigma \nabla_\sigma G_{\mu\nu;\mu'\nu'} - \nabla_\mu \nabla_\nu G^\sigma{}_{\sigma;\mu'\nu'} + \nabla_\mu \nabla^\sigma G_{\sigma\nu;\mu'\nu'} \\ & + \nabla_\nu \nabla^\sigma G_{\mu\sigma;\mu'\nu'} - 2 G_{\mu\nu;\mu'\nu'} + 2 g_{\mu\nu} G^\sigma{}_{\sigma;\mu'\nu'} \end{aligned} \tag{146}$$

The solution to this equation is obtained by decomposing G onto a basis of 5 irreducible $SO(1,d)$-tensors, which may all be expressed in terms of the metric $g_{\mu\nu}$ and the derivatives of the chordal distance $\partial_\mu u$, $\partial_\mu \partial_{\nu'} u$ etc. One finds that three linear combinations of these 5 tensors correspond to diffeomorphisms, so that we have

$$\begin{aligned} G_{\mu\nu;\mu'\nu'} = & (\partial_\mu \partial_{\mu'} u \ \partial_\nu \partial_{\nu'} u + \partial_\mu \partial_{\nu'} u \ \partial_\nu \partial_{\mu'} u) G(u) + g_{\mu\nu} g_{\mu'\nu'} H(u) \\ & + \nabla_{(\mu} S_{\nu);\mu'\nu'} + \nabla_{(\mu'} S_{\mu\nu);\nu')} \end{aligned} \tag{147}$$

The functions G precisely obeys the equation for a massless scalar propagator $G_\Delta(u)$ with $\Delta = d$, so that $G(u) = G_d(u)$. The function $H(u)$ is then given by

$$-(d-1)H(u) = 2(1+u)^2 G(u) + 2(d-2)(1+u) \int_\infty^u dv G(v) \tag{148}$$

which may also be expressed in terms of hypergeometric functions. The graviton propagator was derived using the above methods, or alternatively in De Donder gauge in Ref. 79. Propagators for other fields, such as massive tensor and form fields were constructed in Refs. 80 and 81; see also Refs. 82 and 83.

6.4. *Conformal Structure of 1- 2- and 3- Point Functions*

Conformal invariance is remarkably restrictive on correlation functions with 1, 2, and 3 conformal operators.[89] We illustrate this point for correlation functions of superconformal primary operators, which are all scalars.

The *1-point function* is given by

$$\langle \mathcal{O}_\Delta(x) \rangle = \delta_{\Delta,0} \tag{149}$$

Indeed, by translation invariance, this object must be independent of x, while by scaling invariance, an x-independent quantity can have dimension Δ only when $\Delta = 0$, in which case when have the identity operator.

The *2-point function* is given by

$$\langle \mathcal{O}_{\Delta_1}(x_1)\mathcal{O}_{\Delta_2}(x_2) \rangle = \frac{\delta_{\Delta_1,\Delta_2}}{|x_1 - x_2|^{2\Delta_1}} \tag{150}$$

Indeed, by Poincaré symmetry, this object only depends upon $(x_1 - x_2)^2$; by inversion symmetry, it must vanish unless $\Delta_1 = \Delta_2$; by scaling symmetry one fixes the exponent; and by properly normalizing the operators, the 2-point function may be put in diagonal form with unit coefficients.

The *3-point function* is given by

$$\begin{aligned} &\langle \mathcal{O}_{\Delta_1}(x_1)\mathcal{O}_{\Delta_2}(x_2)\mathcal{O}_{\Delta_3}(x_3) \rangle \\ &= \frac{c_{\Delta_1\Delta_2\Delta_3}(g_s, N)}{|x_1 - x_2|^{\Delta-2\Delta_3}|x_2 - x_3|^{\Delta-2\Delta_1}|x_3 - x_1|^{\Delta-2\Delta_2}} \end{aligned} \tag{151}$$

where $\Delta = \Delta_1 + \Delta_2 + \Delta_3$. The coefficient $c_{\Delta_1\Delta_2\Delta_3}$ is independent of the x_i and will in general depend upon the coupling g^2_{YM} of the theory and on the Yang-Mills gauge group through N.

6.5. *SYM Calculation of 2- and 3- Point Functions*

All that is needed to compute the SYM correlation functions of the composite operators

$$\mathcal{O}_\Delta(x) \equiv \frac{1}{n_\Delta} \mathrm{str} X^{i_1}(x) \cdots X^{i_\Delta}(x) \tag{152}$$

to Born level (order g^0_{YM}) is the propagator of the scalar field

$$\langle X^{ic}(x_1) X^{jc'}(x_2) \rangle = \frac{\delta^{ij}\delta^{cc'}}{4\pi^2(x_1 - x_2)^2} \tag{153}$$

where c is a color index running over the adjoint representation of $SU(N)$ while $i = 1, \cdots, 6$ labels the fundamental representation of $SO(6)$. Clearly, the 2- and 3- point functions have the space-time behavior expected from the preceding discussion of conformal invariance. Normalizing the 2-point function as below, we have $n_k^2 = \mathrm{str}(T^{c_1}\cdots T^{c_k})\mathrm{str}(T^{c_1}\cdots T^{c_k})$.

$$\langle \mathcal{O}_{\Delta_1}(x_1)\mathcal{O}_{\Delta_2}(x_2)\rangle = \frac{\delta_{\Delta_1,\Delta_2}}{(x_1-x_2)^{2\Delta_1}}$$
$$\langle \mathcal{O}_{\Delta_1}(x_1)\mathcal{O}_{\Delta_2}(x_2)\mathcal{O}_{\Delta_3}(x_3)\rangle \sim \frac{1}{(x_1-x_2)^{\Delta_{12}}(x_2-x_3)^{\Delta_{23}}(x_3-x_1)^{\Delta_{31}}} \qquad (154)$$

Using the fact that the number Δ_i of propagators emerging from operator $\mathcal{O}_{\Delta_i}$ equals the sum $\Delta_{ij}+\Delta_{ik}$, we find $2\Delta_{ij} = \Delta_i + \Delta_j - \Delta_k$, in agreement with (151). The precise numerical coefficients may be worked out with the help of the contractions of color traces.

6.6. *AdS Calculation of 2- and 3- Point Functions*

On the AdS side, the 2-point function to lowest order is obtained by taking the boundary-to-bulk propagator $K_\Delta(z,\vec{x})$ for a field with dimension Δ and extracting the z_0^Δ behavior as $z_0 \to 0$, which gives

$$\lim_{z_0\to 0} z_0^{-\Delta} K_\Delta(z,\vec{x}) \sim \frac{1}{(\vec{z}-\vec{x})^{2\Delta}} \qquad (155)$$

in agreement with the behavior predicted from conformal invariance.[90]

The 3-point function involves an integral over the intermediate supergravity interaction point, and is given by

$$\mathcal{G}(\Delta_1,\Delta_2,\Delta_3)\int_{S^5} Y_{\Delta_1}Y_{\Delta_2}Y_{\Delta_3}\int_H \frac{d^5z}{z_0^5}\prod_{i=1}^{3} C_{\Delta_i}\left(\frac{z_0}{z_0^2+(\vec{z}-\vec{x}_i)^2}\right)^{\Delta_i} \qquad (156)$$

where $\mathcal{G}(\Delta_1,\Delta_2,\Delta_3)$ stands for the supergravity 3-point coupling and the second factor is the integrals over the spherical harmonics of S^5. To carry out the integral over H, one proceeds in three steps. First, use a translation to set $\vec{x}_3 = 0$. Second, use an inversion about 0, given by $z^\mu \to z^\mu/z^2$ to set $\vec{x}_3' = \infty$. Third, having one point at ∞, one may now use translation invariance again, to obtain for the H-integral

$$\sim (x_{13}')^{2\Delta_1}(x_{23}')^{2\Delta_2}\int_H \frac{d^5z}{z_0^5}\frac{z_0^{\Delta_1+\Delta_2+\Delta_3}}{z^{2\Delta_1}[z_0^2+(\vec{z}-\vec{x}_{13}'-\vec{x}_{23}')^2]^{\Delta_2}} \qquad (157)$$

Carrying out the $\vec{z}$ integral using a Feynman parametrization of the integral and then carrying out the z_0 integral, one recovers again the general space-time dependence of the 3-point function.[90] A more detailed account of the AdS calculations of the 2- and 3-point functions will be given in §8.4.

6.7. *Non-Renormalization of 2- and 3- Point Functions*

Upon proper normalization of the operators $\mathcal{O}_\Delta$, so that their 2-point function is canonically normalized, the three point couplings $c_{\Delta_1,\Delta_2,\Delta_3}(g_{YM}^2, N)$ may be computed in a unique manner. On the SYM side, small coupling g_{YM} perturbation theory yields results for $g_{YM} \ll 1$, but all N. On the AdS side, the only calculation available in practice so far is at the level of classical supergravity, which means the large N limit (where quantum loops are being neglected), as well as large 't Hooft coupling $\lambda = g_{YM}^2 N$ (where α' string corrections to supergravity are being neglected). Therefore, a direct comparison between the two calculations cannot be made because the calculations hold in mutually exclusive regimes of validity.

Nonetheless, one may compare the results of the calculations in both regimes. This involves obtaining a complete normalization of the supergravity three-point couplings $\mathcal{G}(\Delta_1, \Delta_2, \Delta_3)$, which was worked out in Ref. 91. It was found that

$$\lim_{N,\lambda=g_sN\to\infty} c_{\Delta_1,\Delta_2,\Delta_3}(g_s, N)\bigg|_{AdS} = \lim_{N\to\infty} c_{\Delta_1,\Delta_2,\Delta_3}(0, N)\bigg|_{SYM} \qquad (158)$$

Given that this result holds irrespectively of the dimensions Δ_i, it was conjectured in Ref. 91 that this result should be viewed as emerging from a *non-renormalization effect* for 2- and 3-point functions of 1/2 BPS operators. Consequently, it was conjectured that the equality should hold for all couplings, at large N,

$$\lim_{N\to\infty} c_{\Delta_1,\Delta_2,\Delta_3}(g_s, N)\bigg|_{AdS} = \lim_{N\to\infty} c_{\Delta_1,\Delta_2,\Delta_3}(g_{YM}^2, N)\bigg|_{SYM} \qquad (159)$$

and more precisely that $c_{\Delta_1,\Delta_2,\Delta_3}(g_s, N)$ be independent of g_s in the $N \to \infty$ limit.

Independence on g_{YM} of the three point coupling $c_{\Delta_1,\Delta_2,\Delta_3}(g_{YM}^2, N)$ is now a problem purely in $\mathcal{N}=4$ SYM theory, and may be studied there in its own right. This issue has been pursued since by performing calculations of the same correlators to order g_{YM}^2. It was found that to this order, neither the 2- nor the 3-point functions receive any corrections.[92] Consequently, a stronger conjecture was proposed to hold for all N,

$$c_{\Delta_1,\Delta_2,\Delta_3}(g_s, N)\bigg|_{AdS} = c_{\Delta_1,\Delta_2,\Delta_3}(g_{YM}^2, N)\bigg|_{SYM} \qquad (160)$$

Further evidence that this relation holds has been obtained using $\mathcal{N}=1$ superfields[93,94] and $\mathcal{N}=2$ off-shell analytic/harmonic superfield methods.[107,108] The problem has also been investigated using $\mathcal{N}=4$ on-shell superspace methods,[95,96] via the study of nilpotent superconformal invariants, which had been introduced for OSp(1,N) in Ref. 97. Similar non-renormalization effects may be derived for 1/4 BPS operators and their correlators as well.[98] Two and three point correlators have also been investigated for superconformal descendant fields; for the R-symmetry current in Ref. 90 and later in Ref. 100; see also Refs. 99 and 101. Additional references include[102] and 103. A further test of the Maldacena conjecture involving the Weyl anomaly is in Ref. 104.

6.8. *Extremal 3-Point Functions*

We now wish to investigate the dependence of the 3-point function of 1/2 BPS single trace operators

$$\langle \mathcal{O}_{\Delta_1}(x_1)\mathcal{O}_{\Delta_2}(x_2)\mathcal{O}_{\Delta_3}(x_3)\rangle \tag{161}$$

on the dimensions Δ_i a little more closely. Recall that these operators transform under the irreducible representations of $SU(4)_R$ with Dynkin labels $[0,\Delta_i,0]$. As a result, the correlators must vanish whenever $\Delta_i > \Delta_j+\Delta_k$ for any one of the labels $i\neq j,k$, since in this case no $SU(4)_R$ singlet exists. Whenever $\Delta_i\leq\Delta_j+\Delta_k$, for all i,j,k, the correlator is allowed by $SU(4)_R$ symmetry.

These facts may also be seen at Born level in SYM perturbation theory by matching the number of X propagators connecting different operators. If $\Delta_i>\Delta_j+\Delta_k$, it will be impossible to match up the X propagator lines and the diagram will have to vanish.

The case where $\Delta_i=\Delta_j+\Delta_k$ for one of the labels i is of special interest and is referred to as an *extremal correlator.*[105] Although allowed by $SU(4)_R$ group theory, its Born graph effectively factorizes into two 2-point functions, because no X propagators directly connect the vertices operators j and k. Thus, the extremal 3-point function is non-zero. However, the supergravity coupling $\mathcal{G}(\Delta_1,\Delta_2,\Delta_3)\sim\Delta_1-\Delta_2-\Delta_3$ vanishes in the extremal case as was shown in Ref. 91. The reason that all these statements can be consistent with the AdS/CFT correspondence is because the AdS_5 integration actually has a pole at the extremal dimensions, as may indeed be seen by taking a closer look at the integrals,

$$\int_H \frac{d^5z}{z_0^5}\prod_{i=1}^{3}\frac{z_0^{\Delta_i}}{(z_0^2+(\vec{z}-\vec{x}_i)^2)^{\Delta_i}}\sim\frac{1}{\Delta_1-\Delta_2-\Delta_3} \tag{162}$$

Thus, the AdS/CFT correspondence for extremal 3-point functions holds because a zero in the supergravity coupling is compensated by a pole in the AdS_5 integrals.

Actually, the dimensions Δ_i are really integers (which is why "pole" was put in quotation marks above) and direct analytic continuation in them is not really justified. It was shown in Ref. 105 that when keeping the dimensions Δ_i integer, it is possible to study the supergravity integrands more carefully and to establish that while the bulk contribution vanishes, there remains a boundary contribution (which was immaterial for non-extremal correlators). A careful analysis of the boundary contribution allows one to recover agreement with the SYM calculation directly.

6.9. *Non-Renormalization of General Extremal Correlators*

Extremal correlators may be defined not just for 3-point functions, but for general $(n+1)$-point functions. Let $\mathcal{O}_\Delta$ and $\mathcal{O}_{\Delta_i}$ with $i = 1, \cdots, n$ be 1/2 BPS chiral primary operators obeying the relation $\Delta = \Delta_1 + \cdots + \Delta_n$, which generalizes the extremality relation for the 3-point function. We have the *extremal correlation non-renormalization conjecture*, stating the form of the following correlator,[105]

$$\langle \mathcal{O}_\Delta(x)\mathcal{O}_{\Delta_1}(x_1)\cdots\mathcal{O}_{\Delta_n}(x_n)\rangle = A(\Delta_i; N)\prod_{i=1}^{n}\frac{1}{(\vec{x}-\vec{x}_i)^{2\Delta_i}} \tag{163}$$

The conjecture furthermore states that the overall function $A(\Delta_i; N)$ is independent of the points x_i and x and is also independent of the string coupling constant $g_s = g_{YM}^2$. *The conjecture also states that the associated supergravity bulk couplings $\mathcal{G}(\Delta; \Delta_1, \cdots, \Delta_n)$ must vanish.*[105]

There is by now ample evidence for the conjecture and we shall briefly review it here. First, there is evidence from the SYM side. To Born level (order $\mathcal{O}(g_{YM}^0)$, the factorization of the space-time dependence in a product of 2-point functions simply follows from the fact that no X-propagator lines can connect different points x_i; instead all X-propagator lines emanating from any vertex x_i flow into the point x. The absence of $\mathcal{O}(g_{YM}^2)$ perturbative corrections was demonstrated in Ref. 109. Off-shell $\mathcal{N} = 2$ analytic/harmonic superspace methods have been used to show that g_{YM} corrections are absent to all orders of perturbation theory,[107,108]

On the AdS side, the simplest diagram that contributes to the extremal correlator is the contact graph, which is proportional to

$$\mathcal{G}(\Delta;\Delta_1,\cdots,\Delta_n)\int_H \frac{d^5z}{z_0^5}\frac{z_0^\Delta}{(z-\vec{x})^{2\Delta}}\prod_{i=1}^n \frac{z_0^{\Delta_i}}{(z-\vec{x}_i)^{2\Delta_i}} \tag{164}$$

In view of the relation $\Delta = \Delta_1 + \cdots + \Delta_n$, the integration is convergent everywhere in H, except when $\vec{z} \to \vec{x}$ and $z_0 \to 0$, where a simple pole arises in $\Delta - \Delta_1 - \cdots - \Delta_n$. Finiteness of Type IIB superstring theory on $\mathrm{AdS}_5 \times \mathrm{S}^5$(which we take as an assumption here) guarantees that the full correlator must be convergent. Therefore, the associated supergravity bulk coupling must vanish,

$$\mathcal{G}(\Delta;\Delta_1,\cdots,\Delta_n) \sim \Delta - \Delta_1 - \cdots - \Delta_n \tag{165}$$

as indeed stated in the conjecture. Assuming that it makes sense to "analytically continue in the dimensions Δ", one may proceed as follows. The pole of the z-integration and the zero of the supergravity coupling $\mathcal{G}$ compensate one another and the contribution of the contact graph to the extremal correlator will be given by the residue of the pole, which is precisely of the form (163). It is also possible to carefully treat the boundary contributions generated by the supergravity action in the extremal case, to recover the same result.[105]

The analysis of all other AdS graph, which have at least one bulk-to-bulk exchange in them, was carried out in detail in Ref. 105. For the exchange of chiral primaries in the graph, the extremality condition $\Delta = \Delta_1 + \cdots + \Delta_n$ implies that each of the exchange bulk vertices must be extremal as well. A non-zero contribution can then arise only if the associated integral is divergent, produces a pole in the dimensions, and makes the interaction point collapse onto the boundary ∂H. Dealing with all intermediate external vertices in this way, one recovers that all intermediate vertices have collapsed onto $\vec{x}$, thereby reproducing the space-time behavior of (163). The exchange of descendants may be dealt with in an analogous manner.

Assuming non-renormalization of 2- and 3-point functions for all (single and multiple trace) 1/2 BPS operators, and assuming the space-time form (163) of the extremal correlators, it is possible to *prove* that the overall factor $A(\Delta_i;N)$ is independent of $g_s = g_{YM}^2$, as was done in Ref. 105 in a special case. We present only the simplest non-trivial case of $n=3$ and $\Delta = 6$; the general case may be proved by induction. Assuming the space-time form, we have

$$\langle \mathcal{O}_6(x)\mathcal{O}_2(x_1)\mathcal{O}_2(x_2)\mathcal{O}_2(x_3)\rangle = A\prod_{i=1}^{3}\frac{1}{(\vec{x}-\vec{x}_i)^4} \tag{166}$$

We begin with the OPE

$$\mathcal{O}_6(x)\mathcal{O}_2(x_1) \sim \frac{c\mathcal{O}_4(x)+c'[\mathcal{O}_2\mathcal{O}_2]_{\max}(x)}{(x-x_1)^4} + \text{less singular} \tag{167}$$

Using non-renormalization of the 3-point functions $\langle \mathcal{O}_6\mathcal{O}_2\mathcal{O}_4\rangle$ and $\langle \mathcal{O}_6\mathcal{O}_2[\mathcal{O}_2\mathcal{O}_2]_{\max}$, we find that c and c' are independent of the coupling g_{YM}. Now substitute the above OPE into the correlator (166), and use the fact that the 3-point functions $\langle \mathcal{O}_4\mathcal{O}_2\mathcal{O}_2\rangle$ and $\langle \mathcal{O}_2\mathcal{O}_2[\mathcal{O}_2\mathcal{O}_2]_{\max}$ are not renormalized. It immediately follows that A in (166) is independent of the coupling.

6.10. *Next-to-Extremal Correlators*

The space-time dependence of extremal correlators was characterized by its factorization into a product of n 2-point functions. The space-time dependence of *Next-to-extremal correlators* $\langle \mathcal{O}_\Delta(x)\mathcal{O}_{\Delta_1}(x_1)\cdots\mathcal{O}_{\Delta_n}(x_n)\rangle$, with the dimensions satisfying $\Delta = \Delta_1+\cdots+\Delta_n-2$ is characterized by its factorization into a product of $n-2$ two-point functions and one 3-point function. Therefore, the conjectured space-time dependence of *next-to-extremal correlators* is given by Ref. 108 as

$$\begin{aligned}&\langle \mathcal{O}_\Delta(x)\mathcal{O}_{\Delta_1}(x_1)\cdots\mathcal{O}_{\Delta_n}(x_n)\rangle \\ &\quad= \frac{B(\Delta_i;N)}{x_{12}^2(x-x_1)^{2\Delta_1-2}(x-x_2)^{2\Delta_2-2}}\prod_{i=3}^{n}\frac{1}{(x-x_i)^{2\Delta_i}}\end{aligned} \tag{168}$$

where the overall strength $B(\Delta_i;N)$ is independent of g_{YM}. This form is readily checked at Born level and was verified at order $\mathcal{O}(g_{YM}^2)$ by Ref. 106.

On the AdS side, the exchange diagrams, say with a single exchange, are such that one vertex is extremal while the other vertex is not extremal. A divergence arises when the extremal vertex is attached to the operator of maximal dimension Δ and its collapse onto the point x now produces a 3-point correlator times $n-2$ two-point correlators, thereby reproducing the space-time dependence of (168). Other exchange diagrams may be handled analogously. However, there is also a contact graph, whose AdS integration is now *convergent*. Since the space-time dependence of this contact term is qualitatively different from the factorized form of (168), the only manner in which (168) can hold true is if the supergravity bulk coupling associated with next-to-extremal couplings vanishes,

$$\mathcal{G}(\Delta; \Delta_1, \cdots, \Delta_n) = 0 \qquad \text{whenever} \qquad \Delta = \Delta_1 + \cdots + \Delta_n - 2 \quad (169)$$

which is to be included as part of the conjecture.[110] This type of cancellation has been checked to low order in Ref. 111.

6.11. *Consistent Decoupling and Near-Extremal Correlators*

The vanishing of the extremal and next-to-extremal supergravity couplings has a direct interpretation, at least in part, in supergravity. Recall that the operator $\mathcal{O}_2$ and its descendants are dual to the 5-dimensional supergravity multiplet on AdS_5, while the operators $\mathcal{O}_\Delta$ with $\Delta \geq 3$ and its descendants are dual to the Kaluza-Klein excitations on S^5 of the 10-dimensional supergravity multiplet. Now, prior work on gauged supergravity.[112,113] has shown that the 5-dimensional gauged supergravity theory on AdS_5 all by itself exists and is consistent. Thus, there must exist a *consistent truncation of the Kaluza-Klein modes of supergravity on* $AdS_5 \times S^5$ *to only the supergravity on* AdS_5; see also Ref. 114. In a perturbation expansion, this means that if only AdS_5 supergravity modes are excited, then the Euler-Lagrange equations of the full $AdS_5 \times S^5$ supergravity must close on these excitations alone without generating Kaluza-Klein excitation modes. This means that the one 1-point function of any Kaluza-Klein excitation operator in the presence of AdS_5 supergravity alone must vanish, or

$$\mathcal{G}(\Delta, \Delta_1, \cdots, \Delta_n) = 0, \quad \Delta_i = 2, \ i = 1, \cdots, n \quad \text{for all} \quad \Delta \geq 4 \quad (170)$$

When $\Delta > 2n$, the cancellation takes place by $SU(4)_R$ group theory only. For $\Delta = 2n$ and $\Delta = 2n - 2$, we have special cases of extremal and next-to-extremal correlators respectively, but for $4 \leq \Delta \leq 2n - 4$, they belong to a larger class. We refer to these as *near-extremal correlators,*[110]

$$\langle \mathcal{O}_\Delta(x) \mathcal{O}_{\Delta_1}(x_1) \cdots \mathcal{O}_{\Delta_n}(x_n) \rangle, \ \Delta = \Delta_1 + \cdots + \Delta_n - 2m, \ 0 \leq m \leq n-2 \quad (171)$$

The principal result on near-extremal correlators (but which are not of the extremal or next-to-extremal type) is that they do receive coupling dependent quantum corrections, but only through lower point functions.[110] Associated supergravity couplings must vanish,

$$\mathcal{G}(\Delta, \Delta_1, \cdots, \Delta_n) = 0, \quad \Delta = \Delta_1 + \cdots + \Delta_n - 2m, \quad 0 \leq m \leq n - 2 \quad (172)$$

Arguments in favor of this conjecture may be given based on the divergence structure of the AdS integrals and on perturbation calculations in SYM.

6.12. *Problem Sets*

(6.1) Using infinitesimal special conformal symmetry (or global inversion under which $x^\mu \to x^\mu/x^2$) show that $\langle \mathcal{O}_\Delta(x)\mathcal{O}_{\Delta'}(x')\rangle = 0$ unless $\Delta' = \Delta$.

(6.2) Gauge dependent correlators in gauge theories such as $\mathcal{N} = 4$ SYM theory will, in general, depend upon a renormalization scale μ. (a) Show that the general form of the scalar two point function to one loop order is given by

$$\langle X^{ic}(x)X^{jc'}(y)\rangle = \frac{\delta^{cc'}\delta^{ij}}{(x-y)^2}\left(A + B\ln(x-y)^2\mu^2\right)$$

for some numerical constants A and B. (b) Show that the 2-pt function of the gauge invariant operator $\mathcal{O}_2(x) \equiv \mathrm{tr}X^i(x)X^j(x) - \frac{1}{6}\delta^{ij}\sum_k \mathrm{tr}X^k(x)X^k(x)$ is μ-independent. (c) Show that the 2-pt function of the gauge invariant operator $\mathcal{O}_K(x) \equiv \mathrm{tr}X^i(x)X^i(x)$ (the Konishi operator) is μ-dependent. (d) Calculate the 1-loop anomalous dimensions of $\mathcal{O}_2$ and $\mathcal{O}_K$.

(6.3) Consider the Laplace operator Δ acting on scalar functions on the sphere S^d with round $SO(d+1)$-invariant metric and radius R. Compute the eigenvalues of of Δ. Suggestion : Δ is related to the quadratic Casimir operator $L^2 \equiv \sum_{i,j=1}^{d+1} L_{ij}^2$ where L_{ij} are the generators of $d+1$-dimensional angular momentum, i.e. generators of $SO(d+1)$; thus the problem may be solved by pure group theory methods, analogous to those used for rotations on S^2.

(III.4) Continuing on the above problem, show that the eigenfunctions are of the form $c_{i_1\cdots i_p}x^{i_1}\cdots x^{i_p}$, where we have now represented the sphere by the usual equation in $\mathbf{R}^{d+1}$: $\sum_i (x^i)^2 = R^2$ and c is totally symmetric and traceless.

7. Structure of General Correlators

In the previous section, we have concentrated on matching between the SYM side and the AdS side of the Maldacena correspondence correlation functions that were not renormalized or were simply renormalized from their free form. This led us to uncover a certain number of important *non-renormalization effects*, most of which are at the level of conjecture.

However, $\mathcal{N} = 4$ super-Yang-Mills theory is certainly not a free quantum field theory, and generic correlators will receive quantum corrections from their free field values, and therefore will acquire non-trivial coupling $g_s = g_{YM}^2$ dependence. In this section, we analyze the behavior of such correlators. We shall specifically deal with the 4-point function. The relevant dynamical information available from correlators in conformal quantum field theory is contained in the *scaling dimensions of general operators*, in the *operator mixings* between general operators and in the values of the *operator product (OPE) coefficients*. As in the case of the 3-point function, a direct quantitative comparison between the results of weak coupling g_{YM} perturbation theory in SYM and the large N, large 't Hooft coupling $\lambda = g_{YM}^2 N$ limit of supergravity cannot be made, because the domains of validity of the expansions do not overlap. Nonetheless, general properties lead to exciting and non-trivial comparisons, which we shall make here.

7.1. *RG Equations for Correlators of General Operators*

It is a general result of quantum field theory that all renormalizations of local operators are multiplicative. This is familiar for canonical fields; for example the bare field $\phi_0(x)$ and the renormalized field $\phi(x)$ in a scalar field theory are related by the field renormalization factor Z_ϕ via the relation $\phi_0(x) = Z_\phi \phi(x)$. Composite operators often requires *additive renormalizations*; for example the proper definition of the operator $\phi^2(x)$ requires the subtraction of a constant C. If this constant is viewed as multiplying the identity operator I in the theory, then renormalization may alternatively be viewed as multiplicative (by a matrix) on an array of two operators I and $\phi^2(x)$ as follows,

$$\begin{pmatrix} I \\ \phi^2(x) \end{pmatrix}_0 = \begin{pmatrix} I & 0 \\ -C & Z_{\phi^2} \end{pmatrix} \begin{pmatrix} I \\ \phi^2(x) \end{pmatrix} \tag{173}$$

The general rule is that operators will renormalize with operators with the same quantum numbers but of lesser or equal dimension.

In more complicated theories such as $\mathcal{N} = 4$ super-Yang-Mills theory, renormalization will continue to proceed in a multiplicative way. If we denote a basis of (local gauge invariant polynomial) operators by $\mathcal{O}_I$, and their bare counterparts by $\mathcal{O}_{0I}$, then we have the following multiplicative renormalization formula

$$\mathcal{O}_{0I}(x) = \sum_J Z_I{}^J \mathcal{O}_J(x) \tag{174}$$

The field renormalization matrix $Z_I{}^J$ may be arranged in block lower triangular form, in ascending value of the operator dimensions, generalizing

(173). Consider now a general correlator of such operators

$$G_{I_1,\cdots,I_n}(x_i;g,\mu) \equiv \langle \mathcal{O}_{I_1}(x_1)\cdots\mathcal{O}_{I_n}(x_n)\rangle \tag{175}$$

and its bare counterpart $G_{0I_1,\cdots,I_n}(x_i;g_0,\Lambda)$, in a theory in which we schematically represent the dimensionless and dimensionful couplings by g, and their bare counterparts by g_0. The renormalization scale is μ and the UV cutoff is Λ. Multiplicative renormalization implies the following relation between the renormalized and bare correlators

$$\begin{aligned} &G_{0I_1,\cdots,I_n}(x_i;g_0,\Lambda) \\ &\quad = \textstyle\sum_{J_1,\cdots,J_n} Z_{I_1}{}^{J_1}\cdots Z_{I_n}{}^{J_n}(g,\mu,\Lambda)G_{J_1,\cdots,J_n}(x_i;g,\mu) \end{aligned} \tag{176}$$

Keeping the bare parameters g_0 and Λ fixed and varying the renormalization scale μ, we see that the lhs is independent of μ. Differentiating both sides with respect to μ is the standard way of deriving the renormalization group equations for the renormalized correlators, and we find

$$\begin{aligned} &\left(\tfrac{\partial}{\partial \ln\mu} + \beta\tfrac{\partial}{\partial g}\right)G_{I_1,\cdots,I_n}(x_i;g,\mu) \\ &\quad -\textstyle\sum_{j=1}^{n}\sum_J \gamma_{I_j}{}^J G_{I_1,\cdots,I_{j-1},J,I_{j+1},\cdots,I_n}(x_i;g,\mu) = 0 \end{aligned} \tag{177}$$

where the RG β-function and anomalous dimension matrix $\gamma_I{}^J$ are defined by

$$\beta(g) \equiv \left.\frac{\partial g}{\partial \ln\mu}\right|_{g_0,\Lambda} \qquad \gamma_I{}^J(g) \equiv -\sum_K (Z^{-1})_I{}^K \left.\frac{\partial Z_K{}^J}{\partial \ln\mu}\right|_{g_0,\Lambda} \tag{178}$$

For each I, only a finite number of J's are non-zero in the sum over J. The diagonal entries $\gamma_I{}^I$ contribute to the anomalous dimension of the operator $\mathcal{O}_I$, while the off-diagonal entries are responsible for operator mixing. Operators that are eigenstates of the dimension operator D (at a given coupling g) correspond to the eigenvectors of the matrix γ.

7.2. *RG Equations for Scale Invariant Theories*

Considerable simplifications occur in the RG equations for scale invariant quantum field theories. Scale invariance requires in particular that $\beta(g_*) = 0$, so that the theory is at a fixed point g_*. In rare cases, such as is in fact the case for $\mathcal{N} = 4$ SYM, the theory is scale invariant for all couplings. If no dimensionful couplings occur in the Lagrangian, either from masses or from vacuum expectation vales of dimensionful fields, $\gamma_I{}^J$ is constant and the RG equation becomes a simple scaling equation

$$\frac{\partial}{\partial \ln \mu} G_{I_1,\cdots,I_n}(x_i; g_*, \mu) - \sum_{j=1}^{n} \sum_{J} \gamma_{I_j}{}^{J}(g_*) G_{I_1,\cdots,I_{j-1},J,I_{j+1},\cdots,I_n}(x_i; g_*, \mu) = 0 \tag{179}$$

In conformal theories, the dilation generator may be viewed as a Hamiltonian of the system conjugate to radial evolution.[115] Therefore, in unitary scale invariant theories, the dilation generator should be self-adjoint, and hence the anomalous dimension matrix should be Hermitian.[vi] As such, $\gamma_I{}^J$ must be diagonalizable with real eigenvalues, γ_i. Standard scaling arguments then give the behavior of the correlators

$$G_{I_1,\cdots,I_n}(\lambda x_i; g_*, \lambda^{-1}\mu) = \lambda^{-\Delta_1} \cdots \lambda^{-\Delta_n} G_{I_1,\cdots,I_n}(x_i; g_*, \mu) \tag{180}$$

where the full dimensions Δ_i are given in terms of the canonical dimension δ_i by $\Delta_i = \gamma_i + \delta_i$.

7.3. *Structure of the OPE*

One of the most useful tools of local quantum field theory is the Operator Product Expansion (OPE) which expresses the product of two local operators in terms of a sum over all local operators in the theory,

$$\mathcal{O}_I(x)\mathcal{O}_J(y) = \sum_K C_{IJ}{}^K(x-y; g, \mu)\mathcal{O}_K(y) \tag{181}$$

The OPE should be understood as a relations that holds when evaluated between states in the theory's Hilbert space or when inserted into correlators with other operators,

$$\langle \mathcal{O}_I(x)\mathcal{O}_J(y) \textstyle\prod_L \mathcal{O}_L(z_L)\rangle = \sum_K C_{IJ}{}^K(x-y; g, \mu)\langle \mathcal{O}_K(y) \prod_L \mathcal{O}_L(z_L)\rangle \tag{182}$$

From the latter, together with the RG equations for the correlators, one deduces the RG equations for the OPE coefficient functions $C_{IJ}{}^K$,

$$\left(\frac{\partial}{\partial \ln \mu} + \beta \frac{\partial}{\partial g}\right) C_{IJ}{}^K = \textstyle\sum_L \left(\gamma_I{}^L C_{LJ}{}^K + \gamma_J{}^L C_{IL}{}^K - C_{IJ}{}^L \gamma_L{}^K\right) \tag{183}$$

In a scale invariant theory, we have $\beta = 0$ and $\gamma_I{}^J$ constant. Furthermore, if the theory is unitary, $\gamma_I{}^J$ may be diagonalized in terms of operators $\mathcal{O}_{\Delta_I}$

[vi] In non-unitary theories, the matrix $\gamma_I{}^J$ may be put in Jordan diagonal form, and this form will produce dependence on μ through $\ln \mu$ terms. A fuller discussion is given in Ref. 116.

of definite dimension Δ_I. The OPE then simplifies considerably and we have,[120]

$$\mathcal{O}_{\Delta_I}(x)\mathcal{O}_{\Delta_J}(y) = \sum_K \frac{c_{\Delta_I \Delta_J \Delta_K}}{(x-y)^{\Delta_I+\Delta_J-\Delta_K}} \mathcal{O}_{\Delta_K}(y) \tag{184}$$

The operator product coefficients $c_{\Delta_I \Delta_J \Delta_K}$ are now independent of x and y, but they will depend upon the coupling constants and parameters of the theory, such as g_{YM} and N.

7.4. *Perturbative Expansion of OPE in Small Parameter*

Conformal field theories such as $\mathcal{N} = 4$ SYM have coupling constants g_{YM}, θ_I, N and the theory is (super)-conformal for any value of these parameters. In particular, the scaling dimensions are fixed but may depend upon these parameters in a non-trivial way,

$$\Delta_I = \Delta(g_{YM}, \theta_I, N) \tag{185}$$

The dependence of the composite operators on the canonical fields will in general also involve these coupling dependences.

It is interesting to analyze the effects of small variations in any of these parameters on the structure of the OPE and correlation functions. Especially important is the fact that infinitesimal changes in Δ_I produce *logarithmic dependences* in the OPE. To see this, assume that

$$\Delta_I = \Delta_I^0 + \delta_I \qquad |\delta_I| \ll \Delta_I \tag{186}$$

and now observe that to first order in δ_I, we have

$$\mathcal{O}_{\Delta_I}(x)\mathcal{O}_{\Delta_J}(y) = \sum_K \frac{c_{\Delta_I \Delta_J \Delta_K} \mathcal{O}_{\Delta_K}(y)}{(x-y)^{\Delta_I^0+\Delta_J^0-\Delta_K^0}} \left[1 - (\delta_I + \delta_J - \delta_K) \ln |x-y|\mu\right] \tag{187}$$

In the special case where the dimensions Δ_I^0 and Δ_J^0 are unchanged, because the operators are protected (e.g. BPS) then isolating the logarithmic dependence allows one to compute δ_K and thus the correction to the dimension of operators occurring in the OPE. A useful reference on anomalous dimensions and the OPE, though not in conformal field theory, is in Ref. 121.

7.5. *The 4-Point Function – The Double OPE*

Recall that the AdS/CFT correspondence maps supergravity fields into single-trace 1/2 BPS operators on the SYM side. Thus, the only correlators that can be computed directly are the ones with one 1/2 BPS operator insertions. To explore even the simplest renormalization effects of non-BPS operators, such as their change in dimension, via the AdS/CFT correspondence, we need to go beyond the 3-point function. The simplest case is the 4-point function, which indeed can yield information on the anomalous dimensions of single and double trace operators.

Thanks to conformal symmetry, the 4-point function may be factorized into a factor capturing its overall non-trivial conformal dependence times a function $F(s,t)$ that depends only upon 2 conformal invariants s, t of 4 points,

$$\langle \mathcal{O}_{\Delta_1}.(x_1)\mathcal{O}_{\Delta_2}(x_2)\mathcal{O}_{\Delta_3}(x_3)\mathcal{O}_{\Delta_4}(x_4)\rangle = \frac{1}{|x_{13}|^{\Delta_1+\Delta_3}|x_{24}|^{\Delta_2+\Delta_4}} F(s,t) \quad (188)$$

The conformal invariants of the 4-point function s and t may be chosen as follows

$$s = \frac{1}{2}\frac{x_{13}^2 x_{24}^2}{x_{12}^2 x_{34}^2 + x_{14}^2 x_{23}^2} \qquad t = \frac{x_{12}^2 x_{34}^2 - x_{14}^2 x_{23}^2}{x_{12}^2 x_{34}^2 + x_{14}^2 x_{23}^2} \quad (189)$$

The fact that there are only 2 conformal invariants may be seen as follows. By a translation, take $x_4 = 0$; under an inversion, we then have $x_4' = \infty$ and we may use translations again to choose $x_3' = 0$. There remain 3 Lorentz invariants, x_1^2, x_2^2 and $x_1 \cdot x_2$, and thus 2 independent scale-invariant ratios. Note that 2 is also the number of Lorentz invariants of a massless 4-point function in momentum space.

A specific representation for the function F may be obtained by making use of the OPE twice in the 4-point function, one on the pair $\mathcal{O}_{\Delta_1}(x_1)\mathcal{O}_{\Delta_3}(x_3)$ and once on the pair $\mathcal{O}_{\Delta_2}(x_2)\mathcal{O}_{\Delta_4}(x_4)$. One obtains the *double OPE*, first introduced in Ref. 122

$$\begin{aligned}&\langle \mathcal{O}_{\Delta_1},(x_1)\mathcal{O}_{\Delta_2}(x_2)\mathcal{O}_{\Delta_3}(x_3)\mathcal{O}_{\Delta_4}(x_4)\rangle \\ &\quad = \sum_{\Delta\Delta'} \frac{c_{\Delta_1\Delta_3\Delta}}{|x_{13}|^{\Delta_1\Delta_3\Delta}} \frac{1}{|x_{12}|^{\Delta+\Delta'}} \frac{c_{\Delta_2+\Delta_4-\Delta'}}{|x_{24}|^{\Delta_2+\Delta_4-\Delta'}}\end{aligned} \quad (190)$$

The OPE coefficients $c_{\Delta_1\Delta_3\Delta}$ appeared in the simple OPE of the operators $\mathcal{O}_{\Delta_1}$ and $\mathcal{O}_{\Delta_3}$. General properties of the OPE and double OPE have been studied recently in Refs. 126, 123, 124, 125, and 127 and from a perturbative point of view in Ref. 128; see also Ref. 129.

7.6. *4-pt Function of Dilaton/Axion System*

The possible intermediary fields and operators are restricted by the $SU(4)_R$ tensor product formula for external operators.[vii] Assuming external operators (such as the 1/2 BPS primaries) in representations $[0, \Delta, 0]$ and $[0, \Delta', 0]$, their tensor product decomposes as

$$[0, \Delta, 0] \otimes [0, \Delta', 0] = \oplus_{\mu=0}^{\Delta'} \oplus_{\nu=0}^{\Delta'-\mu} [\nu, \Delta + \Delta' - 2\mu - 2\nu, \nu] \qquad (191)$$

For example, the product of two AdS_5 supergravity primaries in the representation $\mathbf{20'} = [0, 2, 0]$ is given by (the subscript A denotes antisymmetrization)

$$\mathbf{20'} \otimes \mathbf{20'} = \mathbf{1} \oplus \mathbf{15}_A \oplus \mathbf{20'} \oplus \mathbf{84} \oplus \mathbf{105} \oplus \mathbf{175}_A \qquad (192)$$

Actually, the simplest group theoretical structure emerges when taking two $SU(4)_R$ and Lorentz singlets which are $SU(2,2|4)$ descendants. We consider the system of dimension $\Delta = 4$ half-BPS operators dual to the dilaton and axion fields in the bulk;

$$\mathcal{O}_\phi = \mathrm{tr} F_{\mu\nu} F^{\mu\nu} + \cdots \qquad \mathcal{O}_C = \mathrm{tr} F_{\mu\nu} \tilde{F}^{\mu\nu} + \cdots \qquad (193)$$

The further advantage of this system is that the classical supergravity action is simple,

$$S[G, \Phi, C] = \frac{1}{2\kappa_5^2} \int_H \sqrt{G} \left[-R_G + \Lambda + \tfrac{1}{2} \partial_\mu \Phi \partial^\mu \Phi + \tfrac{1}{2} e^{2\Phi} \partial_\mu C \partial^\mu C \right] \qquad (194)$$

In the AdS/CFT correspondence, κ_5^2 may be related to N by $\kappa_5^2 = 4\pi^2/N^2$. This system was first examined in Refs. 130 and 131. An investigation directly of the correlator of half-BPS chiral primaries may be found in Ref. 132.

7.7. *Calculation of 4-point Contact Graph*

The 4-point function receives contributions from the *contact graph* and from a number of *exchange graphs*, which we now discuss in turn. The most general 4-point contact term is given by the following integral,

$$D_{\Delta_1 \Delta_2 \Delta_3 \Delta_4}(x_i) \equiv \int_H \frac{d^{d+1}z}{z_0^{d+1}} \prod_{i=1}^{4} \left(\frac{z_0}{z_0^2 + (\vec{z} - \vec{x}_i)^2} \right)^{\Delta_i} \qquad (195)$$

[vii]Actually, the possible intermediary fields and operators are restricted by the full $SU(2,2|4)$ superconformal algebra branching rules. Since no $\mathcal{N} = 4$ off-shell superfield formulation is available, however, it appears very difficult to make direct use of this powerful fact.

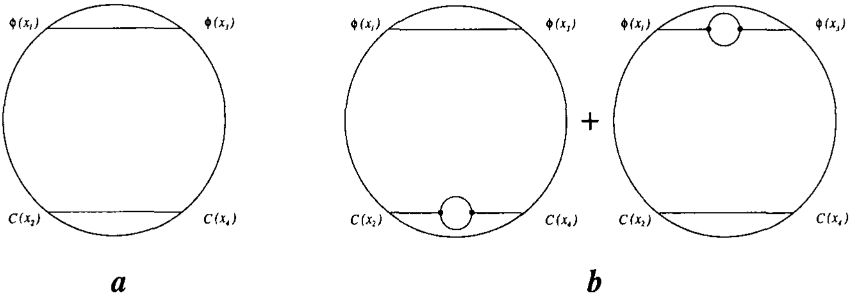


Figure 6. Disconnected contributions to the correlator $\langle \mathcal{O}_\Phi \mathcal{O}_C \mathcal{O}_\Phi \mathcal{O}_C \rangle$ to order $1/N^2$

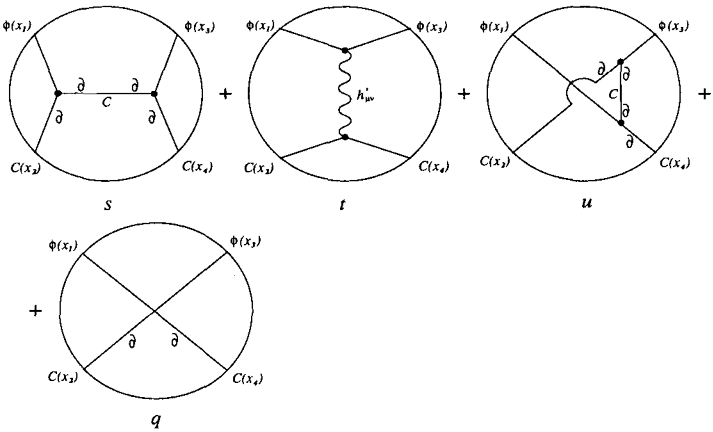


Figure 7. Connected contributions to the correlator $\langle \mathcal{O}_\Phi \mathcal{O}_C \mathcal{O}_\Phi \mathcal{O}_C \rangle$ to order $1/N^2$

This integral is closely related to the momentum space integration of the box graph. In fact, we shall not need this object in all its generality, but may restrict to the case $D_{\Delta\Delta\Delta'\Delta'}$. The calculation in the general case is given in Refs. 133 and 134; see also Ref. 135.

$D_{\Delta_1\Delta_3\Delta_2\Delta_4} =$

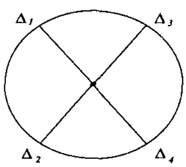


Figure 8. Definition of the contact graph function D

To compute this object explicitly, it is convenient to factor out the overall non-trivial conformal dependence. This may be done by first translating x_1 to 0, then performing an inversion and then translating also x'_3 to 0. The result may be expressed in terms of

$$x \equiv x'_{13} - x'_{14} \qquad y \equiv x'_{13} - x'_{12} \tag{196}$$

and is found to be

$$D_{\Delta\Delta\Delta'\Delta'}(x_i) = x_{12}^{2\Delta'} x_{13}^{2\Delta} x_{14}^{2\Delta'} \times \int_H \frac{d^{d+1}z}{z_0^{d+1}} \frac{z_0^{2\Delta+2\Delta'}}{z^{2\Delta}(z-x)^{2\Delta'}(z-y)^{2\Delta'}} \tag{197}$$

Introducing two Feynman parameters, and carrying out the z-integration, the integral may be re-expressed as

$$\begin{aligned} D_{\Delta\Delta\Delta'\Delta'}(x_i) = & \frac{x_{12}^{2\Delta'} x_{13}^{2\Delta} x_{14}^{2\Delta'}}{(x^2+y^2)^{\Delta'}} \frac{\pi^{d/2}\Gamma(\Delta+\Delta'-d/2)}{2^{\Delta'}\Gamma(\Delta)\Gamma(\Delta')} \\ & \times \int_0^\infty d\rho \int_{-1}^{+1} d\lambda \, \frac{\rho^{\Delta-1}(1-\lambda^2)^{\Delta-1}}{[1+\rho(1-\lambda^2)]^{\Delta}} \frac{1}{(s+\rho+\rho\lambda t)^{\Delta'}} \end{aligned} \tag{198}$$

Remarkably, for positive integers Δ and Δ', the integral for any Δ, Δ' and d may be re-expressed in terms of successive derivatives of a universal function $I(s,t)$,

$$I(s,t) \equiv \int_{-1}^{+1} d\lambda \frac{1}{1+\lambda t - s(1-\lambda^2)} \ln \frac{1+\lambda t}{s(1-\lambda^2)} \tag{199}$$

in the following way,

$$\begin{aligned} D_{\Delta\Delta\Delta'\Delta'}(x_i) = & (-)^{\Delta+\Delta'} \frac{x_{12}^{2\Delta'} x_{13}^{2\Delta} x_{14}^{2\Delta'}}{(x^2+y^2)^{\Delta'}} \frac{\pi^{d/2}\Gamma(\Delta+\Delta'-d/2)}{\Gamma(\Delta)^2\Gamma(\Delta')^2} \\ & \times \left(\frac{\partial}{\partial s}\right)^{\Delta'-1} \left\{ s^{\Delta-1} \left(\frac{\partial}{\partial s}\right)^{\Delta-1} I(s,t) \right\} \end{aligned} \tag{200}$$

While the function $I(s,t)$ is not elementary, its asymptotic behavior is easily obtained.

In the *direct channel* or *t-channel*, we have $|x_{13}| \ll |x_{12}|$ and $|x_{24}| \ll |x_{12}|$, so that we have both $s, t \to 0$. Of principal interest will be the contribution which contains logarithms of s, and this part is given by (for the full asymptotics, Ref. 134); see also Ref. 138:

$$I^{\log},(s,t) = -\ln s \sum_{k=0}^{\infty} a_k(t) s^k, \qquad a_k(t) = \int_{-1}^{+1} d\lambda \frac{(1-\lambda^2)^k}{(1+\lambda t)^{k+1}} \tag{201}$$

In the two *crossed channels*, we have $s \to 1/2$: in the *s-channel* $|x_{12}|, |x_{34}| \ll |x_{13}|$ for which $t \to -1$; in the *u-channel* $|x_{23}|, |x_{14}| \ll |x_{34}|$ for which $t \to +1$. Of principal interest will be the contribution which contains logarithms of $(1 - t^2)$, and this part is given by (for the full asymptotics, see Ref. 134),

$$I^{\log}(s,t) = -\ln(1-t^2)\sum_{k=0}^{\infty}(1-2s)^k\alpha_k(t), \ \alpha_k(t) = \sum_{\ell=0}^{\infty}\frac{\Gamma(\ell+\frac{1}{2})(1-t^2)^\ell}{\Gamma(\frac{1}{2})(2\ell+k+1)\ell!} \tag{202}$$

7.8. *Calculation of the 4-point Exchange Diagrams*

A direct approach to the calculation of the exchange graphs for scalar and gravitons is to insert the scalar or graviton propagators computed previously and then to perform the integrals over the 3-point interaction vertices. This approach was followed in Refs. 78, 133, 134. However, it is also possible to exploit the special space-time properties of conformal symmetry to take a more convenient approach discussed in Ref. 136. This approach consists in first computing the 3-point interaction integral with two boundary-to-bulk propagators (say to vertices 1 and 3) with the bulk-to-bulk propagator between the same interaction vertex and an arbitrary bulk point. Conformal invariance and the assumption of integer dimension $\Delta \geq d/2$ makes this into a very simple object. We shall follow the last method to evaluate the exchange graphs.

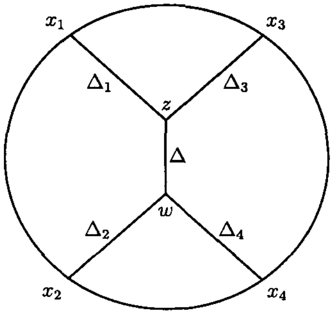


Figure 9. The t-channel exchange graph

For the *scalar exchange diagram*, we need to compute the following integral[viii]

$$A(w,x_1,x_3) = \int_H \frac{d^{d+1}z}{z_0^{d+1}} G_\Delta(w,z) K_{\Delta_1}(z,x_1) K_{\Delta_3}(z,x_3) \tag{203}$$

As in the past, we simplify the integral by using translation invariance to translate x_1 to 0, and then performing an inversion. As a result,

$$A(w,x_1,x_3) = |x_{13}|^{-2\Delta_3} I(w'-x'_{13}), \quad I(w) = \int_H \frac{d^5 z}{z_0^5} G_\Delta(w,z) \frac{z_0^{\Delta_1+\Delta_3}}{z^{2\Delta_3}} \tag{204}$$

We now use the fact that G_Δ is a Green function and satisfies $(\Box_w + \Delta(\Delta - d))G_\Delta(w,z) = \delta(w,z)$, so that

$$(\Box_w + \Delta(\Delta - d))I(w) = \frac{w_0^{\Delta_1+\Delta_3}}{w^{2\Delta_3}} \tag{205}$$

In terms of the scale invariant combination $\zeta = w_0^2/w^2$, we have $I(w) = w_0^{\Delta_{13}} f_S(\zeta)$, $\Delta_{13} = \Delta_1 - \Delta_3$ and the function f_S now satisfies the following differential equation

$$\begin{aligned} 4\zeta^2(\zeta-1)f_S'' + 4\zeta\big[(\Delta_{13}+1)\zeta - \Delta_{13} + d/2 - 1\big]f_S' \\ +(\Delta - \Delta_{13})(\Delta + \Delta_{13} - d)f_S = \zeta^{\Delta_3} \end{aligned} \tag{206}$$

Making the change of variables $\sigma = 1/\zeta$, we find that the new differential equation is manifestly of the hypergeometric type and is solved by

$$f_S(\zeta) = F\left(\frac{\Delta - \Delta_{13}}{2}, \frac{d - \Delta - \Delta_{13}}{2}; \frac{d}{2}; 1 - \frac{1}{\zeta}\right) \tag{207}$$

The other linearly independent solution to the hypergeometric equation is singular as $\zeta \to 1$, which is unacceptable since the original integral was perfectly regular in this limit (which corresponds to $\vec{w} \to 0$).

It is easier, however, to find the solutions in terms of a power series, $f_S(\zeta) = \sum_k f_{Sk}\zeta^k$. Upon substitution into (206), we find solutions that truncate to a finite number of terms in ζ, provided $\Delta_1 + \Delta_3 - \Delta$ is a positive integer. Notice that k need not take integer values, rather $k - \Delta_3$ must be integer. The series truncates from above at $k_{\max} = \Delta_3 - 1$, so that $f_{Sk} = 0$ when $k \geq \Delta_3$, and

$$f_{Sk} = \frac{\Gamma(k)\Gamma(k+\Delta_{13})\Gamma(\frac{1}{2}\{\Delta_1+\Delta_3-\Delta\})\Gamma(\frac{1}{2}\{\Delta+\Delta_1+\Delta_3-d\})}{4\Gamma(\Delta_1)\Gamma(\Delta_3)\Gamma(k+1+\frac{1}{2}\{\Delta_{13}-\Delta\})\Gamma(k+1+\frac{1}{2}\{\Delta_{13}+\Delta-d\})} \tag{208}$$

[viii]In this subsection, we shall not write explicitly the propagator normalization constants C_Δ; however, they will be properly restored in the next subsection.

Still under the assumption that $\Delta_1+\Delta_3-\Delta$ is a positive integer, the series also truncates from below at $k_{\min}=\frac{1}{2}(\Delta-\Delta_{13})$.

It remains to complete the calculation and substitute the above partial result into the full exchange graphs. The required integral is

$$S(x_i)=\int_H dw\sqrt{g}K_{\Delta_2}(w,x_2)K_{\Delta_4}(w,x_4)A(w,x_1,x_3) \tag{209}$$

Remarkably, the expansion terms $w_0^{\Delta_{13}}\zeta^k=w_0^{\Delta_1+\Delta_3+2k}/w^{2k}$ are precisely of the form of the product of two boundary-to-bulk propagators, one with dimension k, the other with dimension $\Delta_{13}+k$. Thus, the scalar exchange diagram may be written as a sum over contact graphs in the following way,

$$S(x_i)=\sum_{k=k_{\min}}^{k_{\max}} f_{Sk}|x_{13}|^{-2\Delta_3+2k}D_{k\,\Delta_{13}+k\,\Delta_2\Delta_4}(x_i) \tag{210}$$

The evaluation of the contact graphs was carried out in the preceding subsection for the special cases $\Delta_1=\Delta_3$ and $\Delta_2=\Delta_4$.

For the *massless graviton exchange diagram*, we need to compute the integral,

$$A_{\mu\nu}(w,x_1,x_2)=\int_H \frac{d^{d+1}z}{z_0^{d+1}}G_{\mu\nu\mu'\nu'}(w,z)T^{\mu'\nu'}(z,x_1,x_3) \tag{211}$$

where the stress tensor is generated by two boundary-to-bulk scalar propagators which we assume both to be of dimension Δ_1,

$$\begin{aligned}T^{\mu'\nu'}(z,x_1,x_3)&=\nabla^{\mu'}K_{\Delta_1}(z,x_1)\nabla^{\nu'}K_{\Delta_1}(z,x_3)\\&-\tfrac{1}{2}g^{\mu'\nu'}\Big[\nabla_{\rho'}K_{\Delta_1}(z,x_1)\nabla^{\rho'}K_{\Delta_1}(z,x_3)\\&\qquad+\Delta_1(\Delta_1-d)K_{\Delta_1}(z,x_1)\nabla^{\rho'}K_{\Delta_1}(z,x_3)\Big]\end{aligned} \tag{212}$$

Under translation of x_1 to 0 and inversion, then using the symmetries of rank 2 symmetric tensors on AdS_5, and finally using the operator W on both sides of Eq. (211), we find

$$\begin{aligned}A_{\mu\nu}(w,x_1,x_3)&=\frac{1}{w^4|x_{13}|^{2\Delta_1}}J_{\mu\kappa}(w)J_{\nu\lambda}(w)I_{\kappa\lambda}(w'-x'_{13})\\I_{\kappa\lambda}(w)&=\left(\frac{\delta_{0\mu}\delta_{0\nu}}{w_0^2}-\frac{1}{d-1}g_{\mu\nu}\right)f_G(\zeta)+\nabla_{(\mu}v_{\nu)}\end{aligned} \tag{213}$$

where the field v_μ represents an immaterial action of a diffeomorphism while the function $f_G(\zeta)$ satisfies the first order differential equation

$$2\zeta(1-\zeta)f'_G(\zeta) - (d-2)f_G(\zeta) = \Delta_1\zeta^{\Delta_1} \tag{214}$$

It is again possible to solve this equation via a power series $f_G(\zeta) = \sum_k f_{Gk}\zeta^k$. The range of k is found to be $d/2 - 1 = k_{\rm min} \leq k \leq k_{\rm max} = \Delta_1 - 1$, provided $\Delta_1 - d/2$ is a non-negative integer and $d > 2$. The coefficients are then given by

$$f_{Gk} = -\frac{\Delta_1\Gamma(k)\Gamma(\Delta_1+1-d/2)}{\Gamma(\Delta_1)\Gamma(k+2-d/2)} \tag{215}$$

The result is particularly simple for the case of interest here when $d = 4$ and Δ_1 integer,

$$f_G(\zeta) = -\frac{\Delta_1}{2\Delta_1-2}(\zeta+\zeta^2+\cdots+\zeta^{\Delta_1-1}) \tag{216}$$

Again, this result may be substituted into the remaining integral in w versus the boundary-to-bulk propagators from the interaction point w to x_2 and x_4, thereby yielding again contributions proportional to contact terms.

7.9. *Structure of Amplitudes*

The full calculations of the graviton exchange amplitudes are quite involved and will not be reproduced completely here.[134] Instead, we quote the contributions to the amplitudes from the correlator $[\mathcal{O}_\phi(x_1)\mathcal{O}_C(x_2)\mathcal{O}_\phi(x_3)\mathcal{O}_C(x_4)]$, where the graviton is exchanged in the t-channel only. The sum of the axion exchange graph I_s in the s-channel, of the axion exchange I_u in the u-channel and of the quartic contact graph I_q is listed separately from the graviton contribution I_g,

$$\begin{aligned}
I_s + I_u + I_q &= \frac{6^4}{\pi^8}\Big[64x_{24}^2 D_{4455} - 32D_{4444}\Big] \qquad (217)\\
I_{\rm grav} &= \frac{6^4}{\pi^8}\Big[8(\frac{1}{s}-2)x_{24}^2 D_{4455} + \frac{64}{9s}\frac{x_{24}^2}{x_{13}^2}D_{3355} + \frac{16}{3s}\frac{x_{24}^2}{x_{13}^4}D_{2255}\\
&\qquad + 18D_{4444} - \frac{46}{9x_{13}^2}D_{3344} - \frac{40}{9x_{13}^4}D_{2244} - \frac{8}{3x_{13}^6}D_{1144}\Big]
\end{aligned}$$

The most interesting information is contained in the power singularity part of this amplitude as well as in the part containing logarithmic singularities. Both are obtained from the singular parts of the universal function $I(s,t)$ in terms of which the contact functions $D_{\Delta_1\Delta_2\Delta_4\Delta_4}$ may be expressed.

7.10. *Power Singularities*

In the s-channel and u-channel, no power singularities occur in the supergravity result. This is consistent with the fact that there are no power singular terms in the OPE of $\mathcal{O}_\phi$ with $\mathcal{O}_C$, since the resulting composite operator would have $U(1)_Y$ hypercharge 4, and the lowest operator with those quantum numbers has dimension 8. (More details on this kind of argument will be given in §7.12.)

In the t-channel, where $|x_{13}|, |x_{24}| \ll |x_{12}|$, we have $s, t \to 0$, with $s \sim t^2$. The power singularities in this channel come entirely from the graviton exchange part, given by

$$I_{\text{grav}}\Big|_{\text{sing}} = \frac{2^{10}}{35\pi^6}\frac{1}{x_{13}^8 x_{24}^8}\left[s(7t^2+6t^4)+s^2(-7+3t^2)-8s^3\right] \tag{218}$$

To compare this behavior with the singularities expected from the OPE, we derive first the behavior of the variables s and t in the t-channel limit,

$$s \sim \frac{x_{13}^2 x_{24}^2}{2x_{12}^4} \qquad\qquad t \sim -\frac{x_{13}\cdot J(x_{12})\cdot x_{24}}{x_{12}^2} \tag{219}$$

where $J_{ij}(x) \equiv \delta_{ij} - 2x_i x_j/x^2$ is the conformal inversion Jacobian tensor. Therefore, the leading singularity in the graviton exchange contribution may be written as

$$I_{\text{grav}}\Big|_{\text{sing}} = \frac{2^6}{5\pi^6}\frac{1}{x_{13}^6 x_{24}^6}\frac{4(x_{13}\cdot J(x_{12})\cdot x_{24})^2 - x_{13}^2 x_{24}^2}{x_{12}^8} \tag{220}$$

with further subleading terms suppressed by additional powers of x_{13}^2/x_{12}^2 and x_{24}^2/x_{12}^2. The leading contribution above describes the exchanges of an operator of dimension 4, whose tensorial structure is that of the stress tensor.

Note that there is also a term corresponding to the exchange of the identity operator, with behavior $x_{13}^{-8}x_{24}^{-8}$, which derives from the disconnected contribution to the correlator in Fig 5 (a). Note that there is no contribution in the singular terms that corresponds to the exchange of an operator of dimension 2. One candidate would be $\mathcal{O}_2$ which is a Lorentz scalar; however, it is a **20′** under $SU(4)_R$, and therefore not allowed in the OPE of two singlets. The other candidate is the Konishi operator, which is both a Lorentz and $SU(4)_R$ singlet. The fact that it is not seen here is consistent with the fact that its dimension becomes very large $\sim \lambda^{1/4}$ in the limit $\lambda \to \infty$ and is dual to a massive string excitation.

7.11. *Logarithmic Singularities*

The logarithmic singularities in the t-channel are produced by both the scalar exchange and contact graphs as well as by the graviton exchange graph.[134] They are given by

$$I_s + I_u + I_q\Big|_{\log} = \frac{960}{\pi^6}\frac{\ln s}{x_{13}^8 x_{24}^8}\sum_{k=0}^{\infty} s^{k+4}(k+1)^2(k+2)^2(k+3)^2(3k+4)a_{k+3}(t)$$

$$I_{\text{grav}}\Big|_{\log} = \frac{24}{\pi^6}\frac{\ln s}{x_{13}^8 x_{24}^8}\sum_{k=0}^{\infty} s^{k+4}\frac{\Gamma(k+4)}{\Gamma(k+1)}\Big\{(k+4)^2(15k^2+55k+42)a_{k+4}(t) - 2(5k^2+20k+16)(3k^2+15k+22)a_{k+3}(t)\Big\} \tag{221}$$

To leading order, these expressions simplify as follows,

$$\begin{aligned} I_s + I_u + I_q\Big|_{\log} &= +\frac{2^7\cdot 3^3}{7\pi^6 x_{12}^{16}}\ln\frac{x_{13}^2 x_{24}^2}{x_{12}^4} \\ I_{\text{grav}}\Big|_{\log} &= -\frac{2^7\cdot 3}{7\pi^6 x_{12}^{16}}\ln\frac{x_{13}^2 x_{24}^2}{x_{12}^4} \end{aligned} \tag{222}$$

Assembling all logarithmic contributions for the various correlators, we get[116]

$$\begin{aligned} \langle \mathcal{O}_\phi\mathcal{O}_\phi\mathcal{O}_\phi\mathcal{O}_\phi\rangle_{\log} &= -\frac{208}{21N^2}\frac{1}{x_{12}^{16}}\ln\frac{x_{13}^2x_{24}^2}{x_{12}^4} && \text{t}-\text{channel} \\ \langle \mathcal{O}_C\mathcal{O}_C\mathcal{O}_C\mathcal{O}_C\rangle_{\log} &= -\frac{208}{21N^2}\frac{1}{x_{12}^{16}}\ln\frac{x_{13}^2x_{24}^2}{x_{12}^4} && \text{t}-\text{channel} \\ \langle \mathcal{O}_\phi\mathcal{O}_C\mathcal{O}_\phi\mathcal{O}_C\rangle_{\log} &= +\frac{128}{21N^2}\frac{1}{x_{12}^{16}}\ln\frac{x_{13}^2x_{24}^2}{x_{12}^4} && \text{t}-\text{channel} \\ \langle \mathcal{O}_\phi\mathcal{O}_C\mathcal{O}_\phi\mathcal{O}_C\rangle_{\log} &= -\frac{8}{N^2}\frac{1}{x_{13}^{16}}\ln\frac{x_{12}^2x_{34}^2}{x_{13}^4} && \text{s}-\text{channel} \end{aligned} \tag{223}$$

Here, the overall coupling constant factor of κ_5^2 has been converted to a factor of $1/N^2$ with the help of the relation $\kappa_5^2 = 4\pi^2/N^2$, a relation that will be explained and justified in (239). Further investigations of these log singularities may be found in Ref. 137.

7.12. *Anomalous Dimension Calculations*

We shall use the supergravity calculations of the 4-point functions for the operators $\mathcal{O}_\phi$ and $\mathcal{O}_C$ to extract anomalous dimensions of double-trace operators built out of linear combinations of $[\mathcal{O}_\phi\mathcal{O}_\phi]$, $[\mathcal{O}_C\mathcal{O}_C]$ and

$[\mathcal{O}_\phi \mathcal{O}_C]$. This was done in Ref. 116. by taking the limits in various channels of the three 4-point functions $\langle \mathcal{O}_\phi(x_1)\mathcal{O}_\phi(x_2)\mathcal{O}_\phi(x_3)\mathcal{O}_\phi(x_4)\rangle$ $\langle \mathcal{O}_C(x_1)\mathcal{O}_C(x_2)\mathcal{O}_C(x_3)\mathcal{O}_C(x_4)\rangle$ and $\langle \mathcal{O}_\phi(x_1)\mathcal{O}_C(x_2)\mathcal{O}_\phi(x_3)\mathcal{O}_C(x_4)\rangle$. For example, we extract the following simple behavior from the *s-channel limit* $x_{12}, x_{34} \to 0$ of the correlator $\langle \mathcal{O}_\phi(x_1)\mathcal{O}_C(x_2)\mathcal{O}_\phi(x_3)\mathcal{O}_C(x_4)\rangle$,

$$\mathcal{O}_\phi(x_1)\mathcal{O}_C(x_2) = A_{\phi c}(x_{12}\mu)[\mathcal{O}_\phi \mathcal{O}_C]_\mu(x_2) + \cdots \tag{224}$$

where μ is an arbitrary renormalization scale for the composite operators and $A_{\phi c}$ is the corresponding logarithmic coefficient function, whose precise value in the large N, large λ limit is available from the logarithmic singularities of the correlator, and is given by

$$A_{\phi c}(x_{12}\mu) = 1 - \frac{16}{N^2}\ln(x_{12}\mu) \tag{225}$$

This leading behavior receives further corrections both in inverse powers of N and λ.

From the *t-channel* and *u-channel* of the same correlators, we extract the leading terms in the OPE of two $\mathcal{O}_\phi$'s and of two $\mathcal{O}_C$'s as follows,

$$\begin{aligned} &\mathcal{O}_\phi(x_1)\mathcal{O}_\phi(x_3) \\ &\quad = S(x_1,x_3) + C_{\phi\phi}[\mathcal{O}_\phi\mathcal{O}_\phi]_\mu + C_{\phi c}[\mathcal{O}_C\mathcal{O}_C]_\mu + C_{\phi T}[TT]_\mu + \cdots \\ &\mathcal{O}_C(x_1)\mathcal{O}_C(x_3) \\ &\quad = S(x_1,x_3) + C_{c\phi}[\mathcal{O}_\phi\mathcal{O}_\phi]_\mu + C_{cc}[\mathcal{O}_C\mathcal{O}_C]_\mu + C_{cT}[TT]_\mu + \cdots \end{aligned} \tag{226}$$

where the term $S(x_1, x_3)$ contains all the *power singular terms* in the expansion, and is given schematically by

$$S(x_1,x_3) = \frac{I}{x_{13}^8} + \frac{T(x_3)}{x_{13}^4} + \frac{\partial T(x_3)}{x_{13}^3} + \frac{\partial\partial T(x_3)}{x_{13}^2} + \frac{\partial\partial\partial T(x_3)}{x_{13}} \tag{227}$$

The coefficient functions may be extracted from the logarithmic behavior as before,

$$\begin{aligned} C_{\phi\phi} &= C_{cc} = 1 - \frac{208}{21N^2}\ln(x_{13}\mu) \\ C_{\phi c} &= C_{c\phi} = 1 + \frac{128}{21N^2}\ln(x_{13}\mu) \end{aligned} \tag{228}$$

Unfortunately, the coefficient functions $C_{\phi T}$ and C_{cT} are not known at this time as their evaluation would involve the highly complicated calculation involving two external stress tensor insertions.

To make progress, we make use of a continuous symmetry of supergravity, namely $U(1)_Y$ *hypercharge invariance.* Most important for us here is that the operator

$$\mathcal{O}_B \equiv \frac{1}{\sqrt{2}}\{\mathcal{O}_\phi + i\mathcal{O}_C\} \tag{229}$$

has hypercharge $Y = 2$, which is the unique highest values attained amongst the canonical supergravity fields, as may be seen from the Table 7. We may now re-organize the OPE's of operators $\mathcal{O}_\phi$ and $\mathcal{O}_C$ in terms of $\mathcal{O}_B$ and $\mathcal{O}_B^*$. The OPE of $\mathcal{O}_B$ with $\mathcal{O}_B^*$ contains the identity operator, the stress tensor and its derivatives and powers, as well as the $Y = 0$ operator $[\mathcal{O}_B\mathcal{O}_B^*]$,

$$\mathcal{O}_B(x_1)\mathcal{O}_B^*(x_2) = S(x_1,x_2) + C_{BT}[TT]_\mu + C_{BB^*}[\mathcal{O}_B\mathcal{O}_B^*]_\mu + \cdots \tag{230}$$

while the $Y = 4$ channel of the OPE is given by

$$\mathcal{O}_B(x_1)\mathcal{O}_B(x_2) = (C_{\phi\phi} - C_{\phi c})\mathrm{Re}[\mathcal{O}_B\mathcal{O}_B]_\mu + iA_{\phi c}\mathrm{Im}[\mathcal{O}_B\mathcal{O}_B]_\mu \tag{231}$$

Since the smallest dimensional operator of hypercharge $Y = 4$ is the composite $[\mathcal{O}_B\mathcal{O}_B]$, we see that the power singularity terms $S(x_1,x_3)$ indeed had to be the same for both OPE's in (226). By the same token, the rhs of (231) must be proportional to $[\mathcal{O}_B\mathcal{O}_B]_\mu$, so we must have $C_{\phi\phi} - C_{\phi c} = A_{\phi c}$, which is indeed borne out by the explicit calculational results of (225) and (228). In summary, we have a single simple OPE

$$\mathcal{O}_B(x_1)\mathcal{O}_B(x_2) = A_{\phi c}(x_{12}\mu)[\mathcal{O}_B\mathcal{O}_B]_\mu + \cdots \tag{232}$$

from which the anomalous dimension may be found to be[116]

$$\gamma_{[\mathcal{O}_B\mathcal{O}_B]} = \gamma_{[\mathcal{O}_B^*\mathcal{O}_B^*]} = -16/N^2 \tag{233}$$

There is another operator occurring in this OPE channel of which we know the anomalous dimension. Indeed, the double-trace operator $[\mathcal{O}_2\mathcal{O}_2]_{105}$ is 1/2 BPS, and thus has vanishing anomalous dimension. Its maximal descendant $Q^4\bar{Q}^4[\mathcal{O}_2\mathcal{O}_2]_{105}$ therefore has $Y = 0$ and unrenormalized dimension 8. For more on the role of the $U(1)_Y$ symmetry, see Ref. 117. The study of the OPE via the 4-point function has also revealed some surprising non-renormalization effects, not directly related to the BPS nature of the intermediate operators. In the OPE of two half-BPS **20'** operators, for example, the **20'** intermediate state is *not* chiral. Yet, to lowest order at strong coupling, its dimension was found to be protected; see Ref. 118 and.[119]

7.13. *Check of N-dependence*

The prediction for the anomalous dimension of the operator $[\mathcal{O}_B\mathcal{O}_B]$ to order $1/N^2$ obtained from supergravity calculations holds for infinitely large values of the 't Hooft coupling $\lambda = g_{YM}^2 N$ on the SYM side. As the regimes of couplings for possible direct calculations do not overlap, we cannot directly compare this prediction with a calculation on the SYM side. However, it is very illuminating to reproduce the $1/N^2$ dependence of the anomalous dimension from standard large N counting rules in SYM theory.[116] We proceed by expansing $\mathcal{N} = 4$ SYM in $1/N$, while keeping the 't Hooft coupling fixed (and perturbatively small). The strategy will be to isolate the general structure of the expansion and then to seek the limit where $\lambda \to \infty$.

To be concrete, we study the correlator $\langle \mathcal{O}_\phi \mathcal{O}_c \mathcal{O}_\phi \mathcal{O}_c \rangle$, though our results will apply generally.

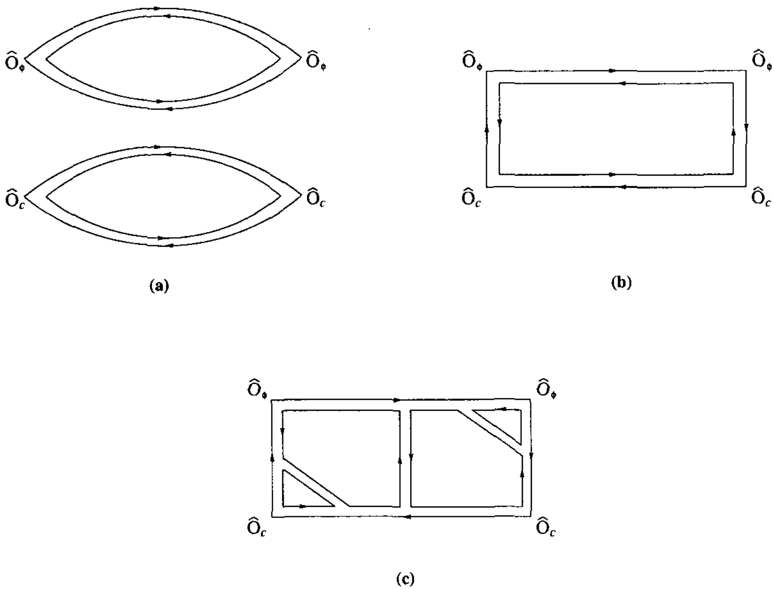


Figure 10. Large N counting for the 4-point function

First, we normalize the individual operators via their 2-point functions, which to leading order in large N requires

$$\mathcal{O}_c = \frac{1}{N}\mathrm{tr}F_{\mu\nu}\tilde{F}^{\mu\nu} + \cdots \qquad \mathcal{O}_\phi = \frac{1}{N}\mathrm{tr}F_{\mu\nu}F^{\mu\nu} + \cdots \tag{234}$$

In computing the 4-point function $\langle \mathcal{O}_\phi \mathcal{O}_c \mathcal{O}_\phi \mathcal{O}_c \rangle$, there will first be a *disconnected contribution* of the form $\langle \mathcal{O}_\phi \mathcal{O}_\phi \rangle \langle \mathcal{O}_c \mathcal{O}_c \rangle$, which thus contributes precisely to order N^0. The simplest connected contribution is a Born graph with a single gluon loop; the operator normalizations contribute N^{-4}, while the two color loops contribute N^2, thereby suppressing the connected contribution by a factor of N^{-2} compared to the disconnected one. This Born graph has no logarithms because it is simply the product of 4 propagators. Perturbative corrections with internal interaction vertices will however generate logarithmic corrections, and thus contributions to anomalous dimensions. In the large N limit, planar graphs will dominate and the only corrections are due to non-trivial λ-dependence with the same expansion order N^{-2} and the connected contribution will take the form

$$\langle \mathcal{O}_\phi \mathcal{O}_c \mathcal{O}_\phi \mathcal{O}_c \rangle_{\text{conn}} = \frac{1}{N^2} f(\lambda) + O(\frac{1}{N^4}) \tag{235}$$

For the anomalous dimensions, a similar expansion will hold,

$$\gamma(N, \lambda) = \frac{1}{N^2} \bar{\gamma}(\lambda) + O(\frac{1}{N^4}) \tag{236}$$

The above results were established perturbatively in the 't Hooft coupling. To compare with the supergravity results, f and $\bar{\gamma}$ should admit well-behaved $\lambda \to \infty$ limits. Our supergravity calculation in fact established that $\bar{\gamma}_{[\mathcal{O}_B \mathcal{O}_B]}(\lambda = \infty) = -16$, a result that could of course not have been gotten from Feynman diagrams in SYM theory.

The calculation of AdS four point functions in weak coupling perturbation theory was carried out in Refs. 139 and 140; string corrections to 4-point functions were considered early on in Refs. 141 and 142; further 4-point function calculations in the AdS setting may be found in Refs. 144, 143, and 145. More general correlators of 4-point functions and higher corresponding to the insertion of currents and tensor forms may be found in Refs. 149, 147, 148, and 150. Finally, an approach to correlation functions based on the existence of a higher spin field theory in Anti-de Sitter spacetime may be found in a series of papers.[151] Finally, effects of instantons on SYM and AdS/CFT correlators were explored recently in Refs. 156, 152, 154, 153, 157, and 155. Possible constraints on correlators in AdS/CFT and $\mathcal{N} = 4$ SYM from S-duality have been investigated by Ref. 158. Finally, very recently, correlators have been evaluated exactly for strings propagating in AdS_3 in Ref. 232.

8. How to Calculate CFT_d Correlation Functions from AdS_{d+1} Gravity

The main purpose of this chapter is to discuss the techniques used to calculate correlation functions in $\mathcal{N} = 4$, $d = 4$ SYM field theory from Type IIB $D = 10$ supergravity. We will begin with a quick summary of the basic ideas of the correspondence between the two theories. These were discussed in more detail in earlier sections, but we wish to make this chapter self-contained. Other reviews we recommend to readers are the broad treatment of Ref. 7 and the 1999 TASI lectures of Klebanov[159] in which the AdS/CFT correspondence is motivated from the viewpoints of D-brane and black hole solutions, entropy and absorption cross-sections.

The $\mathcal{N} = 4$ SYM field theory is a 4-dimensional gauge theory with gauge group $SU(N)$ and R-symmetry or global symmetry group $SO(6) \sim SU(4)$. Elementary fields are all in the adjoint representation of $SU(N)$ and are represented by traceless Hermitean $N \times N$ matrices. There are 6 elementary scalars $X^i(x)$, 4 fermions $\psi^a(x)$, and the gauge potential $A_j(x)$. The theory contains a unique coupling constant, the gauge coupling g_{YM}. It is known that the only divergences of elementary Green's functions are those of wave function renormalization which is unobservable and gauge-dependent. The β-function $\beta(g_{YM})$ vanishes, so the theory is conformal invariant. The bosonic symmetry group of the theory is the direct product of the conformal group $SO(2,4) \sim SU(2,2)$ and the R-symmetry $SU(4)$. These combine with 16 Poincaré and 16 conformal supercharges to give the superalgebra $SU(2,2|4)$ which is the over-arching symmetry of the theory.

Observables in a gauge theory must be gauge-invariant quantities, such as:

(1) Correlation functions of gauge invariant local composite operators — the subject on which we focus,
(2) Wilson loops — not to be discussed, See, for example, 160, 161, 162.

Our primary interest is in correlation functions of the chiral primary operators[ix]

$$\text{tr}X^k \equiv N^{\frac{1-k}{2}} \text{tr}\left(X^{\{i_1} X^{i_2} ... X^{i_k\}}\right) - \text{traces} \tag{237}$$

[ix] the normalization factor $N^{\frac{1-k}{2}}$ is chosen so that all correlation functions of these operators are of order N^2 for large N.

These operators transform as rank k symmetric traceless $SO(6)$ tensors – irreducible representations whose Dynkin designation is $[0, k, 0]$. For $k = 2, 3, 4$ the dimensions of these representations are $20, 50, 105$, respectively.

Ex. 1:. What is the dimension of the [0,5,0] representation?

The $\mathrm{tr}X^k$ are lowest weight states of **short** representations of $SU(2,2|4)$. The condition for a short representation is the relation $\Delta_{\mathrm{tr}X^k} = k$ between scale dimension Δ and $SO(6)$ rank. Since the latter must be an integer, the former is quantized. The scale dimension of chiral primary operators (and all descendents) is said to be "protected" It is given for all g_{YM} by its free-field value (i.e. the value at $g_{YM} = 0$). This is to be contrasted with the many composite operators which belong to **long** representations of $SU(2,2|4)$. For example, the Konishi operator $K(x) = \mathrm{tr}[X^i X^i]$ is the primary of a long representation. In the weak coupling limit, it is known[163] that $\Delta_K = 2 + 3g_{YM}^2 N/4\pi^2 + \mathcal{O}(g_{YM}^4)$. The existence of a gauge invariant operator with anomalous dimension is one sign that the field theory is non-trivial, not a cleverly disguised free theory.

In Sec. 3.4 it was discussed how $SU(2,2|4)$ representations are "filled out" with descendent states obtained by applying SUSY generators with $\Delta = \frac{1}{2}$ to the primaries. Descendents can be important. For example, the descendents of the lowest chiral primary $\mathrm{tr}X^2$ include the 15 SO(6) currents, the 4 supercurrents, and the stress tensor.

Some years ago 't Hooft taught us (for a review, see Ref. 164) that it is useful to express amplitudes in an $SU(N)$ gauge theory in terms of N and the 't Hooft coupling $\lambda = g_{YM}^2 N$. Any Feynman diagram can be redrawn as a sum of color-flow diagrams with definite Euler character χ (in the sense of graph theory). n-point functions of the operators $\mathrm{tr}X^k$ are of the form

$$N^\chi F(\lambda, x_i) = N^\chi [f_0(x_i) + \lambda f_1(x_i) + \cdots] \tag{238}$$

The right side shows the beginning of a weak coupling expansion. One can see that planar diagrams (those with $\chi = 2$) dominate in the large N-limit.

The extremely remarkable fact of the AdS/CFT correspondence is that the planar contribution to n-point correlation functions of operators $\mathrm{tr}X^k$ and descendents can be calculated (in the limit $N \to \infty$, $\lambda >> 1$) from **classical** supergravity, a strong coupling limit of a QFT_4 without gravity from classical calculations in a $D = 5$ gravity theory. Results are interpreted as the sum of the series in (238). Information about operators in long representations can be obtained by including string scale effects. It is known that their scale dimensions are of order $\lambda^{\frac{1}{4}}$ in the limit above. They decouple from supergravity correlators.

This claim brings us to the supergravity side of the duality, namely to type IIB, $D = 10$ supergravity which has the product space-time $\mathrm{AdS}_5 \times \mathrm{S}^5$ as a classical "vacuum solution". The first hint of some relation to $\mathcal{N} = 4$ SYM theory is the match of the isometry group $SO(2,4) \times SO(6)$ with the conformal and R-symmetry groups of the field theory. The vacuum solution is also invariant under $16 + 16$ supercharges and thus has the same $SU(2,2|4)$ superalgebra as the field theory.

Type IIB supergravity is a complicated theory whose structure was discussed in Secs. 4.4 and 4.5. Here we describe only the essential points necessary to understand the correspondence with $\mathcal{N} = 4$ SYM theory. Since the supergravity theory is the low energy limit of IIB string theory, the 10D gravitational coupling may be expressed in terms of the dimensionless string coupling g_s and the string scale α' (of dimension l^2). The relation is $\kappa_{10}^2 = 8\pi G_{10} = 64\pi^7 g_s^2 \alpha'^4$. The length scale of the AdS_5 and S_5 factors of the vacuum space-time is L with $L^4 = 4\pi\alpha'^2 g_s N$. The integer N is determined by the flux of the self-dual 5−form field strength on S_5. The volume of S_5 is $\pi^3 L^3$ so the effective $5D$ gravitational constant is

$$\frac{\kappa_5^2}{8\pi} = G_5 = \frac{G_{10}}{\mathrm{Vol}(S_5)} = \frac{\pi L^3}{2N^2} \tag{239}$$

Among the bosonic fields of the theory, we single out the 10D metric g_{MN} and 5-form F_{MNPQR}, which participate in the vacuum solution, and the dilaton ϕ and axion C. Other fields consistently decouple from these and the subsystem is governed by the truncated action (in Einstein frame)

$$\begin{aligned} S_{\mathrm{IIB}} = \tfrac{1}{16\pi G_{10}} \int d^{10}z \sqrt{g_{10}} \{ R_{10} - \tfrac{1}{2\cdot 5!} F_{MNPQR} F^{MNPQR} \\ - \tfrac{1}{2} \partial_M \phi \partial^M \phi - \tfrac{1}{2} e^{2\phi} \partial_M C \partial^M C \} \end{aligned} \tag{240}$$

Actually there is no covariant action which gives the self-dual relation $F_5 = *F_5$ as an Euler-Lagrange equation, and the field equations from S_{IIB} must be supplemented by this extra condition.

Using x^i, $i = 0,1,2,3$ as Cartesian coordinates of Minkowski space with metric $\eta_{ij} = (-+++)$ and y^a, $a = 1,2,3,4,5,6$ as coordinates of a flat transverse space, we write the following ansatz for the set of fields above:

$$\begin{aligned} & ds_{10}^2 = \tfrac{1}{\sqrt{H(y^a)}} \eta_{ij} dx^i dx^j + \sqrt{H(y^a)} \delta_{ab} dy^a dy^b \\ & F = dA + *dA \qquad\qquad A = \tfrac{1}{H(y^a)} dx^0 \wedge dx^1 \wedge dx^2 \wedge dx^3 \\ & \phi = C \equiv 0 \end{aligned} \tag{241}$$

Remarkably the configuration above is a solution of the equations of motion provided that $H(y^a)$ is a harmonic function of y^a, *i.e.*

$$\sum_{a=1}^{6} \frac{\partial^2}{\partial y^a \partial y^a} H = 0 \tag{242}$$

Ex. 2:. Verify that the above is a solution. Compute the connection and curvature of the metric as an intermediate step. See the discussion of the Cartan structure equations in Section 9 for some guidance.

The appearance of harmonic functions is typical of D-brane solutions to supergravity theories. The solutions (241) are $\frac{1}{2}-BPS$ solutions which support 16 conserved supercharges. This fact may be derived by studying the transformation rules of Type IIB supergravity to find the Killing spinors. A quite general harmonic function is given by

$$H = 1 + \sum_{I=1}^{M} \frac{L_I^4}{(y - y_I)^4} \qquad\qquad L_I^4 = 4\pi\alpha'^2 g_s N_I \tag{243}$$

This describes a collection of M parallel stacks of $D3$-branes, with N_I branes located at position $y^a = y_I^a$ in the transverse space. This "multi-center" solution of IIB supergravity defines a 10-dimensional manifold with M infinitely long throats as $y \to y_I$ and which is asymptotically flat as $y \to \infty$. The curvature invariants are non-singular as $y \to y_I$, and these loci are simply degenerate horizons. The solution has an AdS/CFT interpretation as the dual of a Higgs branch vacuum state of $\mathcal{N} = 4$ SYM theory, a vacuum in which conformal symmetry is spontaneously broken. However, we are jumping too far ahead.

Let's consider the simplest case of a single stack of N D3-branes at $y_I = 0$. We replace the y^a by a radial coordinate $r = \sqrt{y^a y^a}$ plus 5 angular coordinates y^α on an S_5. At the same time we take the near-horizon limit. The physical and mathematical arguments for this limit are rather complex and discussed in Sec 5.2 above, in Ref. 7 and elsewhere. We simply state that it is the throat region of the geometry that determines the physics of AdS/CFT. We therefore restrict to the throat simply by dropping the 1 in the harmonic function $H(r)$. Thus we have the metric

$$ds_{10}^2 = \frac{r^2}{L^2} \eta_{ij} dx^i dx^j + \frac{L^2 dr^2}{r^2} + L^2 d\Omega_5^2 \tag{244}$$

where $d\Omega_5^2$ is the $SO(6)$ invariant metric on the unit S_5. The metric describes the product space $\mathrm{AdS}_5 \times \mathrm{S}^5$. The coordinates (x^i, r) are collectively called z_μ below. These coordinates give the Poincaré patch of the induced metric on the hyperboloid embedded in 6-dimensions.[7]

$$Y_0^2 + Y_5^2 - Y_1^2 - Y_2^2 - Y_3^2 - Y_4^2 = L^2 \tag{245}$$

Ex. 3:. Show that the curvature tensor in the z_μ directions has the maximal symmetric form $R_{\mu\nu\rho\lambda} = -\frac{1}{L^2}(g_{\mu\rho}g_{\nu\lambda} - g_{\mu\lambda}g_{\nu\rho})$.

The bulk theory may now be viewed as a supergravity theory in the AdS$_5$ space-time with an infinite number of 5D fields obtained by Kaluza-Klein analysis on the internal space S_5. We will discuss the KK decomposition process and the properties of the 5D fields obtained from it. The main point is to emphasize the 1:1 correspondence between these bulk fields and the composite operators of the $\mathcal{N} = 4$ SYM theory discussed above.

The linearized field equations of fluctuations about the background (244) were analyzed in Ref. 50. All fields of the $D = 10$ theory are expressed as series expansions in appropriate spherical harmonics on S_5. Typically the independent 5D fields are mixtures of KK modes from different 10D fields. For example the scalar fields which correspond to the chiral primary operators are superpositions of the trace h_α^α of metric fluctuations on S_5 with the S_5 components of the 4-form potential $A_{\alpha\beta\gamma\delta}$. The independent 5D fields transform in representations of the isometry group $SU(4) \sim SO(6)$ of S_5 which are determined by the spherical harmonics.

The analysis of Ref. 50 leads to a graviton multiplet plus an infinite set of KK excitations. We list the fields of the graviton multiplet, together with the dimensionalities of the corresponding $SO(6)$ representations: graviton $h_{\mu\nu}$, 1, gravitini ψ_μ, $\mathbf{4} \oplus \mathbf{4}^*$, 2-form potentials $A_{\mu\nu}$, $\mathbf{6}_c$, gauge potentials A_μ, $\mathbf{15}$, spinors λ, $\mathbf{4} \oplus \mathbf{4}^* \oplus \mathbf{20} \oplus \mathbf{20}^* = 48$, and finally scalars ϕ, $\mathbf{20}' \oplus \mathbf{10} \oplus \mathbf{10}^* \oplus \mathbf{1}_c = 42$. In this notation $\mathbf{10}^*$ denotes the conjugate of the complex irrep $\mathbf{10}$, while $\mathbf{6}_c$ denotes a doubling of the real 6-dimensional (defining) representation of $SO(6)$.

Each of these fields is the base of a KK tower. For the scalar primaries one effectively has the following expansion, after mixing is implemented,

$$\phi(z,y) = \sum_{k=2}^{\infty} \phi_k(z) Y^k(y) \tag{246}$$

Here $Y^k(y)$ denotes a spherical harmonic of rank k, so that $\phi_k(z)$ is a scalar field on AdS$_5$ which transforms[x] in the $[0,k,0]$ irrep of $SO(6)$. In

[x]indices for components of this irrep are omitted on both ϕ_k and Y^k.

the same way that every scalar field on Minkowski space contains an infinite number of momentum modes, each ϕ_k contains an infinite number of modes classified in a unitary irreducible representation of the AdS_5 isometry group $SO(2,4)$. We will describe these irreps briefly. For more information, see Refs. 24, 165–167. The group has maximal compact subgroup $SO(2) \times SO(4)$ and irreps are denoted by (Δ, s, s'). The generator of the $SO(2)$ factor is identified with the energy in the physical setting, and Δ is the lowest energy eigenvalue that occurs in the representation. The quantum numbers s, s' designate the irrep of $SO(4)$ in which the lowest energy components transform. Unitarity requires the bounds

$$\Delta \geq 2+s+s' \qquad \text{if}\, ss'>0 \qquad \Delta \geq 1+s \qquad \text{if}\, s'=0. \tag{247}$$

In general Δ need not be integer, but our KK scalars ϕ_k transform in the irrep $[0, \Delta = k, 0]$ in which the energy and internal symmetry eigenvalues are locked, a condition which gives a short representation of $SU(2,2|4)$.

Each $\phi_k(z)$ satisfies an equation of motion of the form

$$(\Box_{AdS} - M^2)\phi_k = \text{ nonlinear interaction terms} \tag{248}$$

The symbol $\Box$ is the invariant Laplacian on AdS_5,

Ex. 4:. Obtain its explicit form from the metric in (244).

Each KK mode has a definite mass $M^2 = m^2/L^2$ and the dimensionless m^2 is essentially determined by $SO(6)$ group theory[xi] to be $m^2 = k(k-4)$. Formulas of this type are important in the AdS/CFT correspondence, because the energy quantum number, $\Delta = k$ in this case, is identified with the scale dimension of the dual operator in the $\mathcal{N}=4$ SYM theory. Later we will see how this occurs.

Since the superalgebra $SU(2,2|4)$ operates in the dimensionally reduced bulk theory all KK modes obtained in the decomposition process can be classified in representations of $SU(2,2|4)$. It turns out that one gets exactly the set of short representation discussed above for the composite operators of the field theory. There is thus a 1:1 correspondence between the KK fields of Type IIB $D=10$ supergravity and the composite operators (in short representations) of $\mathcal{N}=4$ SYM theory. The ϕ_k we have been discussing

[xi]In the simplest case of the dilaton field, whose linearized 10D field equation is uncoupled, the masses in the KK decomposition are simply given by the eigenvalues of the Laplacian on S_5, namely $m^2 = k(k+4)$. The mass formula which follows differs because of the mixing discussed above.

are dual to the chiral primary operators $\mathrm{tr}X^k$. Within the lowest $k = 2$ multiplet, the 15 bulk gauge fields A_μ are dual to the conserved currents $\mathcal{J}_i$ of the $SO(6)$ R-symmetry group, and the AdS$_5$ metric fluctuation $h_{\mu\nu}$ is dual to the field theory stress tensor T_{ij}.

Critics of the AdS/CFT correspondence legitimately ask whether results are due to dynamics or simply to symmetries. It thus must be admitted that the operator duality just discussed was essentially ensured by symmetry. The superalgebra representations which can occur in the KK reduction of a gravity theory whose "highest spin" field is the metric tensor g_{MN} are strongly constrained. In the present case of $SU(2,2|4)$ there was no choice but to obtain the series of short representations which were found. So what we have uncovered so far is just the working of the same symmetry algebra in two different physical settings, a field theory without gravity in 4 dimensions and a gravity theory in 5 dimensions. The more dynamical aspects of the correspondence involve the interactions of the dimensionally reduced bulk theory, *e.g.* the nonlinear terms in (248). It is notoriously difficult to find these terms,[xii] but fortunately enough information has been obtained to give highly non-trivial tests of AdS/CFT, some of which are discussed later.

8.1. *AdS$_{d+1}$ Basics—Geometry and Isometries*

We now begin our discussion of how to obtain information on correlation functions in conformal field theory from classical gravity. For applications to the "realistic" case of $\mathcal{N} = 4$ SYM theory, we will need details of Type IIB supergravity, but we can learn a lot from toy models of the bulk dynamics. In most cases we will use Euclidean signature models in order to simplify the discussions and calculations.

Consider the Euclidean signature gravitational action in $d+1$ dimensions

$$S = \frac{-1}{16\pi G} \int d^{d+1}z\sqrt{g}(R - \Lambda) \tag{249}$$

with $\Lambda = -d(d-1)/L^2$. The maximally symmetric solution is Euclidean AdS$_{d+1}$ which should be more properly called the hyperbolic space H_{d+1}. The metric can be presented in various coordinate systems, each of which brings out different features. For now we will use the upper half-space

[xii]Except in subsectors such as that of the 15 A_μ where non-abelian gauge invariance in 5 dimensions governs the situation.

description

$$
\begin{aligned}
ds^2 &= \tfrac{L^2}{z_0^2}(dz_0^2 + \textstyle\sum_{i=1}^d dz_i^2) \\
&= \bar{g}_{\mu\nu} dz^\mu dz^\nu
\end{aligned}
\tag{250}
$$

Ex. 5:. Calculate the curvatures $R_{\mu\nu} = \frac{-d}{L^2}\bar{g}_{\mu\nu}$, $R = \frac{-d(d+1)}{L^2}$.

The space is conformally flat and one may think of the coordinates as a $(d+1)$-dimensional Cartesian vector which we will variously denote as $z_\mu = (z_0, z_i) = (z_0, \vec{z})$, with $z_0 > 0$. Scalar products $z \cdot w$ and invariant squares z^2 involve a sum over all $d+1$ components, *e.g.* $z \cdot w = \delta^{\mu\nu} z_\mu w_\nu$.

The plane $z_0 = 0$ is at infinite geodesic distance from any interior point. Yet it is technically a boundary. Data must be specified there to obtain unique solutions of wave equations on the spacetime, as we will see later. We will usually set the scale $L = 1$. Equivalently, all dimensionful quantities are measured in units set by L.

The continuous isometry group of Euclidean AdS_{d+1} is $SO(d+1,1)$. This consists of rotations and translations of the z_i with $\frac{1}{2}d(d-1) + d$ parameters, scale transformations $z_\mu \to \lambda z_\mu$ with 1 parameter, and special conformal transformations whose infinitesimal form is $\delta z_\mu = 2c \cdot z z_\mu - z^2 c_\mu$, with $c_\mu = (0, c_i)$ and thus d parameters. The total number of parameters is $(d+2)(d+1)/2$ which is the dimension of the group $SO(d+1,1)$.

Ex. 6:. Verify explicitly the Killing condition $D_\mu K_\nu + D_\nu K_\mu = 0$ for all infinitesimal transformations. The covariant derivative D_μ includes the Christoffel connection for the metric (250).

Ex. 7:. (Extra credit !) Since $\mathrm{AdS}_{d+1}$is conformally flat, it has the same conformal group $SO(d+2,1)$ as flat $(d+1)$-dimensional Euclidean space. There are $d+2$ additional conformal Killing vectors $\bar{K}_\mu$ which satisfy $D_\mu \bar{K}_\nu + D_\nu \bar{K}_\mu - \frac{2}{d+1}\bar{g}_{\mu\nu} D^\rho \bar{K}_\rho = 0$. Find them!

The AdS_{d+1}space also has the important **discrete** isometry of **inversion**. We will discuss this in some detail because it has applications to the computation of AdS/CFT correlation functions and in conformal field theory itself. Under inversion the coordinates z_μ transform to new coordinates z'_μ by $z_\mu = z'_\mu / z'^2$, and it is not hard to show that the line element (250) is invariant under this transformation.

Ex. 8:. Show this explicitly.

Inversion is also a discrete conformal isometry of flat Euclidean space.

The Jacobian of the transformation is also useful,

$$\frac{\partial z_\mu}{\partial z'_\nu} = \frac{1}{z'^2} J_{\mu\nu}(z) \tag{251}$$
$$J_{\mu\nu}(z) = J_{\mu\nu}(z') = \delta_{\mu\nu} - \frac{2z_\mu z_\nu}{z^2}$$

The Jacobian tells us how a tangent vector of the manifold transforms under inversion.

Ex. 9:. View $J_{\mu\nu}(z)$ as a matrix. Show that it satisfies $J_{\mu\rho}(z)J_{\rho\nu}(z) = \delta_{\mu\nu}$ and has d eigenvalues $+1$ and 1 eigenvalue -1.

$J_{\mu\nu}$ is thus a matrix of the group $O(d+1)$ which is not in the proper subgroup $SO(d+1)$. As an isometry, inversion is an improper reflection which cannot be continuously connected to the identity in $SO(d+1,1)$.

Ex. 10:. (Important but tedious !) Let z_μ, w_μ denote two vectors with z'_μ, w'_μ their images under inversion. Show that

$$\frac{1}{(z-w)^2} = \frac{(z')^2\,(w')^2}{(z'-w')^2} \tag{252}$$

$$J_{\mu\nu}(z-w) = J_{\mu\mu'}(z')J_{\mu'\nu'}(z'-w')J_{\nu'\nu}(w') \tag{253}$$

8.2. *Inversion and CFT Correlation Functions*

Although we have derived the properties of inversion in the context of AdS_{d+1}, the manipulations are essentially the same for flat d-dimensional Euclidean space. We simply replace z_μ, w_μ by d-vectors x_i, y_i and take $x_i = \frac{x'_i}{x'^2}$, etc. Inversion is now a conformal isometry and in most cases [xiii] a symmetry of CFT_d. Under the inversion $x_i \to x'_i$, a scalar operator of scale dimension Δ is transformed as $\mathcal{O}_\Delta(x) \to \mathcal{O}'_\Delta(x) = x'^{2\Delta}\mathcal{O}_\Delta(x')$. Correlation functions then transform covariantly under inversion, viz.

$$\begin{aligned}&\langle \mathcal{O}_{\Delta_1}(x_1)\mathcal{O}_{\Delta_2}(x_2)\cdots\mathcal{O}_{\Delta_n}(x_n)\rangle \\ &\quad = (x'_1)^{2\Delta_1}(x'_2)^{2\Delta_2}\cdots(x'_n)^{2\Delta_n}\langle \mathcal{O}_{\Delta_1}(x'_1)\mathcal{O}_{\Delta_2}(x'_2)\cdots\mathcal{O}_{\Delta_n}(x'_n)\rangle\end{aligned} \tag{254}$$

It is well known that the spacetime forms of 2- and 3-point functions

[xiii] Inversion is an improper reflection similar to parity and is not always a symmetry of a field theory action containing fermions.

are unique in any CFT_d, a fact which can be established using the transformation law under inversion. These forms are

$$\langle \mathcal{O}_\Delta(x)\mathcal{O}_{\Delta'}(y)\rangle = \frac{c\delta_{\Delta\Delta'}}{(x-y)^{2\Delta}} \tag{255}$$

$$\langle \mathcal{O}_{\Delta_1}(x)\mathcal{O}_{\Delta_2}(y)\mathcal{O}_{\Delta_3}(3)\rangle = \frac{\tilde{c}}{(x-y)^{\Delta_{12}}(y-z)^{\Delta_{23}}(z-x)^{\Delta_{31}}} \tag{256}$$
$$\Delta_{12} = \Delta_1 + \Delta_2 - \Delta_3, \text{ and cyclic permutations}$$

It follows immediately from the exercise above that they do transform correctly.

Operators such as conserved currents $\mathcal{J}_i$ and the conserved traceless stress tensor T_{ij} are important in a CFT_d. Under inversion $\mathcal{J}_i(x) \to J_{ij}(x')x'^{2(d-1)}\mathcal{J}_j(x')$ with an analogous rule for T_{ij}. The 2-point function of a conserved current takes the form

$$\begin{aligned}\langle \mathcal{J}_i(x)\mathcal{J}_j(y)\rangle &\approx (\partial_i\partial_j - \Box\delta_{ij})\frac{1}{(x-y)^{(2d-4)}} \\ &\sim \frac{J_{ij}(x-y)}{(x-y)^{(2d-2)}}\end{aligned} \tag{257}$$

The exercise above can be used to show this tensor does transform correctly. Here are some new exercises.

Ex. 11:. show that the second line in (257) follows from the manifestly conserved first form and obtain the missing coefficient.

Ex. 12:. Use the projection operator $\pi ij = \partial_i\partial_j - \Box\delta_{ij}$ to write the 2-point correlator of the stress tensor and then convert to a form with manifestly correct inversion properties,

$$\begin{aligned}\langle T_{ij}(x)T_{kl}(y)\rangle &= [2\pi_{ij}\pi_{kl} - 3(\pi_{ik}\pi_{jl} + \pi_{il}\pi_{jk})]\frac{c}{(x-y)^{(2d-4)}} \\ &\sim \frac{J_{ik}(x-y)J_{jl}(x-y) + k \leftrightarrow l - \frac{2}{d}\delta_{ij}\delta_{kl}}{(x-y)^{2d}}\end{aligned} \tag{258}$$

This form is unique. For $d \geq 4$ there are two independent tensor structures for a 3-point function of conserved currents and three structures for the 3-point function of T_{ij}. For more information on the tensor structure of conformal amplitudes, see the work of Osborn and collaborators, for example.[168,169]

It is useful to mention that any finite special conformal transformation can be expressed as a product of (inversion)(translation)(inversion).

Ex. 13:. Show that the finite transformation is $x_i \to (x_i + x^2 a_i)/(1 + 2a \cdot x + a^2x^2)$. Show that the flat Euclidean line element transforms with a conformal factor under this transformation. Show that the commutator of an infinitesimal special conformal transformation and a translation involves a rotation plus scale transformation.

The behavior of amplitudes under rotations and translations is rather trivial to test. Special conformal symmetry is more difficult, but it can be reduced to inversion. Thus the behavior under inversion essentially establishes covariance under the full conformal group.

We will soon put the inversion to good use in our study of the AdS/CFT correspondence, but we first need to discuss how the dynamics of the correspondence works.

8.3. *AdS/CFT Amplitudes in a Toy Model*

Let us consider a toy model of a scalar field $\phi(z)$ in an AdS_{d+1} Euclidean signature background. The action is

$$S = \frac{1}{8\pi G}\int d^{d+1}z\sqrt{g}\left(\frac{1}{2}\partial_\mu\phi\partial^\mu\phi + \frac{1}{2}m^2\phi^2 + \frac{1}{3}b\phi^3 + \cdots\right) \tag{259}$$

We will outline the general prescription for correlation functions due to Witten[3] and then give further details. The first step is to solve the non-linear classical field equations

$$\frac{\delta S}{\delta\phi} = (-\Box + m^2)\phi + b\phi^2 + \cdots = 0 \tag{260}$$

with the boundary condition

$$\begin{aligned} \phi(z_0,\vec{z}) &\underset{z_0\to 0}{\longrightarrow} z_0^{d-\Delta}\bar{\phi}(\vec{z}) \\ \Delta &= \tfrac{d}{2} + \tfrac{1}{2}\sqrt{d^2+4m^2} \end{aligned} \tag{261}$$

This is a modified Dirichlet boundary value problem with boundary data $\bar{\phi}(\vec{z})$. The scaling rate $z_0^{d-\Delta}$ is that of the leading Frobenius solution of the linearized version[xiv] of (260).

Exact solutions of the non-linear equation (260) with general boundary data are beyond present ability, so we work with the iterative solution

$$\phi_0(z) = \int d^d\vec{x}K_\Delta(z_0,\vec{z}-\vec{x})\bar{\phi}(\vec{x}) \tag{262}$$

$$\phi(z) = \phi_0(z) + b\int d^{d+1}w\sqrt{g}G(z,w)\phi_0^2(w) + \cdots \tag{263}$$

[xiv]To simplify the discussion we restrict throughout to the range $m^2 > -\frac{d^2}{4}$ and consider $\Delta > \frac{1}{2}d$. See Ref. 170 for an extension to the region $\frac{1}{2}d \geq \Delta \geq \frac{1}{2}(d-2)$ close to the unitarity bound. See also paper by P. Minces and V. O. Rivelles, JHEP **0112**, 010 (2001) [hep-th/0110189].

The linear solution ϕ_0 involves the bulk-to-boundary propagator

$$K_\Delta(z_0,\vec{z}) = C_\Delta\left(\frac{z_0}{z_0^2+\vec{z}^2}\right)^\Delta \qquad C_\Delta = \frac{\Gamma(\Delta)}{\pi^{\frac{d}{2}}\Gamma(\Delta-\frac{d}{2})}, \quad (264)$$

which satisfies $(\Box + m^2)K_\Delta(z_0,\vec{z}) = 0$. Interaction terms require the bulk-to-bulk propagator $G(z,w)$ which satisfies $(-\Box_z + m^2)G(z,w) = \delta(z,w)/\sqrt{g}$ and is given by the hypergeometric function

$$\begin{aligned} G_\Delta(u) &= \tilde{C}_\Delta(2u^{-1})^\Delta F\left(\Delta,\Delta-d+\frac{1}{2};2\Delta-d+1;-2u^{-1}\right) \quad (265)\\ \tilde{C}_\Delta &= \frac{\Gamma(\Delta)\Gamma(\Delta-\frac{d}{2}+\frac{1}{2})}{(4\pi)^{(d+1)/2}\Gamma(2\Delta-d+1)}\\ u &= \frac{(z-w)^2}{2z_0w_0}. \end{aligned}$$

This differs from the form given in Sec. 6.3 by a quadratic hypergeometric transformation, see Ref. 133.

For several purposes in dealing with the AdS/CFT correspondence it is appropriate to insert a cutoff at $z_0 = \epsilon$ in the bulk geometry and consider a true Dirichlet problem at this boundary. This is the situation of 19th century boundary value problems where Green's formula gives a well known relation between G and K. Essentially K is the normal derivative at the boundary of G. The cutoff region has less symmetry than full AdS. Exact expressions for G and K in terms of Bessel functions in the $\vec{p}$-space conjugate to $\vec{z}$ are straightforward to obtain, but the Fourier transform back to $z_0, \vec{z}$ is unknown. See Sec. 8.5 below.

The next step is to substitute the solution $\phi(z)$ into the action (259) to obtain the on-shell action $S[\bar{\phi}]$ which is a functional of the boundary data. The key dynamical statement of the AdS/CFT correspondence is that $S[\bar{\phi}]$ is the generating functional for correlation functions of the dual operator $\mathcal{O}(\vec{x})$ in the boundary field theory, so that

$$\langle\mathcal{O}(\vec{x_1})\cdots\mathcal{O}(\vec{x_n})\rangle = (-)^{n-1}\cdot\frac{\delta}{\delta\bar{\phi}(\vec{x_1})}\cdots\frac{\delta}{\delta\bar{\phi}(\vec{x_n})}S[\bar{\phi}]\Big|_{\bar{\phi}=0} \quad (266)$$

Another way to state things is that the boundary data for bulk fields play the role of sources for dual field theory operators. The integrals in the on-shell action diverge at the boundary and must be cut off either as discussed above or by a related method.[171,172] However we will proceed formally here.

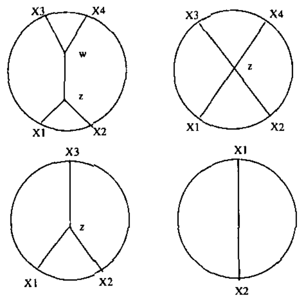


Figure 11. Some Witten Diagrams

From the expansion of $S[\bar{\phi}]$ in powers of $\bar{\phi}$, one obtains a diagrammatic algorithm (in terms of Witten diagrams) for the correlation functions. Some examples are given in Figure 11. In these diagrams the interior and boundary of each disc denote the interior and boundary of the AdS geometry. The rules for interpretation and computation associated with the diagrams are as follows:

(a) boundary points $\vec{x}_i$ are points of flat Euclidean$_d$ space where field theory operators are inserted.
(b) bulk points $z, w \,\epsilon\, AdS_{d+1}$ and are integrated as $\int d^{d+1}z\sqrt{\bar{g}(z)}$
(c) Each bulk-to-boundary line carries a factor of K_Δ and each bulk-to-bulk line a factor of $G(z,w)$
(d) An n-point vertex carries a coupling factor from the interaction terms of the bulk Lagrangian, e.g. $\mathcal{L} = \frac{1}{3}b\phi^3 + \frac{1}{4}c\phi^4 + \cdots$ with the same combinatoric weights as for Feynman-Wick diagrams. This is most clearly derived using the cutoff discussed above.

Let us examine this construction more closely beginning with the linear solution for bulk fields.

Ex. 14:. Show that the linearized field equation can be written as

$$(z_0^2\partial_0^2 - (d-1)z_0\partial_0 + z_0^2\nabla^2 - m^2)\phi = 0 \tag{267}$$

and that $K(z_0, \vec{z})$ given above is a solution. Plot $K(z_0, \vec{z})$ as a function of $|\vec{z}|$ for several fixed values of z_0. Note that it becomes more and more like $\delta(\vec{z})$ as $z_0 \to 0$.

The exercise shows that $\phi_0(z)$ in (262) is indeed a solution of (267) and suggests that it satisfies the right boundary condition. Let's verify that it has the correct normalization at the boundary. Because of translation symmetry there is no loss of generality in taking $\vec{z} = 0$. We then have

$$
\begin{aligned}
\phi(z_0, 0) &= C_\Delta \int d^d\vec{x} (\tfrac{z_0}{z_0^2+\vec{x}^2})^\Delta \bar{\phi}(\vec{x}) \\
&= C_\Delta z_0^{d-\Delta} \int d^d\vec{y} (\tfrac{1}{1+\vec{y}^2})^\Delta \bar{\phi}(z_0\vec{y}) \\
&\underset{z_0 \to 0}{\longrightarrow} C_\Delta z_0^{d-\Delta} I_\Delta \bar{\phi}(0) \\
I_\Delta &= \int \frac{d^d\vec{y}}{(1+\vec{y}^2)^\Delta}
\end{aligned}
\tag{268}
$$

Thus we do satisfy the boundary condition (261) provided that $C_\Delta = \frac{1}{I_\Delta}$ and the integral does indeed give the value of C_Δ in (264).

8.4. *How to Calculate 3-Point Correlation Functions*

Two-point correlations do not contain a bulk integral and turn out to require a careful cutoff procedure which we discuss later. For these reasons 3-point functions are the prototype case, and we now discuss them in some detail. The basic integral to be done is:

$$
\begin{aligned}
A(\vec{x}, \vec{y}, \vec{z}) &= \int \frac{dw_0 d^d\vec{w}}{w_0^{d+1}} \left(\frac{w_0}{(w-\vec{x})^2}\right)^{\Delta_1} \left(\frac{w_0}{(w-\vec{y})^2}\right)^{\Delta_2} \left(\frac{w_0}{(w-\vec{z})^2}\right)^{\Delta_3} \\
(w - \vec{x})^2 &\equiv w_0^2 + (\vec{w} - \vec{x})^2
\end{aligned}
\tag{269}
$$

Let us first illustrate the use of the method of inversion. We change integration variable by $w_\mu = w'_\mu / w'^2$ and at the same time refer boundary points to their inverses, *i.e.* $\vec{x} = \vec{x}\,'/(\vec{x}\,')^2$ and the same for $\vec{y}, \vec{z}$. The bulk-to-boundary propagator transform very simply

$$
K_\Delta(w, \vec{x}) = |\vec{x}\prime|^{2\Delta} K_\Delta(w', \vec{x}\,') \tag{270}
$$

with the prefactor associated with a field theory operator $\mathcal{O}_\Delta(\vec{x})$ clearly in evidence. The AdS volume element is invariant, *i.e.* $d^{d+1}w/w_0^{d+1} = d^{d+1}w'/w_0'^{d+1}$ since inversion is an isometry.

Ex. 15:. Use results of previous exercises to prove these important facts.

We then find that

$$
A(\vec{x}, \vec{y}, \vec{z}) = |\vec{x}\,'|^{2\Delta_1} |\vec{y}\,'|^{2\Delta_2} |\vec{z}\,'|^{2\Delta_3} A(\vec{x}\,', \vec{y}\,', \vec{z}\,') \tag{271}
$$

Thus the AdS/CFT procedure produces a 3-point function which transforms correctly under inversion. See (254).

This is a very general property which holds for **all** AdS/CFT correlators. Suppose you wish to calculate $\langle \mathcal{J}_i^a \mathcal{J}_j^b \mathcal{J}_k^c \rangle$. The Witten amplitude is the product (see Ref. 90) of 3 vector bulk-to-boundary propagators, each given by

$$G_{\mu i}(w, \vec{x}) = \frac{1}{2} c_d \frac{w_0^{d-1}}{(w - \vec{x})^{d-1}} J_{\mu i}(w - \vec{x}), \tag{272}$$

in which the Jacobian (251) appears. The bulk indices are contracted with a vertex rule from the Yang-Mills interaction $f^{abc} A^a_\mu A^b_\nu \partial_\mu A^c_\nu$. If you try to do the change of variable in detail, you get a mess. But the process is guaranteed to produce the correct inversion factors for the conserved currents, namely $|\vec{x}\,'|^{2(d-1)} J_{ii'}(\vec{x}\,')$, etc, because inversion is an isometry of AdS$_{d+1}$and all pieces of the amplitude conspire to preserve this symmetry.

Ex. 16:. Show that $G_{\mu i}(w, \vec{x})$ satisfies the bulk Maxwell equation

$$\partial_\mu \sqrt{\bar{g}} \bar{g}^{\mu\nu} \left(\partial_\nu G_{\rho i}(w, \vec{x}) - \partial_\rho G_{\nu i}(w, \vec{x}) \right) = 0 \tag{273}$$

where $\partial_\mu = \partial / \partial w_\mu$. Express $G_{\mu i}(w, \vec{x})$ in terms of the inverted $G_{\mu' i'}(w', \vec{x}\,')$.

We can conclude that all AdS/CFT amplitudes are conformal covariant! A transformation of the $SO(d+1,1)$ isometry group of the bulk is dual to an $SO(d+1,1)$ conformal transformation on the boundary. Since there is a unique covariant form for scalar 3-point functions, given in (254), the AdS/CFT integral $A(\vec{x}, \vec{y}, \vec{z})$ is necessarily a constant multiple of this form. Our exercise also shows conclusively that a scalar field of AdS mass m^2 is dual to an operator $\mathcal{O}_\Delta(\vec{x})$ of dimension Δ given by (261).

We still need to **do** the bulk integral to obtain the constant $\tilde{c}$. It is hard to do the integral in the original form (269) because it contains 3 denominators and the restriction $w_0 > 0$. But we can simplify it by using inversion in a somewhat different way. We use translation symmetry to move the point $\vec{z} \longrightarrow 0$, *i.e.* $A(\vec{x}, \vec{y}, \vec{z}) = A(\vec{x} - \vec{z}, \vec{y} - \vec{z}, 0) \equiv A(\vec{u}, \vec{v}, 0)$. The integral for $A(\vec{u}, \vec{v}, 0)$ is similar to (269) except that the third propagator is simplified,

$$\left(\tfrac{w_0}{(w - \vec{z})^2} \right)^{\Delta_3} \longrightarrow \left(\tfrac{w_0}{w^2} \right)^{\Delta_3} = (w_0')^{\Delta_3}. \tag{274}$$

There is no denominator in the inverted frame since $\vec{z} = 0 \longrightarrow \vec{z}\,' = \infty$. After inversion the integral is

$$A(\vec{u}, \vec{v}, 0) = \tfrac{1}{|\vec{u}|^{2\Delta_1} |\vec{v}|^{2\Delta_2}} \int \tfrac{d^{d+1} w'}{(w_0')^{d+1}} \left[\tfrac{w_0'}{(w' - \vec{u}\,')^2} \right]^{\Delta_1} \left[\tfrac{w_0'}{(w' - \vec{v}\,')^2} \right]^{\Delta_2} (w_0')^{\Delta_3} \tag{275}$$

The integral can now be done by conventional Feynman parameter methods, which give

$$A(\vec{u},\vec{v},0) = \frac{1}{|\vec{u}|^{2\Delta_1}|\vec{v}|^{2\Delta_2}}\frac{a}{|\vec{u}\,'-\vec{v}\,'|^{\Delta_1+\Delta_2-\Delta_3}} \tag{276}$$

where a stands for

$$\frac{\pi^{d/2}}{2}\frac{\Gamma(\frac{1}{2}(\Delta_1+\Delta_2-\Delta_3))\Gamma(\frac{1}{2}(\Delta_2+\Delta_3-\Delta_1))\Gamma(\frac{1}{2}(\Delta_3+\Delta_1-\Delta_2))\Gamma(\frac{1}{2}(\Delta_1+\Delta_2+\Delta_3-d))}{\Gamma(\Delta_1)\Gamma(\Delta_2)\Gamma(\Delta_3)}$$

Ex. 17:. Repristinate the original variables $\vec{x},\vec{y},\vec{z}$ to obtain the form (256) with $\tilde{c} = a$.

The major application of this result was already discussed in Sec. 6.7. A Princeton group[91] obtained the cubic couplings b_{klm} of the Type IIB supergravity modes on $AdS_5 \times S^5$ which are dual to the chiral primary operators $\mathrm{tr}X^k$, etc. of $\mathcal{N}=4$ SYM theory. They combined these couplings with the Witten integral above and observed that the AdS/CFT prediction

$$\langle \mathrm{tr}X^k(\vec{x})\mathrm{tr}X^l(\vec{y})\mathrm{tr}X^m(\vec{z})\rangle = b_{klm}c_kc_lc_mA(\vec{x},\vec{y},\vec{z}) \tag{277}$$

for the large N, large λ supergravity limit agreed with the **free field** Feynman amplitude for these correlators. They conjectured a broader non-renormalization property. It was subsequently confirmed in weak coupling studies in the field theory that order g^2, g^4 and non-perturbative instanton contributions to these correlations vanished for all N and all gauge groups. General all orders arguments for non-renormalization have also been developed. The non-renormalization of 3-point functions of chiral primaries (and their descendents) was a surprise and the first major new result about $\mathcal{N} = 4$ SYM obtained from AdS/CFT. (See the references cited in Sec 6.7.)

8.5. 2-*Point Functions*

This is an important case, but more delicate, since a cutoff procedure is required to obtain a concrete result from the formal integral expression. Since 3-point functions do not require a cutoff, one way to bypass this problem is to study the 3-point function $\langle \mathcal{J}_i(z)\mathcal{O}_\Delta(x)\mathcal{O}^*_\Delta(y)\rangle$ of a conserved current and a scalar operator $\mathcal{O}_\Delta(x)$ assumed to carry one unit of $U(1)$ charge.[xv] The Ward identity relates $\langle \mathcal{J}_i\mathcal{O}_\Delta\mathcal{O}^*_\Delta\rangle$ to $\langle \mathcal{O}_\Delta\mathcal{O}^*_\Delta\rangle$. There is a

[xv] When no ambiguity arises we will denote boundary points by x, y, z etc. rather than $\vec{x},\vec{y},\vec{z}$.

unique conformal tensor for $\langle \mathcal{J}_i \mathcal{O}_\Delta \mathcal{O}^*_\Delta \rangle$ in any CFT$_d$, namely

$$\langle J_i(z) \mathcal{O}_\Delta(x) \mathcal{O}^*_\Delta(y) \rangle \tag{278}$$
$$= -i\xi \frac{1}{(x-y)^{2\Delta-d+2}} \frac{1}{(x-z)^{d-2}(y-z)^{d-2}} \left[\frac{(x-z)i}{(x-z)^2} - \frac{(y-z)i}{(y-z)^2} \right]$$

and the Ward identity is

$$\begin{aligned} \frac{\partial}{\partial z_i} \langle J_i(z) \mathcal{O}_\Delta(x) \mathcal{O}^*_\Delta(y) \rangle &= i[\delta(x-z) - \delta(y-z)] \langle \mathcal{O}_\Delta(x) \mathcal{O}^*_\Delta(y) \rangle \\ &= i[\delta(x-z) - \delta(y-z)] \frac{2\pi^{d/2}}{\Gamma(d/2)} \xi \frac{1}{(x-y)^{2\Delta}} \end{aligned} \tag{279}$$

Ex. 18:. Derive (279) from (278).

To implement the gravity calculation of $\langle \mathcal{J}_i \mathcal{O}_\Delta \mathcal{O}^*_\Delta \rangle$ we extend the bulk toy model (259) to include a $U(1)$ gauge coupling

$$L = \frac{1}{4} F_{\mu\nu} F^{\mu\nu} + \bar{g}^{\mu\nu} (\partial_\mu + iA_\mu) \phi^* (\partial_\nu - iA_\nu) \phi \tag{280}$$

In application to the duality between Type IIB sugra and $\mathcal{N} = 4$ SYM, the $U(1)$ would be interpreted as a subgroup of the $SO(6)$ R-symmetry group. The cubic vertex leads to the AdS integral

$$\langle \mathcal{J}_i(z) \mathcal{O}_\Delta(x) \mathcal{O}^*_\Delta(y) \rangle = -i \int \frac{d^{d+1}w}{w_0^{d+1}} G_{\mu i}(w, \vec{z}) w_0^2 K_\Delta(w, \vec{x}) \overleftrightarrow{\frac{\partial}{\partial w_\mu}} K_\Delta(w, \vec{y}) \tag{281}$$

Ex. 19:. The integral can be done by the inversion technique, please do it.

The result is the tensor form (278) with coefficient

$$\xi = \frac{(\Delta - d/2)\Gamma(\frac{d}{2})\Gamma(\Delta)}{\pi^{d/2}\Gamma(\Delta - d/2)} \tag{282}$$

Using (279) we thus obtain the 2-point function

$$\langle \mathcal{O}_\Delta(x) \mathcal{O}^*_\Delta(y) \rangle = \frac{(2\Delta - d)\Gamma(\Delta)}{\pi^{d/2}\Gamma(\Delta - d/2)} \frac{1}{(x-y)^{2\Delta}} \tag{283}$$

We now discuss a more direct computation[2,90] of 2-point correlators from a Dirichlet boundary value problem in the AdS bulk geometry with cutoff at $z_0 = \epsilon$. This method illustrates the use of a systematic cutoff, and it may be applied to (some) 2-point functions in holographic RG flows for which the 3-point function $\langle \mathcal{J}_i \mathcal{O}_\Delta \mathcal{O}^*_\Delta \rangle$ cannot readily be calculated.

The goal is to obtain a solution of the linear problem

$$\begin{aligned} (\Box - m^2) \phi(z_0, \vec{z}) &= 0 \\ \phi(\epsilon, \vec{z}) &= \bar{\phi}(\vec{z}) \end{aligned} \tag{284}$$

The result will be substituted in the bilinear part of the toy model action to obtain the on-shell action. After partial integration we obtain the boundary integral

$$S[\bar{\phi}] = \frac{1}{2\epsilon^{d-1}} \int d^d\vec{z}\bar{\phi}(\vec{z})\partial_0\phi(\epsilon,\vec{z}) \tag{285}$$

Since the cutoff region $z_0 \geq \epsilon$ does not have the full symmetry of AdS, an exact solution of the Dirichlet problem is impossible in x-space, so we work in p-space. Using the Fourier transform

$$\phi(z_0,\vec{z}) = \int d^d\vec{p}e^{i\vec{p}\cdot\vec{z}}\phi(z_0,\vec{p}) \tag{286}$$

we find the boundary value problem

$$\begin{gathered}[z_0^2\partial_0^2 - (d-1)z_0\partial_0 - (p^2z_0^2+m^2)]\phi(z_0,\vec{p}) = 0 \\ \phi(\epsilon,\vec{p}) = \bar{\phi}(\vec{p})\end{gathered} \tag{287}$$

where $\bar{\phi}(\vec{p})$ is the transform of the boundary data. The differential equation is essentially Bessel's equation, and we choose the solution involving the function $z_0^{d/2}K_\nu(pz_0)$, where $\nu = \Delta - d/2, p = |\vec{p}|$, which is exponentially damped as $z_0 \to \infty$ and behaves as $z_0^{d-\Delta}$ as $z_0 \to 0$. The second solution $z_0^{d/2}I_\nu(pz_0)$ is rejected because it increases exponentially in the deep interior. The normalized solution of the boundary value problem is then

$$\phi(z_0,\vec{p}) = \frac{z_0^{d/2}K_\nu(pz_0)}{\epsilon^{d/2}K_\nu(p\epsilon)}\bar{\phi}(\vec{p}), \tag{288}$$

The on-shell action in p-space is

$$S[\bar{\phi}] = \frac{1}{2\epsilon^{d-1}} \int d^dpd^dq(2\pi)^d\delta(\vec{p}+\vec{q})\phi(\epsilon,\vec{p})\partial_0\phi(\epsilon,q) \tag{289}$$

which leads to the cutoff correlation function

$$\begin{aligned}\langle O_\Delta(\vec{p})O_\Delta(\vec{q})\rangle_\epsilon &= -\frac{\delta^2 S}{\delta\bar{\phi}(\vec{p})\delta\bar{\phi}(p)} \\ &= -\frac{(2\pi)^d\delta(\vec{p}+\vec{q})}{\epsilon^{d-1}}\frac{d}{d\epsilon}\ln(\epsilon^{d/2}K_\nu(p\epsilon))\end{aligned} \tag{290}$$

To extract a physical result, we need the boundary asymptotics of the Bessel function $K_\nu(p\epsilon)$. The values of $\nu = \Delta - d/2$ which occur in most applications of AdS/CFT are integer. The asymptotics were worked out for continuous ν in the Appendix of Ref. 90 with an analytic continuation to the final answer. Here we assume integer ν, although an analytic continuation will be necessary to define Fourier transform to x-space. The behavior of $K_\nu(u)$ near $u = 0$ can be obtained from a standard compendium on special functions such as.[173] For integer ν, the result can be written schematically as

$$K_\nu(u) = u^{-\nu}(a_0 + a_1 u^2 + a_2 u^4 + \cdots) + u^\nu ln(u)\,(b_0 + b_1 u^2 + b_2 u^4 + \cdots) \tag{291}$$

where the a_i, b_i are functions of ν given in Ref. 173. This expansion may be used to compute the right side of (290) leading to

$$\begin{aligned}\langle \mathcal{O}_\Delta(\vec{p})\mathcal{O}_\Delta(\vec{q})\rangle_\epsilon = \tfrac{(2\pi)^d \delta(\vec{p}+\vec{q})}{\epsilon^d}. \, [-\tfrac{d}{2} + \nu(1 + c_2\epsilon^2 p^2 + c_4 \epsilon^4 p^4 + \cdots) \\ - \tfrac{2\nu b_0}{a_0}\epsilon^{2\nu} p^{2\nu} ln(p\epsilon)(1 + d_2 \epsilon^2 p^2 + \cdots)]\end{aligned} \tag{292}$$

where the new constants c_i, d_i are simply related to a_i, b_i. From Ref. 173 we obtain the ratio

$$\frac{2\nu b_0}{a_0} = \frac{(-)^{(\nu-1)}}{2^{(2\nu-2)}\Gamma(\nu)^2} \tag{293}$$

which is the only information explicitly needed.

This formula is quite important for applications of AdS/CFT ideas to both conformal field theories and RG flows where similar formulas appear. The physics is obtained in the limit as $\epsilon \to 0$, and we scale out the factor $\epsilon^{2(\Delta-d)}$ which corresponds to the change from the true Dirichlet boundary condition to the modified form (261) for the full AdS space. We also drop the conventional momentum conservation factor $(2\pi)^d\delta(\vec{p} + \vec{q})$ and study

$$\begin{aligned}\langle \mathcal{O}_\Delta(p)\mathcal{O}_\Delta(-p)\rangle = \tfrac{\beta_0 + \beta_1\epsilon^2 p^2 + \cdots + \beta_\nu(\epsilon p)^{2(\nu-1)}}{\epsilon^{2\Delta-d}} \\ - \tfrac{2\nu b_0}{a_0} p^{2\nu}\ln(p\epsilon) + \mathcal{O}(\epsilon^2)\end{aligned} \tag{294}$$

The first part of this expression is a sum of non-negative integer powers p^{2m} with singular coefficients in ϵ. The Fourier transform of p^{2m} is $\Box^m\delta(\vec{x} - \vec{y})$, a pure contact term in the $\vec{x}$-space correlation. Such terms are usually physically uninteresting and scheme dependent in quantum field theory. Indeed it is easy to see that the singular powers $\epsilon^{2(m-\Delta)+d}$ carried by the terms corresponds to their dependence on the ultraviolet cutoff $\Lambda^{2(\Delta-m)-d}$ in a field theory calculation. This gives rise to the important observation that the $\epsilon-$cutoff in AdS space which cuts off long distance effects in the bulk corresponds to an ultraviolet cutoff in field theory. Henceforth we drop the polynomial contact terms in (294).

The physical p-space correlator is then given by

$$\langle \mathcal{O}_\Delta(p)\mathcal{O}_\Delta(-p)\rangle = -\frac{2\nu b_0}{a_0} p^{2\nu}\ln p. \tag{295}$$

This has an absorptive part which is determined by unitarity in field theory. Its Fourier transform is proportional to $1/(x-y)^{2\Delta}$ which is the correct CFT behavior for $\langle \mathcal{O}_\Delta \mathcal{O}_\Delta \rangle$. The precise constant can be obtained using differential regularization[174] or by analytic continuation in ν from the region where the Fourier transform is defined.

The result agrees exactly with the 2-point function calculated from the Ward identity in (283).

8.6. *Key AdS/CFT Results for $\mathcal{N}=4$ SYM and CFT$_d$ Correlators*

We can now summarize the important results discussed in this chapter and earlier ones for CFT$_d$ correlation functions from the AdS/CFT correspondence.

(i) the non-renormalization of $\langle \mathrm{tr}X^k \mathrm{tr}X^l \mathrm{tr}X^m \rangle$ in $\mathcal{N}=4$ SYM theory

(ii) 4-point functions are less constrained than 2- and 3- point functions in any CFT. In general they contain arbitrary functions $F(\xi,\eta)$ of two invariant variables, the cross ratios

$$\xi = \frac{x_{13}^2 x_{24}^2}{x_{12}^2 x_{34}^2} \qquad \eta = \frac{x_{14}^2 x_{24}^2}{x_{12}^2 x_{34}^2} \qquad x_{ij} = x_i - x_j \tag{296}$$

One way to extract the physics of 4-point functions is to use the operator product expansion. This is written

$$\mathcal{O}_\Delta(x)\mathcal{O}_{\Delta'}(y) \underset{x\to y}{\longrightarrow} \sum_p \frac{a_{\Delta\Delta'\Delta_p}}{(x-y)^{\Delta+\Delta'-\Delta_p}} \mathcal{O}_{\Delta_p}(y) \tag{297}$$

which is interpreted to mean that at short distance inside any correlation function, the product of two operators acts as a sum of other local operators with power coefficients. For simplicity we have indicated only the contributions of primary operators. Thus, in the limit where $|x_{12}|, |x_{34}| \ll |x_{13}|$, a 4-point function must factor as

$$\begin{aligned} &\langle \mathcal{O}_{\Delta_1}(x_1)\mathcal{O}_{\Delta_2}(x_2)\mathcal{O}_{\Delta_3}(x_3)\mathcal{O}_{\Delta_4}(x_4)\rangle \\ &\qquad \approx \sum_p \frac{a_{12p}}{(x_{12})^{\Delta_1+\Delta_2-\Delta_p}} \frac{c_p}{(x_{13})^{2\Delta_p}} \frac{a_{34p}}{(x_{34})^{\Delta_3+\Delta_d-\Delta_p}} \end{aligned} \tag{298}$$

One must expect that AdS/CFT amplitudes satisfy this property and indeed they do in a remarkably simple way. The amplitude of a Witten diagram for exchange of the bulk field $\phi_p(z)$ dual to $\mathcal{O}_{\Delta_p}(\vec{z})$ factors with the correct coefficients c_p, a_{12p}, a_{34p} determined from 2- and 3-point functions. This holds for singular powers, *e.g.* $\Delta_1 + \Delta_2 - \Delta_p > 0$.

The AdS/CFT amplitude also contains a $\ln(\xi)$ term in its short distance asymptotics. This is the level of the *OPE* at which $\mathcal{O}_p =: \mathcal{O}_{\Delta_1}(\vec{x})\mathcal{O}_{\Delta_2}(\vec{y}) :$ contributes. In $\mathcal{N}=4$ SYM theory the normal product is a double trace operator, *e.g.* $: \mathrm{tr}X^k(y)\mathrm{tr}X^l(y) :$, which has components in irreps of $SO(6)$ contained in the direct product $(0,k,0)\otimes(0,l,0)$. The irreducible components are generically primaries of long representations of $SU(2,2|4)$. Their scale dimensions are not fixed, and have a large N expansion of the form $\Delta_{kl} = k + l + \gamma_{kl}/N^2 + \cdots$. The contribution $\Delta\gamma_{kl}$ can be read from the $\ln(\xi)$ term of the 4-point function. It is a strong coupling prediction of AdS/CFT, which cannot yet be checked by field theoretic methods.
(iii) Another surprising fact about $\mathcal{N}=4$ SYM correlators suggested by the AdS/CFT correspondence is that **extremal** n-point functions are not renormalized. The extremal condition for 4-point functions is $\Delta_1 = \Delta_2 + \Delta_3 + \Delta_4$. The name extremal comes from the fact that the correlator vanishes by $SO(6)$ symmetry for any larger value of Δ_1. As discussed in detail in Secs. 6.8 and 6.9, the absence of radiative corrections was suggested by the form of the supergravity couplings and Witten integrals. This prediction was confirmed by weak coupling calculation and general arguments in field theory. Field theory then suggested that next-to-external correlators ($\Delta_1 = \Delta_2 + \Delta_3 + \Delta_4 - 2$) were also not renormalized, and this was subsequently verified by AdS/CFT methods.

It is clear that the AdS/CFT correspondence is a new principle which stimulated an interplay of work involving both supergravity and field theory methods. As a result we have much new information about the $N=4$ SYM theory. It confirms that AdS/CFT has quantitative predictive power, so we can go ahead and apply it in other settings.

9. Holographic Renormalization Group Flows

We have already seen that AdS/CFT has taught us a great deal of useful information about $\mathcal{N}=4$ SYM theory as a CFT$_4$. But years of elegant work in CFT$_2$ has taught us to consider both the pure conformal theory and its deformation by relevant operators. The deformed theory exhibits RG flows in the space of coupling constants of the relevant deformations. For general dimension d we can also consider the CFT$_d$ perturbed by relevant operators. For $\mathcal{N}=4$ SYM theory, the perturbed Lagrangian would take the form

$$\mathcal{L} = \mathcal{L}_{\mathcal{N}=4} + \frac{1}{2}m^2_{ij}\mathrm{tr}X^iX^j + \frac{1}{2}M_{ab}\mathrm{tr}\psi^a\psi^b + b_{ijk}\mathrm{tr}X^iX^jX^k. \tag{299}$$

For $d > 2$ there is the additional option of Coulomb and Higgs phases in which gauge symmetry is spontaneously broken. The Lagrangian is not changed, but certain operators acquire vacuum expectation values, e.g. $\langle X^i \rangle \neq 0$ in $\mathcal{N} = 4$ SYM. In all these cases conformal symmetry is broken because a scale is introduced. The resulting theories have the symmetry of the Poincaré group in d dimensions which is smaller than the conformal group $SO(1, d+1)$. Our purpose in this chapter is to explore the description of such theories using $D = d+1$ dimensional gravity. We will focus on relevant operator deformations.

9.1. *Basics of RG Flows in a Toy Model*

The basic ideas for the holographic description of field theories with RG flow were presented in Refs. 175, 176. We will discuss these ideas in a simple model in which Euclidean $(d+1)$-dimensional gravity interacts with a single bulk scalar field with action

$$S = \frac{1}{4\pi G} \int d^{d+1}\sqrt{g}\left[-\frac{1}{4}R + \frac{1}{2}\partial_\mu\phi\partial^\mu\phi + V(\phi)\right] \tag{300}$$

We henceforth choose units in which $4\pi G = 1$. In these units ϕ is dimensionless and all terms in the Lagrangian have dimension 2. We envisage a potential $V(\phi)$ which has one or more critical points, *i.e.* $V'(\phi_i) = 0$, at which $V(\phi_i) < 0$. We consider both maxima and minima. See Figure 12.

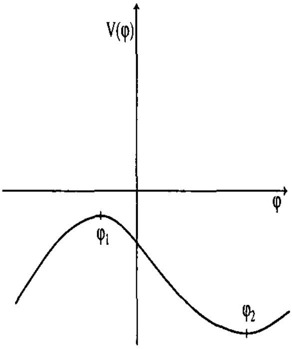


Figure 12. Potential $V(\phi)$

The Euler-Lagrange equations of motion of our system are

$$\frac{1}{\sqrt{g}}\partial_\mu(\sqrt{g}g^{\mu\nu}\partial_\nu\phi) - V'(\phi) = 0 \tag{301}$$

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = 2\left[\partial_\mu\phi\partial_\nu\phi - g_{\mu\nu}\left(\frac{1}{2}(\partial\phi)^2 + V(\phi)\right)\right] = 2T_{\mu\nu} \tag{302}$$

For each critical point ϕ_i there is a trivial solution of the scalar equation, namely $\phi(z) \equiv \phi_i$. The Einstein equation then reduces to

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = -2g_{\mu\nu}V(\phi_i). \tag{303}$$

This is equivalent to the Einstein equation of the action (249) if we identify $\Lambda_i = 4V(\phi_i) = -d(d-1)/L_i^2$. Thus constant scalar fields with AdS$_{d+1}$geometries of scale L_i are solutions of our model. We refer to them as critical solutions.

However, more general solutions in which the scalar field is not constant are needed to describe the gravity duals of RG flows in field theory. Since the symmetries must match on both sides of the duality, we look for solutions of the $D = d + 1$-dimensional bulk equations with d-dimensional Poincaré symmetry. The most general such configuration is

$$\begin{aligned} ds^2 &= e^{2A(r)}\delta_{ij}dx^i dx^j + dr^2 \\ \phi &= \phi(r) \end{aligned} \tag{304}$$

This is known as the domain wall ansatz. The coordinates separate into a radial coordinate r plus d transverse coordinates x^i with manifest Poincaré symmetry. Several equivalent forms which differ only by change of radial coordinate also appear in the literature.

Domain wall metrics have several modern applications, and it is worth outlining a method to compute the connection and curvature. Symbolic manipulation programs are very useful for this purpose, but analytic methods can also be useful, and we discuss a method which uses the Cartan structure equations. A similar method works quite well for brane metrics such as (241). One proceeds as follows using the notation of differential forms:

(1) The first step is to choose a basis of frame 1-forms $e^a = e^a_\mu dx^\mu$ such that the metric is given by the inner product $ds^2 = e^a\delta_{ab}e^b$.

(2) The torsion-free connection 1-form is then defined by $de^a + \omega^{ab}\wedge e^b = 0$ with the condition $\omega^{ab} = -\omega^{ba}$. The connection is valued in the Lie algebra of $SO(d+1)$.

(3) The curvature 2-form is

$$R^{ab} = d\omega^{ab} + \omega^{ac} \wedge \omega^{cb} = \frac{1}{2} R^{ab}_{cd}\, e^c \wedge e^d. \tag{305}$$

The general formulas for ω_μ^{ab} and $R_{\mu\nu}^{ab}$ which appear in textbooks can be deduced from these definitions. However, for a reasonably simple metric ansatz and suitable choice of frame, it is frequently more convenient to use the definitions and compute directly. It takes some experience to learn to use the d and $\wedge$ operations efficiently. One must also remember to convert from frame to coordinate components of the curvature as needed.

For the domain wall metric a convenient frame is given by the transverse forms $e^{\hat{i}} = e^{A(r)} dx^i$, $i = 1 \cdots d$, and the radial form $e^D = dr$.

Ex. 20:. Use the Cartan structure equations with the frame 1-forms above to obtain the domain wall connection forms:

$$\omega^{\hat{i}\hat{j}} = 0 \quad \omega^{D\hat{i}} = A'(r) e^{\hat{i}} \tag{306}$$

Find next the curvature 2-forms:

$$\begin{aligned} R^{\hat{i}\hat{j}} &= -A'^2 e^{\hat{i}} \wedge e^{\hat{j}} \\ R^{\hat{i}D} &= -(A'' + A'^2) e^{\hat{i}} \wedge e^D \end{aligned} \tag{307}$$

Next obtain the curvature tensor (with coordinate indices)

$$\begin{aligned} R^{ij}_{kl} &= -A'^2 \left(\delta^i_k \delta^j_l - \delta^i_l \delta^j_k \right) \\ R^{iD}_{jD} &= -(A'' + A'^2) \delta^i_j \\ R^{ij}_{kD} &= 0 \end{aligned} \tag{308}$$

The final task is to find the Ricci tensor components

$$\begin{aligned} R_{ij} &= -e^{2A}(A'' + dA'^2) \delta_{ij} \\ R_{DD} &= -d(A'' + A')^2 \\ R_{iD} &= 0 \end{aligned} \tag{309}$$

Ex. 21:. If you still have some energy compute the non-vanishing components of the Christoffel connection, namely

$$\mathcal{G}^D_{ij} = -e^{2A} A' \delta_{ij} \qquad \mathcal{G}^i_{jD} = A' \delta^i_j \tag{310}$$

The fact that certain connection and curvature components vanish could have been seen in advance, since there are no possible Poincaré invariant tensors with the appropriate symmetries. We can introduce a new radial coordinate z, defined by $\frac{dz}{dr} = e^{-A(r)}$. This brings the domain wall metric to conformally flat form. It's Weyl tensor thus vanishes.

We now ask readers to manipulate the Einstein equation $G^\mu_\nu \equiv R^\mu_\nu - \frac{1}{2}\delta^\mu_\nu R = 2T^\mu_\nu$ for the domain wall and deduce a simple condition on the scale factor $A(r)$.

Ex. 22:. Deduce that

$$\begin{aligned} G^D_D &= \tfrac{d(d-1)}{2} A'^2 = 2T^D_D \\ G^i_j &= \delta^i_j (d-1)\left(A'' + \tfrac{1}{2} d A'^2\right) = 2T^i_j \end{aligned} \tag{311}$$

Compute $G^i_i - G^D_D$ for any fixed diagonal component (no sum on i)and deduce that

$$A'' = \frac{2}{d-1}\left(T^i_i - T^D_D\right) = -\frac{2}{d-1}\phi'^2 \tag{312}$$

Thus we certainly have $A'' < 0$ in the dynamics of the toy model. However there is a much more general result, namely $T^i_i - T^D_D < 0$ for any Poincaré invariant matter configuration in all conventional models for the bulk dynamics, for example, several scalars with non-linear σ-model kinetic term. In Lorentzian signature, the condition above is one of the standard energy conditions of general relativity. Later we will see the significance of the fact that $A''(r) < 0$.

Ex. 23:. Complete the analysis of the Einstein and scalar equations of motion for the domain wall and obtain the equations

$$\begin{aligned} A'^2 &= \tfrac{2}{d(d-1)}[\phi'^2 - 2V(\phi)] \\ \phi'' + dA'\phi' &= \tfrac{dV(\phi)}{d\phi} \end{aligned} \tag{313}$$

It is frequently the case that the set of equations obtained from a given ansatz for a gravity-matter system is not independent because of the Bianchi identity. Indeed in our system the derivative of the A'^2 equation combines simply with the the scalar equation to give (312). We can thus view the system (313) as independent.

It is easy to see how the previously discussed critical solutions fit into the domain wall framework. At each critical point ϕ_i of the potential, the scalar equation is satisfied by $\phi(r) \equiv \phi_i$. The A'^2 equation then gives $A(r) = \pm\frac{r}{L_i} + a_0$. The integration constant a_0 has no significance since it can be eliminated by scaling the coordinates x^i in (304). The sign above is a matter of convention and we choose the positive sign. The metric (304) is then equivalent to our previous description of AdS$_{d+1}$with the change of radial coordinate $z_0 = L_i e^{-\frac{r}{L_i}}$. With this sign convention we find that $r \to +\infty$ is the boundary region and $r \to -\infty$ is the deep interior.

Our main goal now is to discuss more general solutions of the system (313) in a potential of the type shown in Figure 12. We are interested in solutions which interpolate between two critical points, producing a domain wall geometry which approaches the boundary region of an AdS space with scale L_1 as $r \to +\infty$ and the deep interior of another AdS with scale L_2 as $r \to -\infty$. Such geometries are dual to field theories with RG flow.

To develop this interpretation let's first look at the quadratic approximation to the potential near a critical point,

$$V(\phi) \approx V(\phi_i) + \frac{1}{2}\frac{m_i^2}{L_i^2}h^2, \tag{314}$$

where we use the fluctuation $h = \phi - \phi_i$ and the scaled mass $m_i^2 = L_i^2 V''(\phi_i)$ with $V(\phi_i) = -d(d-1)/4L_i^2$. Let's recall the basic AdS/CFT idea that the boundary data for a bulk scalar field is the source for an operator in quantum field theory. We apply this to the fluctuation $h(r, \vec{x})$ which will be interpreted as the bulk dual of an operator $\mathcal{O}_\Delta(\vec{x})$ whose scale dimension is related to the mass m_i^2 by (261). Given the discussion of Sec. 8.3 it is reasonable to suppose that a general solution of the non-linear scalar equation of motion (301) will approach the critical point with the following boundary asymptotics for the fluctuation,

$$\begin{aligned} h(r,\vec{x}) &\underset{r\to\infty}{\longrightarrow} e^{(\Delta-d)r}\tilde{h}(\vec{x}) \\ &= e^{(\Delta-d)r}(\bar{\phi} + \bar{h}(\vec{x})). \end{aligned} \tag{315}$$

in which $\tilde{h}(\vec{x})$ contains $\bar{\phi}$, describing the boundary behavior of the domain wall profile plus a remainder $\bar{h}(\vec{x})$. We can form the on-shell action $S[\bar{\phi}+\bar{h}]$ which is a functional of this boundary data.[xvi]

A neat way to package the statement that the bulk on-shell action generates correlation functions in the boundary field theory is through the generating functional relation

$$\langle e^{-[S_{\rm CFT}+\int d^d\vec{x}\mathcal{O}_\Delta(\vec{x})(\bar{\phi}+\bar{h}(\vec{x}))]}\rangle = e^{-S[\bar{\phi}+\bar{h}]} \tag{316}$$

[xvi]A complete discussion should include the bulk metric which is coupled to $\phi(r, \vec{x})$. We have omitted this for simplicity. See Refs. 177, 178 for a recent general treatment.

in which $\langle\cdots\rangle$ on the left side indicates a path integral in the field theory. This is a simple generalization of a formula which we have implicitly used in Sec. 8.3 for CFT correlators, and $S_{\rm CFT}$ must still appear. The natural procedure in the present case is to define correlation functions by

$$\frac{(-)^{n-1}\delta^n}{\delta\bar{h}(\vec{x_1})\cdots\delta\bar{h}(\vec{x_n}))}S[\bar{\phi}+\bar{h}]\Big|_{\bar{h}=0}. \tag{317}$$

The term $\Delta S \equiv \int d^d\vec{x}\mathcal{O}_\Delta(\vec{x})\bar{\phi}$ then remains in the QFT Lagrangian and describes an operator deformation of the CFT with coupling constant $\bar{\phi}$. If $0 > m^2 > -\frac{d^2}{4} < 0$, that is if the critical point ϕ_i is a local maximum which is not too steep, then $d > \Delta > \frac{1}{2}d$, and we are describing a **relevant deformation** of a CFT_{UV}, one which will give a new long distance realization of the field theory. It is worth remarking that the lower bound agrees exactly with the stability criterion[179,180] for field theory in Lorentzian AdS_{d+1}. It is the lower mass limit for which the energy of normalized scalar field configurations is conserved and positive.

If the critical point is a local minimum, then $m_i^2 > 0$, and the dual operator has dimension $\Delta > d$. We thus have the deformation of the CFT by an **irrelevant** operator, exactly as describes the approach of an RG flow to a CFT_{IR} at long distance. We thus see the beginnings of a gravitational description of RG flows in quantum field theory!

9.2. *Interpolating Flows, I*

Interpolating flows are solutions of the domain wall equations (313) in which the scalar field $\phi(r)$ approaches the maximum ϕ_1 of $V(\phi)$ in Fig. 12 as $r \to +\infty$ and the minimum ϕ_2, as $r \to -\infty$. The associated metric approaches an AdS geometry in these limits as discussed in the previous section. Exact solutions of the second order non-linear system (313) are difficult (although we discuss an interesting method in the next section). However, we can learn a lot by linearizing about each critical point.

We thus set $\phi(r) = \phi_i + h(r)$ and $A' = \frac{1}{L_i} + a'(r)$ and work with the quadratic approximate potential in (314). (See Footnote xiv.) The linearized scalar equation of motion and its general solution are

$$h'' + \frac{d}{L_i}h' - \frac{m_i^2}{L_i^2}h = 0 \tag{318}$$

$$h(r) = Be^{(\Delta_i - d)r/L_i} + Ce^{-\Delta_i r/L} \tag{319}$$

$$\Delta_i = \left(d + \sqrt{d^2 + 4m_i^2}\right)/2 \tag{320}$$

One may then linearize the scale factor equation in (313) to find $a' = \mathcal{O}(h^2)$ so that the scale factor $A(r)$ is not modified to linear order

Ex. 24:. Verify the statements above.

The basic idea of linearization theory is that there is an exact solution of the nonlinear equations of motion that is well approximated by a linear solution near a critical point. Thus as $r \to +\infty$, we assume that the exact solution behaves as

$$\phi(r) \underset{r \gg 0}{\approx} \phi_1 + B_1 e^{(\Delta_1 - d)r/L_1} + C_1 e^{-\Delta_1 r/L_1}. \tag{321}$$

The fluctuation must disappear as $r \to +\infty$. For a generic situation in which the dominant B term is present, this requires $\frac{d}{2} < \Delta_1 < d$ or $m_1^2 < 0$. Hence the critical point associated with the boundary region of the domain wall must be a local maximum, and everything is consistent with an interpretation as the dual of a QFT_d which is a relevant deformation of an ultraviolet CFT_d.

Near the critical point ϕ_2, which is a minimum, we have $m_2^2 > 0$ so $\Delta_2 > d$. This critical point must be approached at large negative r, where the exact solution is approximated by

$$\phi(r) \underset{r \ll 0}{\approx} \phi_2 + B_2 e^{(\Delta_2 - d)r/L_2} + C_2 e^{-\Delta_x r/L_2}. \tag{322}$$

The second term diverges, so we must choose the solution with $C_2 = 0$. Thus the domain wall approaches the deep interior region with the scaling rate of an irrelevant operator of scale dimension $\Delta_2 > d$ exactly as required for infrared fixed points by RG ideas on field theory.

The non-linear equation of motion for $\phi(r)$ has two integration constants. We must fix one of them to ensure $C = 0$ as $r \to -\infty$. The remaining freedom is just the shift $r \to r + r_0$ and has no effect on the physical picture. A generic solution with $C = 0$ in the IR would be expected to approach the UV critical point at the dominant rate $Be^{(\Delta_1 - d)r/L_1}$ which we have seen to be dual to a relevant operator deformation of the CFT_{UV}. It is possible (but exceptional) that the $C = 0$ solution in the IR would have vanishing B term in the UV and approach the boundary as $C_1 e^{-\Delta_1 r/L_1}$. In this case the physical interpretation is that of the deformation of the CFT_{UV} by a vacuum expectation value, $\langle \mathcal{O}_{\Delta_1} \rangle \sim C_1 \neq 0$. See Refs. 181, 182, 183, 170.

The domain wall flow "sees" the AdS_{IR} geometry only in the deep interior limit. To discuss the CFT_{IR} and its operator perturbations in themselves, we must think of extending this interior region out to a complete AdS_{d+1}geometry with scale $L_{IR} = L_2$.

The interpolating solution we are discussing is plotted in Figure 13. The scale factor $A(r)$ is concave downward since $A''(r) < 0$ from (312). This means that slopes of the linear regions in the deep interior and near boundary are related by $1/L_{IR} > 1/L_{UV}$ (where we have set $L_{UV} = L_1$). Hence,

$$V_{IR} = \frac{-d(d-1)}{4L_{IR}^2} < V_{UV} = \frac{-d(d-1)}{4L_{UV}^2}. \tag{323}$$

Thus the flow from the boundary to the interior necessarily goes to a deeper critical point of $V(\phi)$. Recall that the condition $A''(r) < 0$ is very general and holds in any physically reasonable bulk theory, *e.g.* a system of many scalars ϕ^I and potential $V(\phi^I)$. Thus any Poincaré invariant domain wall interpolating between AdS geometries is **irreversible**.

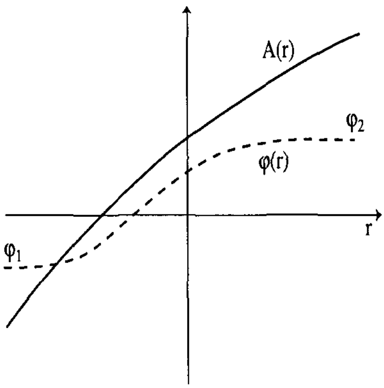


Figure 13. Profile of the scale factor $A(r)$

The philosophy of the AdS/CFT correspondence suggests that any conspicuous feature of the bulk dynamics should be dual to a conspicuous property of quantum field theory. The irreversibility property reminds us of Zamolodchikov's c-theorem[184] which implies that RG flow in QFT_2 is irreversible. We will discuss the c-theorem and its holographic counterpart later. Our immediate goals are to present a very interesting technique for exact solutions of the non-linear flow equations (313) and to discuss a "realistic" application of supergravity domain walls to deformations of $\mathcal{N}=4$ SYM theory.

9.3. *Interpolating Flows, II*

The domain wall equations

$$\phi'' + dA'\phi' = \frac{dV(\phi)}{d\phi} \tag{324}$$

$$A'^2 = \frac{2}{d(d-1)}\left(\phi'^2 - 2V(\phi)\right) \tag{325}$$

constitute a non-linear second order system with no apparent method of analytic solution. Nevertheless, a very interesting procedure which does give exact solutions in a number of examples has emerged from the literature.[185–188]

Given the potential $V(\phi)$, suppose we could solve the following differential equation in field space and obtain an auxiliary quantity, the superpotential $W(\phi)$:

$$\frac{1}{2}\left(\frac{dW}{d\phi}\right)^2 - \frac{d}{d-1}W^2 = V(\phi) \tag{326}$$

We then consider the following set of first order equations

$$\frac{d\phi}{dr} = \frac{dW}{d\phi} \tag{327}$$

$$\frac{dA}{dr} = -\frac{2}{d-1}W\left(\phi(r)\right) \tag{328}$$

These decoupled equations have a trivial structure and can be solved sequentially, the first by separation of variables, and the second by direct integration. (We assume that the two required integrals are tractable.) It is then easy to show that **any solution of the first order system (326, 327, 328) is also a solution of the original second order system (324, 325).**

Ex. 25:. Prove this!

It is also elementary to see that any critical point of $W(\phi)$ is also a critical point of $V(\phi)$ but not conversely.

Ex. 26:. Suppose that $W(\phi)$ takes the form $W \approx -\frac{1}{L_i}(\lambda + \frac{1}{2}\mu h^2)$ near a critical point. Show that λ, μ are related to the parameters of the approximate potential in (314) by $\lambda = \frac{1}{2}(d-1)$ and $m^2 = \mu(\mu - d)$. Show that the solution to the flow equation (327) approaches the critical point at the rate $h \sim e^{-\mu r/L}$ with $\mu = \Delta$, the vev rate, or $\mu = d - \Delta$ the operator deformation rate.

This apparently miraculous structure generalizes to bulk theories with several scalars ϕ^I and Lagrangian

$$L = -\frac{1}{4}R + \frac{1}{2}\partial_\mu \phi^I \partial^\mu \phi^I + V(\phi^I) \tag{329}$$

The superpotential $W(\phi^I)$ is defined to satisfy the partial differential equation

$$\frac{1}{2}\sum_I \left(\frac{\partial W}{\partial \phi^I}\right)^2 - \frac{d}{d-1}W^2 = V \tag{330}$$

The first order flow equations

$$\frac{d\phi^I}{dr} = \frac{\partial W}{\partial \phi^I} \tag{331}$$

$$\frac{dA}{dr} = -\frac{2}{d-1}W \tag{332}$$

automatically give a solution of the second order Euler-Lagrange equations of (329) for Poincaré invariant domain walls.

Ex. 27:. Prove this and derive first order flow equations with the same property for the non-linear σ-model (in which the kinetic term of (329) is replaced by $\frac{1}{2}G_{IJ}(\phi^K)\partial_\mu \phi^I \partial^\mu \phi^J$.

The equations (331) are conventional gradient flow equations. The solutions are paths of steepest descent for $W(\phi^I)$, everywhere perpendicular to the contours $W(\phi^I) =$ const. In applications to RG flows, the $\phi^I(r)$ represent scale dependent couplings of relevant operators in a QFT Lagrangian, so we are talking about gradient flow in the space of couplings—an idea which is frequently discussed in the RG literature!

There are two interesting reasons why there are first order flow equations which reproduce the dynamics of the second order system (324, 325).

(1) They emerge as BPS conditions for supersymmetric domain walls in supergravity theories. For a review, see Ref. 189. The superpotential $W(\phi^I)$ emerges by algebraic analysis of the quantum transformation rule. Bulk solutions have Killing spinors, and bulk supersymmetry is matched in the boundary field theory which describes a supersymmetric deformation of an SCFT.

(2) They are the Hamilton-Jacobi equations for the dynamical system of gravity and scalars.[190] The superpotential $W(\phi^I)$ is the classical Hamilton-Jacobi function, and one must solve (326) or (330) to obtain it from the potential $V(\phi^I)$. This is very interesting theoretically but rather impractical because it is rare that one can actually use the Hamilton-Jacobi formulation to solve a dynamical system explicitly. Numerical and approximate studies have been instructive.[188,191] However, most applications involve superpotentials from BPS conditions in gauged supergravity. One may also employ a toy model viewpoint in which $W(\phi^I)$ is postulated with potential $V(\phi^I)$ defined through (326) or (330).

9.4. *Domain Walls in $D = 5$, $\mathcal{N} = 8$, Gauged Supergravity*

The framework of toy models is useful to illustrate the correspondence between domain walls in $(d+1)$-dimensional gravity and RG flows in QFT_d. However it is highly desirable to have "realistic examples" which describe deformations of $\mathcal{N} = 4$ SYM in the strong coupling limit of the AdS/CFT correspondence. There are two reasons to think first about supersymmetric deformations. As just discussed, the bulk dynamics is then governed by a superpotential $W(\phi^I)$ with first order flow equations. Further the methods of Seiberg dynamics can give control of the long distance non-perturbative behavior of the field theory, so that features of the supergravity description can be checked.

We can only give a brief discussion here. We begin by discussing the relation between $D = 10$ Type IIB sugra dimensionally reduced on $\mathrm{AdS}_5 \times \mathrm{S}^5$ of Ref. 50 and the $D = 5$, $\mathcal{N} = 8$ supergravity theory with gauge group $SO(6)$ first completely constructed in Ref. 192. As discussed in Section 9.3 above, the spectrum of the first theory consists of the graviton multiplet, whose fields are dual due to all relevant and marginal operators of $\mathcal{N} = 4$ SYM plus Kaluza-Klein towers of fields dual to operators of increasing Δ. On the other hand gauged $\mathcal{N} = 8$ supergravity is a theory formulated in 5 dimensions with only the fields of the graviton multiplet above, namely

$$\begin{array}{cccccc} g_{\mu\nu} & \Psi_\mu^a & A_\mu^A & B_{\mu\nu} & X^{abc} & \phi^I \\ 1 & 8 & 15 & 12 & 48 & 42 \end{array} \tag{333}$$

It is a complicated theory in which the scalar dynamics is that of a nonlinear σ-model on the coset $E(6,6)/USp(8)$ with a complicated potential $V(\phi^I)$.

Gauged $\mathcal{N}=8$ supergravity has a maximally symmetric ground state in which the metric is that of AdS$_5$. The global symmetry is $SO(6)$ with 32 supercharges, so that the superalgebra is $SU(2,2|4)$. Symmetries then match the vacuum configuration of Type IIB sugra on AdS$_5 \times S^5$. Indeed $D=5$, $\mathcal{N}=8$ sugra is believed to be the **consistent truncation** of $D=10$ Type IIB sugra to the fields of its graviton multiplet. This means that **every** classical solution of $D=5$, $\mathcal{N}=8$ sugra can be "lifted" to a solution of $D=10$ Type IIB sugra. For example the $SO(6)$ invariant AdS$_5$ ground state solution lifts to the AdS$_5 \times$ S^5 geometry of (244) (with other fields either vanishing or maximally symmetric). There is not yet a general proof of consistent truncation, but explicit lifts of nontrivial domain wall solutions have been given.[193–196] Consistent truncation has been established in other similar theories.[197,198]

In the search for classical solutions with field theory duals it is more elegant, more geometric, and more "braney" to work at the level of $D=10$ Type IIB sugra. There are indeed very interesting examples of Polchinski and Strassler[199] and Klebanov and Strassler.[200] Another example is the multi-center $D3$-brane solution of (241,243) which is dual to a Higgs deformation of $\mathcal{N}=4$ SYM in which the $SU(N)$ gauge symmetry is broken spontaneously to $SU(N_1)\otimes SU(N_2)\otimes\cdots\otimes SU(N_M)$. In these examples the connection with field theory is somewhat different from the emphasis in the present notes. For this reason we confine our discussion to domain wall solutions of $D=5$ $\mathcal{N}=8$ sugra. This is a realistic framework since the $D=5$ theory contains all relevant deformations of $\mathcal{N}=4$ SYM, and experience indicates that $5D$ domain wall solutions can be lifted to solutions of $D=10$ Type IIB sugra.

For domain walls, we can restrict to the metric and scalars of the theory which are governed by the action

$$S=\int d^5z\sqrt{g}\left[-\frac{1}{4}R+\frac{1}{2}G_{IJ}(\phi^K)\partial_\mu\phi^I\partial^\mu\phi^J+V(\phi^k)\right] \tag{334}$$

The 42 scalars sit in the 27-bein matrix $V_{AB}^{ab}(\phi^K)$ of $E(6,6)/USp(8)$. The indices a, b and AB have 8 values. They are anti-symmetrized (with symplectic trace removed) in most expressions we write. The coset metric G_{IJ}, the potential V and other quantities in the theory are constructed from V_{AB}^{ab}. Symmetries govern the construction, but the nested structure of symmetries makes things very complicated.

A simpler question than domain walls is that of critical points of $V(\phi^k)$. The AdS/CFT correspondence requires that every stable critical point with $V < 0$ corresponds to a CFT_4. Stability means simply that mass eigenvalues of fluctuations satisfy $m^2 > -4$ so that bulk fields transform in unitary, positive energy representations of $SO(2,4)$ (for Lorentz signature).

Even the task of extremizing $V(\phi^k)$ is essentially impossible in a space of 40 variables, (V does not depend on the dilaton and axion fields), so one uses the following simple but practically important trick:[201]

(a) Select a subgroup H of the invariance group $SO(6)$ of $V(\phi^K)$.
(b) The $42\phi^K$ may be grouped into fields ϕ which are singlets of H and others ξ which transform in non-trivial representations of H.
(c) It follows from naive group theory that the expansion of V takes the form $V(\phi,\xi) = V_0(\phi) + V_2(\phi)\xi^2 + \mathcal{O}(\xi^3)$ with no linear term.
(d) Thus, if $\widehat{\phi}$ is a stationary point of $V_0(\phi)$, then $\widehat{\phi}, \xi = 0$ is a stationary point of $V(\phi,\xi)$. The problem is then reduced to minimization in a much smaller space.

The same method applies to all solutions of the equations of motion $\frac{\partial S}{\partial \phi^K} = 0$, and to the Killing spinor problem since that gives a solution to the equations of motion. The general principle is that if S is invariant under G, in this case $G = SO(6)$, and $H \subset G$ is a subgroup, then a consistent H-invariant solution to the dynamics can be obtained by restricting, ab initio, to singlets of H.

All critical points with preserved symmetry $H \supseteq SU(2)$ are known.[202] There are 5 critical points of which 3 are non-supersymmetric and unstable.[192,203] There are two $SUSY$ critical points of concern to us. The first with $H = SO(6)$ and full $\mathcal{N} = 8$ $SUSY$ is the maximally symmetric state discussed above, and the second has $H = SU(2) \otimes U(1)$ and $\mathcal{N} = 2$ $SUSY$. The associated critical bulk solutions are dual to the undeformed $\mathcal{N} = 4$ SYM and the critical IR limit of a particular deformation of $\mathcal{N} = 4$ SYM.

The search for supersymmetric domain walls in $\mathcal{N}=8$ gauged supergravity begins with the fermionic transformation rules[xvii] which have the form:

$$\delta\psi^a_\mu = D_\mu\epsilon^a - \frac{1}{3}W^a_b\gamma_\mu\epsilon^b \tag{335}$$

$$\delta\chi^A = \left(\gamma^\mu P^A_{aI}\partial\phi^I - Q^A_a(\varphi)\right)\epsilon^a \tag{336}$$

where A is an index for the 48 spinor fields χ^{abc}. The ϵ^a are 4-component symplectic Majorana spinors[186] (with spinor indices suppressed and $a = 1, \cdots, 8$). The matrices W^a_b, P^A_{aI} and Q^A_a are functions of the scalars φ^I which are part of the specification of the classical supergravity theory. Killing spinors $\epsilon^a(\vec{x}, r)$ are spinor configurations which satisfy $\delta\psi^a_\mu=0$ and $\delta\chi^A=0$. The process of solving these equations leads both to the $\epsilon^a(\vec{x}, r)$ and to conditions which determine the domain wall geometry which supports them. These conditions, in this case the first order field equations (331, 332), imply that the bosonic equations of motion of the theory are satisfied.

Ex. 28:. For a generic SUSY or sugra theory, show that if there are Killing spinors for a given configuration of bosonic fields, then that configuration satisfies the equations of motion. Hint:

$$\delta S = \int \left[\frac{\delta S}{\delta B}\delta B + \frac{\delta S}{\delta\psi}\delta\psi\right] \equiv 0, \tag{337}$$

where B and ψ denote the boson and fermion fields of the theory and δB and $\delta\psi$ their transformation rules.

We now discuss the Killing spinor analysis to outline how the first order flow equations arise.

Ex. 29:. Using the spin connection of Ex: 20, show that the condition $\delta\psi^a_j=0$ can be written out in detail as

$$\delta\psi^a_j = \partial_j\epsilon - \frac{1}{2}A'(r)\gamma_j\gamma_5\epsilon^a - \frac{1}{3}W^a_b\gamma_j\epsilon^b = 0 \tag{338}$$

We can drop the first term because the Killing spinor must be translation invariant. What remains is a purely algebraic condition, and we can see that the flow equation (328) for the scale factor directly emerges with superpotential $W(\phi)$ identified as one of the eigenvalues of the tensor W^a_b. In

[xvii] Conventions for spinors and γ-matrices are those of Ref. 186 with spacetime signature $+----$.

detail one actually has a symplectic eigenvalue problem, with 4 generically distinct W's as solutions. Each of these is a candidate superpotential. One must then examine the 48 conditions

$$\delta\chi^A = (\gamma^5 P^A_{aI}\partial_r\phi^I - Q^A_a)\epsilon^a = 0 \tag{339}$$

to see if $SUSY$ is supported on any of the eigenspaces. One can see how the gradient flow equation (331) can emerge. Success is not guaranteed, but when it occurs, it generically occurs on one of the four (symplectic) eigenspaces. The $5D$ Killing spinor solution satisfies a γ^5 condition effectively yielding a $4d$ Weyl spinor, giving $\mathcal{N} = 1$ $SUSY$ in the dual field theory. Extended $\mathcal{N} > 1$ $SUSY$ requires further degeneracy of the eigenvalues. (The $\delta\Psi^a_r = 0$ condition which has not yet been mentioned gives a differential equation for the r-dependence of $\epsilon^a(r)$)

Ex. 30:. It is a useful exercise to consider a simplified version of the Killing spinor problem involving one complex (Dirac) spinor with superpotential $W(\phi)$ with one scalar field. The equations are

$$\begin{aligned}(D_\mu - \tfrac{1}{3}iW\gamma_\mu)\epsilon &= 0\\ (-i\gamma^\mu\partial_\mu\phi - \tfrac{dW}{d\phi})\epsilon &= 0\end{aligned} \tag{340}$$

Show that the solution of this problem yields the flow equations (327,328) and

$$\epsilon = e^{\frac{A}{2}}\eta \tag{341}$$

where η is a constant eigenspinor of γ^5. Show that at a critical point of W, there is a second Killing spinor (which depends on the transverse coordinates x^i). See Ref. 69. This appears because of the the doubling of supercharges in superconformal $SUSY$.

Needless to say the analysis is impossible on the full space of 42 scalars. Nor do we expect a solution in general, since many domain walls are dual to non-supersymmetric deformations and cannot have Killing spinors. In Ref. 186, a symmetry reduction to singlets of an $SU(2)$ subgroup of $SO(6)$ was used. After further simplification it was found that $\mathcal{N} = 1$ $SUSY$ with $SU(2) \times U(1)$ global symmetry was supported for flows involving two scalar fields, ϕ_2 a field with $\Delta = 2$ in the $20'$ of $SO(6)$ in the full theory, and ϕ_3 a field with $\Delta = 3$ in the $10 + \overline{10}$ representation. (In,[186] these fields were called ϕ_3, ϕ_1 respectively.) The fields ϕ_2, ϕ_3 have canonical kinetic terms as in (329). Using $\rho = e^{\frac{\phi_2}{\sqrt{6}}}$, the superpotential is

$$W(\phi_2, \phi_3) = \frac{1}{4L\rho^2}\left[\cosh(2\phi_3)(\rho^6 - 2) - 3\rho^6 - 2\right] \tag{342}$$

Ex. 31:. Show that $W(\phi_2, \phi_3)$ has the following critical points:

(i) a maximum at $\phi_2 = 0, \phi_3 = 0$, at which $W = -\frac{3}{2L}$
(ii) a saddle point at $\phi_2 = \frac{1}{\sqrt{6}}\ln 2, \phi_3 = \pm\frac{1}{2}\ln 3$ at which $W = -\frac{2^{2/3}}{L}$. (The two solutions are related by a Z_2 symmetry and are equivalent).

Thus there is a possible domain wall flow interpolating between these two critical points. The flow equation (331) cannot yet be solved analytically for W of (342), but a numerical solution and its asymptotic properties were discussed in Ref. 186. See Figure 14.

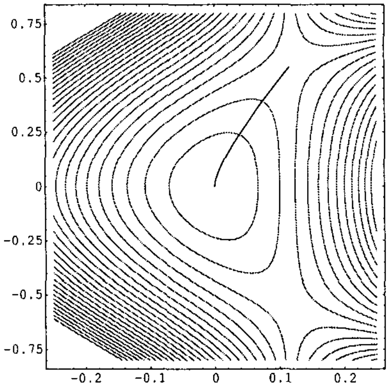


Figure 14. Contour plot of $W(\phi_2, \phi_3)$

In accord with the general discussion of Section 9.2, the solution should be dual to a relevant deformation of $\mathcal{N} = 4$ SYM theory which breaks $SUSY$ to $\mathcal{N} = 1$ and flows to an SCFT$_4$ at long distance. In the next section we discuss this field theory and the evidence that the supergravity description is correct.

In the space of the two bulk fields ϕ_2, ϕ_3 there is a continuously infinite set of gradient flow trajectories emerging from the $\mathcal{N} = 4$ critical point. One must tune the initial direction to find the one which terminates at the $\mathcal{N} = 1$ point. All other trajectories approach infinite values in field space, and the associated geometries, obtained from the flow equation (332) for

$A(r)$ have curvature singularities. There are analytic domain wall flows with $\phi_3 \equiv 0$ (Refs. 204, 295) and in other sectors of the space of scalars of $\mathcal{N} = 8$ sugra, and a number of 2-point correlation functions have been computed.[177,178,207–210] Nevertheless the curvature singularities are at least a conceptual problem for the AdS/CFT correspondence. In any case we do not discuss singular flows here.

9.5. *SUSY Deformations of* $\mathcal{N}=4$ *SYM Theory*

It is useful for several purposes to describe the $\mathcal{N} = 4$ SYM theory in terms of $\mathcal{N} = 1$ superfields. The 4 spinor fields λ^α are regrouped, and λ^4 is paired with gauge potential A_j in a gauge vector superfield V. The remaining $\lambda^{1,2,3}$ may be renamed $\psi^{1,2,3}$ and paired with complex scalars $z^1 = X^1 + iX^4, z^2 = X^2 + iX^5, z^3 = X^3 + iX^6$ to form 3 chiral superfields Φ^i. In the notation of Section 2.5, the Lagrangian consists of a gauge kinetic term plus matter terms

$$\mathcal{L} = \int d^4\theta \mathrm{tr}(\bar{\Phi}^i e^{gV} \Phi^i) + \int d^2\theta g \mathrm{tr}\Phi^3[\Phi^1, \Phi^2] + h.c. \tag{343}$$

The manifest supersymmetry is $\mathcal{N} = 1$ with R-symmetry $SU(3) \otimes U(1)$. Full symmetry is regained after re-expression in components because the Yukawa coupling g is locked to the $SU(N)$ gauge coupling. This formulation is commonly used to explore perturbative issues since the $\mathcal{N} = 1$ supergraph formalism (first reference in Ref. 13) is quite efficient. This formulation suits our main purpose which is to discuss *SUSY* deformations of the theory.

A general relevant $\mathcal{N} = 1$ perturbation of $\mathcal{N} = 4$ SYM is obtained by considering the modified superpotential

$$U = g\mathrm{tr}\Phi^3[\Phi^1, \Phi^2] + \frac{1}{2} M_{\alpha\beta} \mathrm{tr}\Phi^\alpha \Phi^\beta \tag{344}$$

This framework is called the $\mathcal{N} = 1^*$ theory. The moduli space of vacua, $\frac{\partial U}{\partial \Phi^\alpha} = 0$, has been studied[199,211,212] and describes a rich panoply of dynamical realizations of gauge theories, confinement and Higgs-Coulomb phases, and as we shall see, a superconformal phase.

We discuss here the particular deformation with a mass term for one chiral superfield only

$$U = g\mathrm{tr}\Phi^3[\Phi^1, \Phi^2] + \frac{1}{2} m \mathrm{tr}(\Phi^3)^2 \tag{345}$$

The R-symmetry is now the direct product of $SU(2)$ acting on $\Phi^{1,2}$ and $U(1)_R$ with charges $(\frac{1}{2}, \frac{1}{2}, 1)$ for $\Phi^{1,2,3}$. The massive field Φ^3 drops out of the long distance dynamics, leaving the massless fields $\Phi^{1,2}$. We thus find symmetries which match those of the supergravity flow of the last section. However to establish the duality, it needs to be shown that long distance dynamics is conformal. We will briefly discuss the pretty arguments of Leigh and Strassler[213] that show this is the case.

The key condition for conformal symmetry is the vanishing of β-functions for the various couplings in the Lagrangian. In a general $\mathcal{N}=1$ theory with gauge group G and chiral superfields Φ_α in representations R_α of G, the exact $NSVZ$ gauge β-function is

$$\beta(g) = -\frac{g^3}{8\pi^2}\frac{3T(G)-\sum_\alpha T(R_\alpha)(1-2\gamma_\alpha)}{1-\frac{g^2T(G)}{8\pi^2}} \tag{346}$$

where γ_α is the anomalous dimension of Φ^α and $T(R_\alpha)$ is the Dynkin index of the representation. (If T^a are the generators in the representation R_α, then $\mathrm{tr}T^aT^b \equiv T(R_\alpha)\delta^{ab}$). In the present case, in which $G = SU(N)$ and all fields are in the adjoint, we have $T(R_\alpha) = T(G) = N$, and

$$\beta(g) \sim 2N(\gamma_1+\gamma_2+\gamma_3) \tag{347}$$

In addition we need the β-function for various invariant field monomials $\mathrm{Tr}(\Phi^1)^{n_1}(\Phi^2)^{n_2}(\Phi^3)^{n_3}$,

$$\beta_{n_1,n_2,n_3} = 3 - \sum_{\alpha=1}^{3} n_\alpha - \sum_{\alpha=1}^{3} n_\alpha\gamma_\alpha \tag{348}$$

This form is a consequence of the non-renormalization theorem for superpotentials in $\mathcal{N}=1$ $SUSY$. The first two terms are fixed by classical dimensions and the last is due to wave function renormalization. For the two couplings in the superpotential (345) we have

$$\beta_{1,1,1} = \gamma_1+\gamma_2+\gamma_3 \tag{349}$$

$$\beta_{0,0,2} = 1-2\gamma_3. \tag{350}$$

One should view the $\gamma_\alpha(g,m)$ as functions of the two couplings. The conditions for the vanishing of the 3 β-functions have the unique $SU(2)$ invariant solution

$$\gamma_1 = \gamma_2 = -\frac{1}{2}\gamma_3 = -\frac{1}{4} \tag{351}$$

which imposes one relation between g, m, suggesting that the theory has a fixed line of couplings. The $\beta = 0$ conditions are necessary conditions for a superconformal realization in the infrared, and Leigh and Strassler give additional arguments that the conformal phase is realized.

$\mathcal{N} = 1$ superconformal symmetry in 4 dimensions is governed by the superalgebra $SU(2,2,|1)$. This superalgebra has several types of short representations. (See Appendix of Ref. 186). For example, chiral superfields, either elementary or composite, are short multiplets in which scale dimensions and $U(1)_R$ charge are related by $\Delta = \frac{3}{2}r$. For elementary fields $\Delta_\alpha = 1 + \gamma_\alpha$, and one can see that the γ_α values in (351) are correctly related to the $U(1)_R$ charges of the Φ^α.

The observables in the SCFT_{IR} are the correlation functions of gauge invariant composites of the light superfields[xviii] W_α, Φ^1, Φ^2. We list several short multiplets together with the scale dimensions of their primary components

$$\begin{array}{cccccc} \mathcal{O} & \mathrm{tr}\Phi^\alpha\Phi^\beta & \mathrm{tr}W_\alpha\Phi^\beta & \mathrm{tr}W_\alpha W^\alpha & \mathrm{tr}\Phi^+T^A\Phi & \mathrm{tr}(W_\alpha W_{\dot\beta}+\cdots) \\ \Delta & \frac{3}{2} & \frac{9}{4} & 3 & 2 & 3 \end{array} \tag{352}$$

The first 3 operators are chiral, the next is the multiplet containing the $SU(2)$ current, and the last is the multiplet containing $U(1)_R$ current, supercurrent, and stress tensor. Each multiplet has several components.

9.6. *AdS/CFT Duality for the Leigh-Strassler Deformation*

We now discuss the evidence that the domain wall of $\mathcal{N} = 8$ gauged supergravity of Section 9.4 is the dual of the mass deformation of $\mathcal{N} = 4$ SYM of Section 9.5. There are two types of evidence, the match of dimensions of operators, discussed here, and the match of conformal anomalies discussed in the next chapter. Critics may argue that much of the detailed evidence is a consequence of symmetries rather than dynamics. But it is dynamically significant that the potential $V(\Phi^k)$ contains an IR critical point with the correct symmetries and the correct ratio V_{IR}/V_{UV} to describe the IR fixed point of the Leigh-Strassler theory. The AdS/CFT correspondence would be incomplete if $D = 5$ $\mathcal{N} = 8$ sugra did not contain this SCFT_4.

Whether due to symmetries or dynamics, much of the initial enthusiasm for AdS/CFT came from the $1:1$ map between bulk fields of Type IIB sugra and composite operators of $\mathcal{N} = 4$ SYM. The map was established using the relationship between the AdS masses of fluctuations about the

[xviii]the index of the field strength superfield W_a is that of a Lorentz group spinor, while that of Φ^α is that of $SU(2)$ flavor.

$AdS_5 \times S^5$ solution and scale dimensions of operators. The same idea may be applied to fluctuations about the IR critical point of the flow of Section 9.4. One can check the holographic description of the dynamics by computing the mass eigenvalues of all fields in the theory, namely all fields of the graviton multiplet listed in Section 9.4. This task is complicated because the Higgs mechanism acts in several sectors. Scale dimensions are then assigned using the formula in (261) for scalars and its generalizations to other spins. The next step is to assemble component fields into multiplets of the $SU(2,2|1)$ superalgebra. One finds exactly the 5 short multiplets listed at the end of Section 9.5 together with 4 long representations. The detailed match of short multiplets confirms the supergravity description, while the scale dimensions of operators in long representations are non-perturbative predictions of the supergravity description.

It would be highly desirable to study correlation functions of operators in the Leigh-Strassler flow, but this requires an analytic solution for the domain wall, which is so far unavailable.

9.7. *Scale Dimension and AdS Mass*

For completeness we now list the relation between Δ and the mass for the various bulk fields which occur in a supergravity theory. For $d = 4$ some results were given in Ref. 167. For the general case of for AdS_{d+1}, the relations are given below with references. There are exceptional cases in which the lower root of $\pm$ is appropriate.

(1) scalars:[3] $\Delta_\pm = \frac{1}{2}(d \pm \sqrt{d^2 + 4m^2})$,
(2) spinors:[214] $\Delta = \frac{1}{2}(d + 2|m|)$,
(3) vectors $\Delta_\pm = \frac{1}{2}(d \pm \sqrt{(d-2)^2 + 4m^2})$,
(4) p-forms:[147] $\Delta = \frac{1}{2}(d \pm \sqrt{(d-2p)^2 + 4m^2})$,
(5) first-order $(d/2)$-forms (d even): $\Delta = \frac{1}{2}(d + 2|m|)$,
(6) spin-3/2 :[215,216] $\Delta = \frac{1}{2}(d + 2|m|)$,
(7) massless spin-2:[217] $\Delta = \frac{1}{2}(d + \sqrt{d^2 + 4m^2})$.

10. The c-Theorem and Conformal Anomalies

In this chapter we develop a theme introduced in Sec. 9.2, the irreversibility of domain walls in supergravity and the suggested connection with the c-theorem for RG flows in field theory. The c-theorem is related to the conformal anomaly. We discuss this anomaly for 4d field theory and the elegant way it is treated in the AdS/CFT correspondence. This suggests a simple form for a holographic c-function, and monotonicity follows from the equation $A''(r) < 0$. It follows that any RG flow which can be described by the AdS/CFT correspondence satisfies the c-theorem. The holographic computation of anomalies agrees with field theory for both the undeformed $\mathcal{N} = 4$ SYM theory and the $\mathcal{N} = 1$ Leigh-Strassler deformation.

10.1. *The c-Theorem in Field Theory*

We briefly summarize the essential content of Zamolodchikov's c-theorem[184] which proves that RG flows in QFT_2 are irreversible. We consider the correlator $\langle T_{zz}(z,\bar{z})T_{zz}(0)\rangle$ in a flow from a CFT_{UV} to a CFT_{IR}. It has the form

$$\langle T_{zz}(z,\bar{z})T_{zz}(0)\rangle = \frac{c(M^2 z\bar{z})}{z^4} \tag{353}$$

where M^2 is a scale that is present since conformal symmetry is broken. The function $c(M^2 z\bar{z})$ has the properties:

(1) $c(M^2 z\bar{z}) \to c_{UV}$ as $|z| \to 0$ and $c(M^2 z\bar{z}) \to c_{IR}$ as $|z| \to \infty$ where c_{UV} and c_{IR} are central charges of the critical theories CFT_{UV} and CFT_{IR}.

(2) $c(M^2 z\bar{z})$ is not necessarily monotonic, but there are other (non-unique) c-functions which decrease monotonically toward the infrared and agree with $c(M^2 z\bar{z})$ at fixed points. Hence $c_{UV} > c_{IR}$ which proves irreversibility of the flow!

(3) the central charges are also measured by the curved space Weyl anomaly in which the field theory is coupled to a fixed external metric g_{ij} and one has

$$\langle \theta \rangle = -\frac{c}{12} R \tag{354}$$

for both CFT_{UV} and CFT_{IR}.

The intuition for the c-theorem comes from the ideas of Wilsonian renormalization and the decoupling of heavy particles at low energy. Since T_{ij} couples to all the degrees of freedom of a theory, the c-function measures the effective number of degrees of freedom at scale $x = \sqrt{z\bar{z}}$. This number decreases monotonically as we proceed toward longer distance and more and more heavy particles decouple from the low energy dynamics. These are fundamental ideas and we should see if and how they are realized in QFT_4 and $\mathrm{AdS}_5/\mathrm{CFT}_4$.

First we define two projection operators constructed from the basic $\pi_{ij} = \partial_i \partial_j - \delta_{ij} \Box$,

$$\begin{aligned} \Pi^{(0)}_{ijkl} &= \pi_{ij} \pi_{kl} \\ \partial^{(2)}_{ijkl} &= 2\pi_{ij}\pi_{kl} - 3(\pi_{ik}\Pi_{jl} + \pi_{il}\pi_{jk}) \end{aligned} \tag{355}$$

In any QFT$_4$ the $\langle TT \rangle$ correlator then takes the form

$$\langle T_{ij}(x) T_{kl}(0) \rangle = \frac{-1}{48\pi^4} P^{(2)}_{ijkl} \frac{c(m^2 x^2)}{x^4} + P^{(0)}_{ijkl} \frac{f(M^2 x^2)}{x^4} \tag{356}$$

In a flow between two CFT's, the central function[218] $c(m^2x^2)$ approaches central charges c_{UV}, c_{IR} in the appropriate limits, but $f(M^2x^2) \to 0$ in the UV and IR since effects of the trace T^i_i must vanish in conformal limits.

The correlators of T_{ij} can be obtained from a generating functional formally constructed by coupling the flat space theory covariantly to a non-dynamical background metric $g_{ij}(x)$. For example, in a gauge theory one would take

$$S[g_{ij}, A_k] \equiv \frac{1}{4} \int d^4x \sqrt{g} g^{ik} g^{jl} F_{ij} F_{kl} \tag{357}$$

The effective action is then defined as the path integral over elementary fields, *e.g.*

$$e^{-S_{eff}[g]} \equiv \int [dA_i] e^{-S[g,A]} \tag{358}$$

Correlation functions are obtained by functional differentiation, viz.

$$\begin{aligned} &\langle T_{i_1 j_1}(x_1) \cdots T_{i_n j_n}(x_n) \rangle \\ &= \frac{(-)^{n-1} 2^n}{\sqrt{g(x_1)} \cdots \sqrt{g(x_n)}} \frac{\delta^n}{\delta g^{i_1 j_1}(x_1) \cdots g^{i_n j_n}(x_n)} S_{eff}[g] \end{aligned} \tag{359}$$

with $g_{ij} \to \delta_{ij}$.

Consider two background metrics related by a Weyl transformation $g'_{ij}(x) = e^{2\sigma(x)} g_{ij}(x)$. Since the trace of T_{ij} vanishes in a CFT and $\langle T^i_i \rangle = -\delta S / \delta \sigma$, one might expect that $S_{eff}[g] = S_{eff}[g\prime]$. However, $S_{eff}[g]$ is divergent and must be regulated. This must be done even for a free theory (such as the pure $U(1)$ Maxwell theory). In a free theory the correlators of composite operators such as T_{ij} are well defined for separated points but must be regulated since they are too singular at short distance to have a well defined Fourier transform. Regularization introduces a scale and leads to the Weyl anomaly, which is expressed as

$$\langle T_i^i \rangle = \frac{c}{16\pi^2} W_{ijkl}^2 - \frac{a}{16\pi^2} \tilde{R}_{ijkl}^2 + \alpha \Box R + \beta R^2 \tag{360}$$

where the Weyl tensor and Euler densities are

$$\begin{aligned} W_{ijkl}^2 &= R_{ijkl}^2 - 2R_{ij}^2 + \tfrac{1}{3}R^2 \\ (\tfrac{1}{2}\epsilon_{ij}{}^{mn} R_{mnkl})^2 &= R_{ijkl}^2 - 4R_{ij}^2 + R^2 \end{aligned} \tag{361}$$

The anomaly must be local since it comes from ultraviolet divergences, and we have written all possible local terms of dimension 4 above. One can show that βR^2 violates the Wess-Zumino consistency condition while $\Box R$ is the variation of the local term $\int d^4x \sqrt{g} R^2$ in $S_{eff}[g]$. Finite local counter terms in an effective action depend on the regularization scheme and are usually considered not to carry dynamical information. (But see Ref. 219 for a proposed c-theorem based on this term. See Ref. 220 for a more extensive discussion of the Weyl anomaly.)

For the reasons above attention is usually restricted to the first two terms in (360). The scheme-independent coefficients c, a are central charges which characterize a CFT_4. One can show by a difficult argument[221,222] that c for the critical theories CFT_{UV}, CFT_{IR} agrees with the fixed point limits c_{UV}, c_{IR} of $c(M^2x^2)$ in (356). The central charge a is not measured in $\langle TT \rangle$ but agrees with constants a_{UV}, a_{IR} obtained in short and long distance limits of the 3-point function $\langle TTT \rangle$, see Refs. 169, 223.

What can be said about monotonicity? One might expect $c_{UV} > c_{IR}$, since the Weyl central charge is related to $\langle TT \rangle$ and thus closer to the notion of unitarity which was important in Zamolodchikov's proof. However this inequality fails in some field theory models. Cardy[224] conjectured that the inequality $a_{UV} > a_{IR}$ is the expression of the c-theorem in QFT_4. This is plausible since a is related to the topological Euler invariant in common with c for QFT_2, and Cardy showed that the inequality is satisfied in several models. Despite much effort (see Ref. 225 and references therein), there is no generally accepted proof of the c-theorem in QFT_4.

The values of c, a for free fields have been known for years. They were initially calculated by heat kernel methods, as described in Ref. 220. The free field values agree with c_{UV}, a_{UV} in any asymptotically free gauge theory, since the interactions vanish at short distance. For a theory of N_0 real scalars, $N_{\frac{1}{2}}$ Dirac fermions, and N_1 gauge bosons, the results are

$$\begin{aligned} c_{UV} &= \tfrac{1}{120}[N_0 + 5N_{\frac{1}{2}} + 12N_1] \\ a_{UV} &= \tfrac{1}{360}[N_0 + 11N_{\frac{1}{2}} + 62N_1] \end{aligned} \tag{362}$$

In a $SUSY$ gauge theory, component fields assemble into chiral multiplets (2 real scalars plus 1 Majorana (or Weyl) spinor) and vector multiplets (1 gauge boson plus 1 Majorana spinor). For a theory with N_χ chiral and N_V vector multiplets, the numbers above give

$$\begin{aligned} c_{UV} &= \tfrac{1}{24}[N_\chi + 3N_V] \\ a_{UV} &= \tfrac{1}{48}[N_\chi + 9N_V]. \end{aligned} \tag{363}$$

It is worthwhile to present some simple ways to calculate these central charges which are directly accessible to field theorists. Because of the relation to $\langle T_{ij}T_{kl}\rangle$ detailed above, the values of c_{UV} can be easily read from a suitably organized calculation of the free field 1-loop contributions of the various spins. For gauge bosons one must include the contribution of Faddeev-Popov ghosts.

Ex. 32:. Do this. Work directly in x-space at separated points. No integrals and no regularization is required. Organize the result in the form of the first term of (356).

In $SUSY$ gauge theories the stress tensor has a supersymmetric partner, the $U(1)_R$ current R_i. There are anomalies when the theory is coupled to g_{ij} and/or an external vector $V_i(x)$ with field strength V_{ij}. Including both sources one can write the combined anomalies as[222]

$$\begin{aligned} \langle T^i_i\rangle &= \tfrac{c}{16\pi^2}W^2_{ijkl} - \tfrac{a}{16\pi^2}\tilde{R}^2_{ijkl} + \tfrac{c}{6\pi^2}V^2_{ij} \\ \langle \partial_i\sqrt{g}R^i\rangle &= \tfrac{c-a}{24\pi^2}R_{ijkl}\tilde{R}^{ijkl} + \tfrac{5a-3c}{9\pi^2}V_{ij}\tilde{V}^{ij} \end{aligned} \tag{364}$$

Anomalies in the coupling of a gauge theory to external sources may be called external anomalies. There are also internal or gauge anomalies for both $\langle T^i_i\rangle$ and R_i. The gauge anomaly of R_i is described by an additional term in (364) proportional to $\beta(g)F_{ij}\tilde{F}^{ij}$, but this term vanishes in a CFT.

The formula (364) can be used to obtain c, a from 1-loop fermion triangle graphs for both the UV and IR critical theories. The triangle graph for $\langle R^iT_{jk}T_{lm}\rangle$ is linear in the $U(1)_R$ charges $r_{\hat{\alpha}}$ of the fermions in the theory, while the graph for $\langle R^iR_jR_k\rangle$ is cubic. We consider a general $\mathcal{N}=1$

theory with gauge group G and chiral multiplets in representations R_α of G. Comparing standard results for the anomalous divergences of triangle graphs with (364), one finds (see Refs. 222, 226),

$$c - a = -\tfrac{1}{16}(\mathrm{dim}G + \textstyle\sum_\alpha \mathrm{dim}R_\alpha(r_\alpha - 1)) \tag{365}$$

$$5a - 3c \; = \tfrac{9}{16}(\mathrm{dim}G + \textstyle\sum_\alpha \mathrm{dim}R_\alpha(r_\alpha - 1)^3 \tag{366}$$

We incorporate the facts that the $U(1)_R$ charge of the gaugino is $r_\lambda = 1$ while the charge of a fermion ψ_α in a chiral multiplet is related to the charge of the chiral superfield Φ^α by $r_{\hat{\alpha}} = r_\alpha - 1$.

If asymptotic freedom holds, then the CFT_{UV} is free, and one obtains its central charges c_{UV}, a_{UV} using the free field $U(1)_R$ charges, $r_\lambda = 1$ for the gaugino and $r_\alpha = \frac{2}{3}$ for chiral multiplets. The situation is more complex for the CFT_{IR} since the central charges are corrected by interactions. Seiberg and others following his techniques have found a large set of $SUSY$ gauge theories which do flow to critical points in the IR.[18] The $\mathcal{N}=1$ superconformal algebra $SU(2,2|1)$ contains a $U(1)_R$ current S_i which is in the same composite multiplet as the stress tensor. In many models this current is uniquely determined as a combination of the free current R_i plus terms which cancel the internal (gauge) anomalies of the former. Of course, the current S_i must also be conserved classically. Thus the S-charges of each Φ_α arrange so that all terms in the superpotential $U(\Phi_\alpha)$ have charge 2. It is the S-current which is used to show that anomalies match between Seiberg duals. These anomalies can be calculated from 1-loop graphs because the external anomalies are 1-loop exact for currents with no gauge anomaly. This is just the standard procedure of 't Hooft anomaly matching. The charge assigned by the S_i current is $r_\lambda = 1$ for gauginos and uniquely determined values r_α for chiral multiplets. It can be shown[227,222] that c_{IR}, a_{IR} are obtained by inserting these values in (365).

Given this theoretical background it is a matter of simple algebra to obtain the UV and IR central charges and subtract to deduce the following formulas for their change in an RG flow:

$$c_{UV} - c_{IR} = \tfrac{1}{384}\textstyle\sum_\alpha dimR_a(2 - 3r_\alpha)[(7 - 6r_\alpha)^2 - 17] \tag{367}$$

$$a_{UV} - a_{IR} \quad = \tfrac{1}{96}\textstyle\sum_\alpha dimR_\alpha(3r_\alpha - 2)^2(5 - 3r_\alpha) \tag{368}$$

These formulas were applied[226] to test the proposed c-theorem in the very many Seiberg models of $SUSY$ gauge theories with IR fixed points. Results indicated that the sign of $c_{UV} - c_{IR}$ is model-dependent, but $a_{UV} - a_{IR} > 0$ in **all** models. Thus there is a wealth of evidence that the Euler central charge satisfies a c-theorem, even though a fundamental proof is lacking.

Ex. 33:. Serious readers are urged to verify as many statements about the anomalies as they can. For minimal credit on this exercise please obtain the flow formulas (367) from (365).

Let us now apply some of these results to the field theories of most concern to us, namely the undeformed $\mathcal{N} = 4$ theory and its $\mathcal{N} = 1$ mass deformation. We can view the undeformed theory as the UV limit of the flow of its $\mathcal{N} = 1$ deformation. The free R-current assigns the charges $(1, -\frac{1}{3}, -\frac{1}{3}, -\frac{1}{3})$ to the gaugino and chiral matter fermions of the $\mathcal{N} = 1$ description, while the S-current of the mass deformed theory with superpotential in (345) assigns $(1, -\frac{1}{2}, -\frac{1}{2}, 0)$. In both cases these are elements of the Cartan subalgebra of $SU(4)_R$ and have vanishing trace. It is easy to see that the formula (365) for $c - a$ is proportional to this trace and vanishes. The same observation establishes that both currents have no gauge anomaly. The formula (366) then becomes

$$a = c = \frac{9}{32}(N^2 - 1)(1 + \sum (r_\alpha - 1)^3). \tag{369}$$

Applied to the free current and then the S-current, this gives

$$\begin{aligned} a_{UV} = c_{UV} &= \tfrac{1}{4}(N^2 - 1) \\ a_{IR} = c_{IR} &= \tfrac{27}{32}\tfrac{1}{4}(N^2 - 1). \end{aligned} \tag{370}$$

The relation $\Delta = \frac{3}{2}r$ between scale dimension and $U(1)_R$ charge also leads to the assignment of charges we have used. In the UV limit we have the $\mathcal{N} = 4$ theory with chiral superfields W_α, Φ^β with dimensions $\frac{3}{2}, 1$. In the IR limit we must consider the SU(2) invariant split $W_\alpha, \Phi^{1,2}, \Phi^3$, and the Leigh-Strassler argument for a conformal fixed point which requires $\Delta = \frac{3}{2}, \frac{3}{4}, \frac{3}{4}, \frac{3}{2}$. These values give the fermion charges used above. It is no accident that $r = 0$ for the fermion ψ^3. The Φ^3 multiplet drops out at long distance and thus cannot contribute to IR anomalies.

10.2. *Anomalies and the c-Theorem from AdS/CFT*

One of the early triumphs of the AdS/CFT was the calculation of the central charge c for $\mathcal{N} = 4$ SYM from the $\langle TT \rangle$ correlator whose absorptive part was obtained from the calculation of the cross-section for absorption of a graviton wave by the D3-brane geometry.[2] This was reviewed in Ref. 159 and we will take a different viewpoint here.

We will describe in some detail the general approach of Henningson and Skenderis[171] to the holographic Weyl anomaly. This leads to the correct values of the central charges and suggests a simple monotonic c-function.

We focus on the gravity part of the toy model action of Sec 9.1

$$S = \frac{-1}{16\pi G}[\int d^5z\sqrt{g}(R+\frac{12}{L^2}) + \int d^4z\sqrt{\gamma}2K] \tag{371}$$

in which we have added the Gibbons-Hawking surface term which we will explain further below. Lower spin bulk fields can be added and do not change the gravitational part of the conformal anomaly.

One solution of the Einstein equation is the AdS_{d+1} geometry which we previously wrote as

$$ds^2 = e^{\frac{2r}{L}}\delta_{ij}dx^i dx^j + dr^2 \tag{372}$$

We introduce the new radial coordinate $\rho = e^{-\frac{2r}{L}}$ in order to follow the treatment of Ref. 171. The boundary is now at $\rho = 0$.

Ex. 34:. Show that the transformed metric is

$$ds^2 = L^2[\frac{d\rho^2}{4\rho^2} + \frac{1}{\rho}\delta_{ij}dx^i dx^j] \tag{373}$$

This is just AdS_5 in new coordinates. We now consider more general solutions of the form

$$ds^2 = L^2[\frac{d\rho^2}{4\rho^2} + \frac{1}{\rho}g_{ij}(x,\rho)dx^i dx^j] \tag{374}$$

with non-trivial boundary data on the transverse metric, viz.

$$g_{ij}(x,\rho) \underset{\rho\to 0}{\longrightarrow} \bar{g}_{ij}(x) \tag{375}$$

The reason for this generalization may be seen by thinking of the form $\bar{g}_{ij}(x) = \delta_{ij} + h_{ij}(x)$. The first term describes the flat boundary on which the CFT_4 lives, while $h_{ij}(x)$ is the source of the stress tensor T_{ij}. We can use the formalism to compute $\langle T_{ij}\rangle$, $\langle T_{ij}T_{kl}\rangle$, etc.

Ex. 35:. Consider the special case of (374) in which $g_{ij}(x,\rho) = g_{ij}(x)$ depends only on the transverse x^i. Let R_{ijkl}, R_{ij} and R denote Riemann, Ricci and scalar curvatures of the $4d$ metric $g_{ij}(x)$. Show that the $5D$ metric thus defined satisfies the EOM $R_{\mu\nu} = -4g_{\mu\nu}$ if $R_{ij} = 0$. Show that the

$5D$ curvature invariant is

$$R_{\mu\nu\rho\sigma}R^{\mu\nu\rho\sigma} = \frac{\rho^2}{L^4}R_{ijkl}R^{ijkl} - \frac{4\rho}{L^2}R + \frac{40}{L^4} \tag{376}$$

Thus, as observed in Ref. 228, if $R_{ij} = 0$, we have a reasonably generic solution of the $5D$ EOM's with a curvature singularity on the horizon, $\rho \to \infty$.

As we will see shortly we will need to introduce a cutoff at $\rho = \epsilon$ and restrict the integration in (371) to the region $\rho \geq \epsilon$. The induced metric at the cutoff is $\gamma_{ij} = \frac{g_{ij}}{\rho}$. The measure $\sqrt{\gamma}$ appears in the surface term in (371) as does the trace of the second fundamental form

$$K = \gamma^{ij}K_{ij} = -g^{ij}\rho\frac{\partial}{\partial\rho}\left(\frac{g_{ij}(x,\rho)}{\epsilon}\right)\Big|_{\rho=\epsilon} \tag{377}$$

We now consider a particular type of infinitesimal 5D diffeomorphism first considered in this context in Ref. 229:

$$\begin{aligned} \rho &= \rho'(1 - 2\sigma(x')) \\ x^i &= x'^i + a^i(x', \rho') \end{aligned} \tag{378}$$

with

$$a^i(x,\rho) = \frac{L^2}{2}\int_0^\rho d\hat{\rho}g^{ij}(x,\hat{\rho})\partial_j\sigma(x) \tag{379}$$

Ex. 36:. Show that $g'_{55} = g_{55}$ and $g'_{5i} = g_{5i} = 0$ under this diffeomorphism, but that

$$g_{ij} \to g'_{ij} = g_{ij} + 2\sigma(1 - \rho\frac{\partial}{\partial_\rho})g_{ij} + \nabla_i a_j + \nabla_j a_i \tag{380}$$

In the boundary limit, $a_i \to 0$ and $\rho\frac{\partial}{\partial_\rho}g_{ij} \to 0$, so that

$$\bar{g}_{ij}(x) \to \bar{g}'_{ij}(x) = (1 + 2\sigma(x))\bar{g}_{ij}(x) \tag{381}$$

Hence the effect of the 5D diffeomorphism is a Weyl transformation of the boundary metric!

This raises a puzzle. Consider the on-shell action $S[\bar{g}_{ij}]$ obtained by substituting the solution (374) into (371). Since the bulk action and the field equations are invariant under diffeomorphisms, we would expect $S[\bar{g}_{ij}] = S[\bar{g}'_{ij}]$. But AdS/CFT requires that $S[\bar{g}_{ij}] = S_{eff}[\bar{g}_{ij}]$, and we know that, due to the Weyl anomaly, $S_{eff}[\bar{g}_{ij}] \neq S_{eff}[\bar{g}'_{ij}]$.

The resolution of the puzzle is that $S[\bar{g}_{ij}]$ as we defined it is meaningless since it diverges. This isn't the somewhat fuzzy-wuzzy divergence usually blamed on the functional integral for $S_{eff}[\bar{g}_{ij}]$ in quantum field theory. It is very concrete; when you insert a solution of Einstein's equation with the boundary behavior above into (371), the radial integral diverges near the boundary.

Therefore we define a cutoff action $S_\epsilon[\bar{g}_{ij}]$ as the on-shell value of (371) with radial integration restricted to $\rho \geq \epsilon$. One can study its dependence on the cutoff to obtain and subtract a counterterm action $S_\epsilon[\bar{g}_{ij}]_{ct}$ to cancel singular terms as $\epsilon \to 0$. $S_\epsilon[\bar{g}_{ij}]_{ct}$ is an integral over the hypersurface $\rho = \epsilon$ of a local function of the induced metric γ_{ij} and its curvatures, and it is not Weyl invariant. The renormalized action is defined as

$$S_{ren}[\bar{g}] \equiv \lim_{\epsilon \to 0}(S_\epsilon[\bar{g}] - S_e[\bar{g}]_{ct}) \tag{382}$$

We now outline how the calculation of correlation functions and the conformal anomaly proceeds in this formalism and then discuss further necessary details. The variation of S_{ren} is

$$\delta S_{ren}[\bar{g}] \equiv \frac{1}{2}\int d4x\sqrt{\bar{g}}\langle T_{ij}\rangle \delta\bar{g}^{ij}. \tag{383}$$

The variation defines the quantity $\langle T_{ij}(x)\rangle$ which, in the light of (358), is interpreted as the expectation value of the field theory stress tensor in the presence of the source $\bar{g}_{ij}$, and it depends non-locally on the source. Correlation functions in the CFT are then obtained by further differentiation, *e.g.*

$$\langle T_{ij}(x)T_{kl}(y)\rangle = -\frac{2}{\sqrt{\bar{g}(y)}}\frac{\delta}{\delta\bar{g}^{kl}(y)}\langle T_{ij}(x)\rangle\Big|_{\bar{g}_{ij}=\delta_{ij}} \tag{384}$$

The contributions to $\langle T_{ij}(x)\rangle$ come from the surface term in the radial integral in $S_\epsilon[\bar{g}]$ and from $S_\epsilon[\bar{g}]_{ct}$. Possible contributions involving bulk integrals vanish by the equations of motion.

The variation $\delta\bar{g}^{ij}$ is arbitrary; let's choose it to correspond to a Weyl transformation, *i.e.* $\delta\bar{g}^{ij} = -2\bar{g}^{ij}\delta\sigma$. Then (383) gives

$$\langle T^i_i\rangle = \bar{g}^{ij}\langle T_{ij}\rangle = -\frac{\delta S_{ren}[\bar{g}]}{\delta\sigma} \tag{385}$$

which is a standard result in quantum field theory in curved space. The quantity $\langle T^i_i\rangle$ is to be identified with the conformal anomaly of the CFT and must therefore be local. It is local, and the holographic computation gives (as we derive below)

$$\langle T^i_i\rangle = \frac{L^3}{8\pi G}(\frac{1}{8}R^{ij}R_{ij} - \frac{1}{24}R^2) \tag{386}$$

(The 2-point function (384) must be non-local, and it is. See Refs. 177, 178 for recent studies in the present formalism, and 210 for a closely related treatment.)

The holographic result may be compared with the field theory $\langle T_i^i \rangle$ in (360). The absence of the invariant R^2_{ijkl} in (386) requires $c = a$. Thus we deduce that any CFT$_4$ which has a holographic dual in this framework must have central charges which satisfy $c = a$ (at least as $N \to \infty$ when the classical supergravity approximation is valid.) This is satisfied by $\mathcal{N} = 4$ SYM but not by the conformal invariant $\mathcal{N} = 2$ theory with an $SU(N)$ gauge multiplet and $2N$ fundamental hypermultiplets.

Ex. 37:. Show that when $c = a$ the QFT trace anomaly of (360) reduces to

$$\langle T_i^i \rangle = \frac{c}{8\pi^2}(R^{ij}R_{ij} - \frac{1}{3}R^2) \tag{387}$$

Thus agreement with the holographic result (386) requires $c = \frac{\pi L^3}{8G}$. To check this recall that G is the 5D Newton constant, so that $G = \frac{G_{10}}{Vol S_5} = \frac{\pi L^3}{2N^2}$, where the last equality incorporates the requirement that AdS$_5 \times$ S^5 with 5-form flux N is a solution of the field equations of $D = 10$ Type IIB sugra. This gives the anomaly of undeformed $\mathcal{N} = 4$ SYM theory on the nose!

The Henningson-Skenderis method is very elegant and has useful generalizations.[172] It is worth discussing in more detail. The treatment starts with the mathematical result[230] that the general solution of the Einstein equations can be brought to the form (374), and that the transverse metric can be expanded in ρ near the boundary as

$$g_{ij}(x,\rho) = \bar{g}_{ij} + \rho g_{(2)ij} + \rho^2 g_{(4)ij} + \rho^2 \ln \rho h_{(4)ij} + \cdots \tag{388}$$

The tensor coefficients are functions of the transverse coordinates x^i. The tensors $g_{(2)ij}$, $h_{(4)ij}$ can be determined as local functions of the curvature $\bar{R}_{ijkl}$ of the boundary metric $\bar{g}_{ij}$. One just needs to substitute the expansion (388) in the 5D field equations $R_{\mu\nu} = -4g_{\mu\nu}$ and grind out a term-by-term solution.

Ex. 38:. Do this and derive $g_{(2)ij} = \frac{1}{2}(\bar{R}_{ij} - \frac{1}{6}\bar{R}\bar{g}_{ij})$. Very serious readers are encouraged to obtain the more complicated result for $h_{(4)ij}$ given in (A.6) of Ref. 172.

The tensor $g_{(4)ij}$ is only partially determined by this process of near-boundary analysis. Specifically its divergence and trace are local in the curvature $\bar{R}_{ijkl}$, but transverse traceless components are left undetermined. This is sensible since the EOM's are second order, and the single Dirichlet boundary condition does not uniquely fix the solution. At the linearized level the extra condition of regularity at large ρ (the deep interior) is imposed. The transverse traceless part of $g_{(4)ij}$ then depends non-locally on $\bar{g}_{ij}$ and eventually contributes to n-point correlators of T_{ij} in the dual field theory.

The local tensors in (388) are sufficient to determine the divergent part of $S_\epsilon[\bar{g}]$. It is tedious, delicate (but straightforward!) to substitute the expansion in (371), integrate near the boundary and identify the counterterms which cancel divergences. The result is

$$S_\epsilon[\bar{g}]_{ct} = \frac{1}{4\pi G}\int d^4x\sqrt{\gamma}\left(\frac{3}{2L^2} - \frac{\hat{R}}{8} - \frac{L^2\ln\epsilon}{32}(\hat{R}^{ij}\hat{R}_{ij} - \frac{1}{3}\hat{R}^2)\right) \tag{389}$$

where $\hat{R}_{ij}$ and $\hat{R}$ are the Ricci and scalar curvatures of the induced metric $\gamma_{ij} = \frac{g_{ij}(x,\epsilon)}{\epsilon}$. The first two terms in (389) thus have power singularities as $\epsilon \to 0$. Recall the discussion of cutoff dependence in Section 8.5. In the $\rho = z_0^2$ coordinate, the bulk cutoff ϵ should be identified with $1/\Lambda^2$ where Λ is the UV cutoff in QFT. Thus we find the quartic, quadratic, and logarithmic divergences expected in QFT_4! See Appendix B of Ref. 172 for details of the computation of (389).

The next step is to calculate

$$\langle T_i^i(x)\rangle = -\lim_{\epsilon\to 0}\frac{\delta}{\delta\sigma(x)}(S_\epsilon[\bar{g}] - S_\epsilon[\bar{g}]_{ct}). \tag{390}$$

However, one must vary the boundary data $\delta\bar{g}_{ij} = 2\delta\sigma\bar{g}_{ij}$ while maintaining the fact that the interior solution corresponds to that variation. Thus one is really carrying out the diffeomorphism of (378) so that $\delta\epsilon = 2\epsilon\delta\sigma(x)$. All terms of $S_\epsilon[\bar{g}]$ are invariant under the combined change of coordinates and change of shape of the cutoff hypersurface. The first two terms in $S_\epsilon[\bar{g}]_{ct}$ are also invariant. There is the explicit variation $\delta\ln\epsilon = -2\delta\sigma(x)$ in the logarithmic counterterm, and this is the only variation since the boundary integral is the difference of the Weyl^2 and Euler densities and is invariant. Thus we find the result (386) stated earlier in a strikingly simple way!

The method just described may be applied to the calculation of holographic conformal anomalies in any **even** dimension.[171,172] However for **odd** dimension the structure of the near-boundary expansion (388) changes. There is no $\ln\rho$ term and no logarithmic counterterm either. Hence no conformal anomaly in agreement with QFT in odd dimension.

10.3. *The Holographic c-Theorem*

The method just discussed can be extended to apply to the Weyl anomalies of the critical theories at end-points of holographic RG flows. In general we can consider a domain wall interpolating between the region of an AdS_{UV} with scale L_{UV} and the deep interior of an AdS_{IR} with scale L_{IR}. The holographic anomalies are

$$c_{UV} = \frac{\pi}{8G} L_{UV}^3 \quad c_{IR} = \frac{\pi}{8G} L_{IR}^3 \tag{391}$$

The first result can be derived by including relevant scalar fields in the previous method, and latter by applying the method to an entire AdS geometry with scale L_{IR}.

For any bulk domain wall one can consider the following scale-dependent function (and its radial derivative):

$$\begin{aligned} C(r) &= \tfrac{\pi}{8G} \tfrac{1}{A'^3} \\ C'(r) &= \tfrac{\pi}{8G} \tfrac{-3A''}{A'^4} \end{aligned} \tag{392}$$

We have $C'(r) \geq 0$ as a consequence of the condition $A'' \leq 0$ derived from the domain wall EOM's in Section 9.2. Thus $C(r)$ is an essentially perfect holographic c-function:

(1) It decreases monotonically along the flow from $UV \to IR$.
(2) It interpolates between the central charges c_{UV} and c_{IR}.
(3) If perfect, it would be stationary only if conformal symmetry holds. This is true if the domain wall is the solution of the first order flow equations discussed in Sec 9.3 and thus true for SUSY flows.

The moral of the story is that the c-theorem for RG flows, which has resisted proof by field theory methods, is trivial when the theory has a gravity dual since $A'' \leq 0$. See Refs. 175, 186.

Finally, we note that for the mass deformed $\mathcal{N} = 4$ theory the ratio $(\frac{L_{IR}}{L_{UV}})^3 = (\frac{W_{UV}}{W_{IR}})^3 = \frac{27}{32}$. Thus the holographic prediction of $c_{IR} = a_{IR}$ agrees with the field theory result in (370)! See Ref. 231.

There is much more to be said about the active subject of holographic RG flows and many interesting papers that deserve study by interested theorists. We hope that the introduction to the basic ideas contained in these lecture notes will stimulate that study.

Acknowledgements

We are grateful to Massimo Bianchi, Umut Gursoy, Krzysztof Pilch and Kostas Skenderis for many helpful comments on the manuscript. We would also like to thank the organizers of TASI 2001, Steven Gubser, Joe Lykken and the general director K.T. Mahanthappa for their invitation to this stimulating Summer School. Finally, these lectures were based in part on material presented by E.D. at Saclay, Sète and Ecole Normale Superieure, and by D.Z.F. during the Semestre Supercordes at the Institut Henri Poincaré and at the Les Houches summer school. We wish to thank Costas Bachas, Eugène Cremmer, Michael Douglas, Ivan Kostov, André Neveu, Kelley Stelle, Aliocha Zamolodchikov and Jean-Bernard Zuber for invitations to present those lectures. Research supported in part by the National Science Foundation under grants PHY-98-19686 and PHY-00-96515.

References

1. J. Maldacena, "The Large N Limit of Superconformal Field Theories and Supergravity", Adv. Theor. Math. Phys. **2** (1998) 231-252, [hep-th/9711200].
2. S. Gubser, I. Klebanov and A. Polyakov, "Gauge Theory Correlators from Non-Critical String Theory", Phys. Lett. **B428** (1998) 105-114, [hep-th/9802109].
3. E. Witten, "Anti De Sitter Space And Holography," Adv. Theor. Math. Phys. **2** (1998) 253-291, [hep-th/9802150].
4. I. R. Klebanov, Nucl. Phys. B **496**, 231 (1997) [arXiv:hep-th/9702076].
5. S. S. Gubser, I. R. Klebanov and A. A. Tseytlin, Nucl. Phys. B **499**, 217 (1997) [arXiv:hep-th/9703040].
6. S. S. Gubser and I. R. Klebanov, Phys. Lett. B **413**, 41 (1997) [arXiv:hep-th/9708005].
7. O. Aharony, S. S. Gubser, J. Maldacena, H. Ooguri and Y. Oz, "Large N field theories, string theory and gravity," Phys. Rept. **323**, 183 (2000) [arXiv:hep-th/9905111].
8. G. 't Hooft, "A Planar diagram theory for strong interactions," Nucl. Phys. B **72**, 461 (1974).
9. G. 't Hooft, "Dimensional Reduction in Quantum Gravity" gr-qc/9310026 ; L. Susskind, "The Worlds as a Hologram", J. Math. Phys. **36** (1995) 6377, hep-th/9409089.
10. D. Bigatti and L. Susskind, "TASI lectures on the holographic principle," arXiv:hep-th/0002044.
11. J. Wess and J. Bagger, *Supersymmetry and Supergravity*, Princeton Univ. Press 1983.
12. S. Weinberg, *The Theory of Quantized Fields*, Vol III, Supersymmetry, Cambridge University Press, 2000

13. S.J. Gates, M.T. Grisaru, M. Rocek and W. Siegel, *Superspace or One Thousand and One Lessons in Supersymmetry*, Benjamin/Cummings Publ. Company, INC (1983), hep-th/0108200; P. West, *Introduction to Supersymmetry and Supergravity*, World Scientific 1990; J. Lopuszanski, *Symmetry and Supersymmetry in Quantum Field Theory*, World Scientific 1991.
14. M.K. Prasad and C.M. Sommerfield, Phys. Rev. Lett. **35** (1975) 760; E.B. Bogomolnyi, Sov. J. Nucl. Phys. **24** (1976) 449.
15. A. Galperin, E. Ivanov, S. Kalitzin, V. Ogievetsky and E. Sokatchev, "Unconstrained N=2 matter, Yang-Mills and Supergravity theories in harmonic superspace", Class. Quant. Grav. **1** (1984) 469; G.G. Hartwell, P.S. Howe, "(N,p,q) harmonic superspace", Int. J. Mod. Phys. **A10** (1995) 3901; P.S. Howe, "A superspace survey", Class. Quant. Grav. **12** (1995) 1823.
16. E. D'Hoker and D. H. Phong, "Lectures on supersymmetric Yang-Mills theory and integrable systems," in *Theoretical Physics at the end of the XX-th Century*, Proceedings of the CRM Summer School, June 27 – July 10, 1999, Banff, Canada; Y. Saint-Aubin and L. Vinet editors, Springer Verlag (2000); hep-th/9912271.
17. M.A. Shifman and A.I. Vainshtein, Nucl. Phys. **B277** (1986) 456; Nucl. Phys. **B359** (1991) 571; D. Amati, K. Konishi, Y. Meurice, G.C. Rossi and G. Veneziano, Phys. Rep. **162** 169; M. Grisaru, W. Siegel, M. Rocek, "Improved Methods for Supergraphs", Nucl. Phys. **B159** (1979) 429.
18. N. Seiberg, "Naturalness versus supersymmetric nonrenormalization theorems", Phys. Lett. **B318** (1993) 469; Phys. Rev. **D49** (1994) 6857; K.A. Intriligator and N. Seiberg, "Lectures on Supersymmetric Gauge Theories and Electric-Magnetic Duality", Nucl. Phys. Proc. Suppl. **45BC** (1996) 1, hep-th/9509066.
19. K.G. Wilson and J. Kogut, "The renormalization group and the epsilon expansion", Phys. Rep. **12** (1974) 75.
20. R. Grimm, M. Sohnius and J. Wess, "Extended supersymmetry and gauge theories", Nucl. Phys. **B133** (1978) 275.
21. C. Montonen and D. Olive, Phys. Lett. **B72** (1977) 117; P. Goddard, J. Nuyts and D. Olive, Nucl. Phys. **B125** (1977) 1; E. Witten and D. Olive. Phys. Lett. **B78** (1978) 97.
22. S. Minwalla, "Restrictions Imposed by Superconformal Invariance on Quantum Field Theories" Adv. Theor. Math. Phys. **2** (1998) 781; J. Rasmussen, "Comments on N = 4 superconformal algebras," Nucl. Phys. B **593**, 634 (2001) [arXiv:hep-th/0003035].
23. N. Seiberg, "Notes on theories with 16 supercharges," Nucl. Phys. Proc. Suppl. **67**, 158 (1998) hep-th/9705117.
24. V. K. Dobrev and V. B. Petkova, "On The Group Theoretical Approach To Extended Conformal Supersymmetry: Classification Of Multiplets," Lett. Math. Phys. **9**, 287 (1985); S. Ferrara and E. Sokatchev, "Short representations of $SU(2,2|N)$ and harmonic superspace analyticity," hep-th/9912168; S. Ferrara, "Superspace representations of $SU(2,2|N)$ superalgebras and multiplet shortening," arXiv:hep-th/0002141.

25. L. Andrianopoli and S. Ferrara, "Short and long SU(2,2/4) multiplets in the AdS/CFT correspondence," Lett. Math. Phys. **48**, 145 (1999) hep-th/9812067; L. Andrianopoli, S. Ferrara, E. Sokatchev and B. Zupnik, "Shortening of primary operators in N-extended SCFT(4) and harmonic-superspace analyticity," Adv. Theor. Math. Phys. **3**, 1149 (1999) [arXiv:hep-th/9912007]; S. Ferrara and A. Zaffaroni, "Superconformal field theories, multiplet shortening, and the AdS(5)/SCFT(4) correspondence," arXiv:hep-th/9908163.
26. S. Ferrara and E. Sokatchev, "Superconformal interpretation of BPS states in AdS geometries," hep-th/0005151; S. Ferrara and E. Sokatchev, "Conformal primaries of OSp(8/4,R) and BPS states in AdS(4)," hep-th/0003051; S. Ferrara and E. Sokatchev, "Representations of (1,0) and (2,0) superconformal algebras in six dimensions: Massless and short superfields," hep-th/0001178; P. J. Heslop and P. S. Howe, "A note on composite operators in N = 4 SYM," Phys. Lett. B **516**, 367 (2001) [arXiv:hep-th/0106238]; N. Maggiore and A. Tanzini, "Protected operators in N = 2,4 supersymmetric theories," Nucl. Phys. B **613**, 34 (2001) [arXiv:hep-th/0105005].
27. A. V. Ryzhov, "Quarter BPS operators in N = 4 SYM," JHEP **0111**, 046 (2001) [arXiv:hep-th/0109064].
28. J. Polchinski, *String Theory*, Vol II, Cambridge University Press, 1998.
29. F. Gliozzi, J. Scherk and D. I. Olive, "Supersymmetry, Supergravity Theories And The Dual Spinor Model," Nucl. Phys. B **122**, 253 (1977).
30. W. Nahm, "Supersymmetries And Their Representations," Nucl. Phys. B **135**, 149 (1978).
31. E. Cremmer, B. Julia and J. Scherk, "Supergravity Theory In 11 Dimensions," Phys. Lett. B **76**, 409 (1978).
32. M.J. Duff, ed, *The World in Eleven Dimensions, Supergravity, Supermembranes and M-theory*, Institute of Physics Publishing, Bristol and Philadelphia (1999).
33. P. S. Howe and P. C. West, "The Complete N=2, D = 10 Supergravity," Nucl. Phys. B **238**, 181 (1984); J. H. Schwarz and P. C. West, "Symmetries And Transformations Of Chiral N=2 D = 10 Supergravity," Phys. Lett. B **126**, 301 (1983).
34. M.B. Green, J.H. Schwarz and E. Witten, *String Theory*, Vol II, Cambridge University Press 1987.
35. E. D'Hoker, "String Theory", Lectures delivered at the Institute for Advanced Study Princeton (1997), available on the IAS Mathematics homepage www.ias.math web-site; in *Quantum Fields and Strings : A Course for Mathematicians*, P. Deligne, P. Etingof, D.S. Freed, L.C. Jeffrey, D. Kazhdan, J.W. Morgan, D.R. Morrison and E. Witten editors, American Mathematical Society, Institute for Advanced Study (1999), Vol II, p 807 – 1012.
36. E. D'Hoker and D.H. Phong, "The geometry of string perturbation theory", Review of Modern Physics, Vol **60** (1988) 917.
37. E. D'Hoker and D.H. Phong, "Superholomorphic Anomalies and Supermoduli Space", Nucl. Phys. **B292** (1987) 317; E. D'Hoker and D.H. Phong, "Chiral Superstring Amplitudes and the GSO projection", Mod. Phys. Lett. **A4** (1989) 1335; E. D'Hoker and D.H. Phong, "Conformal Scalar Fields and

Chiral Splitting on Super-Riemann Surfaces", Commun. Math. Phys. **125** (1989) 469; E. D'Hoker and D.H. Phong, "Two-Loop Superstrings I, Main Formulas", hep-th/0110247.
38. D. H. Friedan, "Nonlinear Models In Two + Epsilon Dimensions," Annals Phys. **163**, 318 (1985).
39. C. G. Callan, E. J. Martinec, M. J. Perry and D. Friedan, "Strings In Background Fields," Nucl. Phys. B **262**, 593 (1985).
40. J. Dai, R. G. Leigh and J. Polchinski, "New Connections Between String Theories," Mod. Phys. Lett. A **4**, 2073 (1989); J. Polchinski, "Dirichlet-Branes and Ramond-Ramond Charges," Phys. Rev. Lett. **75**, 4724 (1995) [arXiv:hep-th/9510017].
41. C. G. Callan, J. A. Harvey and A. Strominger, "Worldbrane actions for string solitons," Nucl. Phys. B **367**, 60 (1991); C. G. Callan, J. A. Harvey and A. Strominger, "Supersymmetric string solitons," arXiv:hep-th/9112030.
42. G. T. Horowitz and A. Strominger, "Black Strings And P-Branes," Nucl. Phys. B **360**, 197 (1991).
43. J. Polchinski, S. Chaudhuri and C. V. Johnson, "Notes on D-Branes," arXiv:hep-th/9602052; C. V. Johnson, "D-brane primer," arXiv:hep-th/0007170.
44. M. J. Duff, "TASI lectures on branes, black holes and anti-de Sitter space," arXiv:hep-th/9912164; A. W. Peet, arXiv:hep-th/0008241.
45. E. Verlinde, "Global Aspects of Electric-Magnetic Dulaity", Nucl. Phys. **B455** (1995) 211, hep-th/9506011.
46. E. Witten, "Solutions of Four-Dimensional Field Theories via M-theory", Nucl. Phys. **B500** (1997) 3.
47. L. Andrianopoli and S. Ferrara, "K-K excitations on AdS(5) x S(5) as N = 4 *primary* superfields," Phys. Lett. B **430**, 248 (1998) [arXiv:hep-th/9803171].
48. G. 't Hooft, "Magnetic monopoles in unified gauge theories", Nucl. Phys. **B79** (1974) 276; A.M. Polyakov, "Particle Spectrum in the quantum field theory", JETP Letters **20** (1974) 194.
49. O. Aharony, M. Berkooz and E. Silverstein, "Multiple-trace operators and non-local string theories," JHEP **0108**, 006 (2001) [arXiv:hep-th/0105309].
50. M. Gunaydin and N. Marcus, "The Spectrum Of The S**5 Compactification Of The Chiral N=2, D = 10 Supergravity And The Unitary Supermultiplets Of U(2, 2/4)," Class. Quant. Grav. **2**, L11 (1985); H. J. Kim, L. J. Romans and P. van Nieuwenhuizen, "The Mass Spectrum Of Chiral N=2 D = 10 Supergravity On S**5," Phys. Rev. D **32**, 389 (1985).
51. B. Biran, A. Casher, F. Englert, M. Rooman and P. Spindel, "The Fluctuating Seven Sphere In Eleven-Dimensional Supergravity," Phys. Lett. **B134**, 179 (1984); A. Casher, F. Englert, H. Nicolai and M. Rooman, "The Mass Spectrum Of Supergravity On The Round Seven Sphere," Nucl. Phys. **B243**, 173 (1984).
52. M. Gunaydin and N. P. Warner, "Unitary Supermultiplets Of $Osp(8/4, R)$ And The Spectrum Of The S^7 Compactification Of Eleven-Dimensional Supergravity," Nucl. Phys. **B272**, 99 (1986).

53. B. de Wit and H. Nicolai, "On The Relation Between D = 4 And D = 11 Supergravity," Nucl. Phys. **B243**, 91 (1984); B. de Wit, H. Nicolai and N. P. Warner, "The Embedding Of Gauged N=8 Supergravity Into D = 11 Supergravity," Nucl. Phys. **B255**, 29 (1985); B. de Wit and H. Nicolai, "D = 11 Supergravity With Local SU(8) Invariance," Nucl. Phys. **B274**, 363 (1986); B. de Wit and H. Nicolai, "The Consistency Of The S^7 Truncation In D = 11 Supergravity," Nucl. Phys. **B281**, 211 (1987).
54. K. Pilch, P. van Nieuwenhuizen and P. K. Townsend, "Compactification Of D = 11 Supergravity On S^4 (Or 11 = 7 + 4, Too)," Nucl. Phys. **B242**, 377 (1984).
55. M. Gunaydin, P. van Nieuwenhuizen and N. P. Warner, "General Construction Of The Unitary Representations Of Anti-De Sitter Superalgebras And The Spectrum Of The S^4 Compactification Of Eleven-Dimensional Supergravity," Nucl. Phys. **B255**, 63 (1985); P. van Nieuwenhuizen, "The Complete Mass Spectrum Of D = 11 Supergravity Compactified On S^4 And A General Mass Formula For Arbitrary Cosets M(4)," Class. Quant. Grav. **2**, 1 (1985).
56. O. Aharony, Y. Oz and Z. Yin, "M-theory on AdS(p) x S(11-p) and superconformal field theories," Phys. Lett. **B430**, 87 (1998) [hep-th/9803051]; S. Minwalla, "Particles on AdS(4/7) and primary operators on M(2/5) brane worldvolumes," JHEP **9810**, 002 (1998) [hep-th/9803053]; R. G. Leigh and M. Rozali, "The large N limit of the (2,0) superconformal field theory," Phys. Lett. **B431**, 311 (1998) [hep-th/9803068]; E. Halyo, "Supergravity on AdS(4/7) x S(7/4) and M branes," JHEP **9804**, 011 (1998) [hep-th/9803077]; F. Bastianelli and R. Zucchini, "Bosonic quadratic actions for 11D supergravity on AdS(7/4) x S(4/7)," Class. Quant. Grav. **16**, 3673 (1999) [hep-th/9903161]; S. Ferrara and E. Sokatchev, "Representations of superconformal algebras in the AdS(7/4)/CFT(6/3) correspondence," J. Math. Phys. **42**, 3015 (2001) [arXiv:hep-th/0010117]; B. Eden, S. Ferrara and E. Sokatchev, "(2,0) superconformal OPEs in D = 6, selection rules and non-renormalization theorems," JHEP **0111**, 020 (2001) [arXiv:hep-th/0107084].
57. R. Corrado, B. Florea and R. McNees, "Correlation functions of operators and Wilson surfaces in the d = 6, (0,2) theory in the large N limit," Phys. Rev. **D60**, 085011 (1999) [hep-th/9902153]; F. Bastianelli and R. Zucchini, "Three point functions of chiral primary operators in d = 3, N = 8 and d = 6, N = (2,0) SCFT at large N," Phys. Lett. **B467**, 61 (1999) hep-th/9907047.
58. E. D'Hoker and B. Pioline, "Near-Extremal Correlators and Generalized Consistent Truncation for $AdS_{4|7} \times S^{7|4}$", JHEP **0007** (2000) 021, hep-th/0006103.
59. P. S. Howe, G. Sierra and P. K. Townsend, "Supersymmetry In Six-Dimensions," Nucl. Phys. **B221**, 331 (1983); P. Claus, R. Kallosh and A. Van Proeyen, "M 5-brane and superconformal (0,2) tensor multiplet in 6 dimensions," Nucl. Phys. **B518**, 117 (1998) [hep-th/9711161]; C. Grojean and J. Mourad, "Superconformal 6D (2,0) theories in superspace," Class. Quant. Grav. **15**, 3397 (1998) hep-th/9807055; G. Arutyunov and E. Sokatchev, "Implications of Superconformal Symmetry for Interacting (2,0) Tensor Multiplets," hep-th/0201145.

60. A. M. Polyakov, "Gauge fields and space-time," arXiv:hep-th/0110196.
61. B. de Wit and I. Herger, "Anti-de Sitter supersymmetry," Lect. Notes Phys. **541**, 79 (2000) [arXiv:hep-th/9908005].
62. M. Spradlin, A. Strominger and A. Volovich, "Les Houches lectures on de Sitter space," arXiv:hep-th/0110007.
63. M. Flato and C. Fronsdal, "Interacting singletons," Lett. Math. Phys. **44**, 249 (1998) [arXiv:hep-th/9803013].
64. M. Gunaydin and D. Minic, Nucl. Phys. B **523**, 145 (1998) [arXiv:hep-th/9802047]; M. Gunaydin, D. Minic and M. Zagermann, "4D doubleton conformal theories, CPT and II B string on AdS(5) x S(5)," Nucl. Phys. B **534**, 96 (1998) [Erratum-ibid. B **538**, 531 (1998)] [arXiv:hep-th/9806042].
65. E. Angelopoulos, M. Flato, C. Fronsdal and D. Sternheimer, "Massless particles, conformal group and De Sitter universe", Phys. Rev. **D23** (1981) 1278; M. Flato, C. Fronsdal and D. Sternheimer, "Singleton physics," Proceedings of Steklov Mathematical Institute **226** (1999) 185, arXiv:hep-th/9901043; M. Flato, C. Fronsdal and D. Sternheimer, "Singletons, physics in AdS universe and oscillations of composite neutrinos", Lett. Math. Phys. **48** (1999) 109; C. Fronsdal, "Singletons and neutrinos," Lett. Math. Phys. **52**, 51 (2000) [arXiv:hep-th/9911241].
66. I. I. Kogan, "Singletons and logarithmic CFT in AdS/CFT correspondence," Phys. Lett. B **458**, 66 (1999) [arXiv:hep-th/9903162].
67. A. Bilal and C. S. Chu, "D3 brane(s) in AdS(5) x S(5) and N = 4,2,1 SYM," Nucl. Phys. B **547**, 179 (1999) [arXiv:hep-th/9810195].
68. Y. Imamura, "Supersymmetries and BPS configurations on Anti-de Sitter space," Nucl. Phys. B **537**, 184 (1999) [arXiv:hep-th/9807179].
69. E. Shuster, "Killing spinors and supersymmetry on AdS," Nucl. Phys. B **554**, 198 (1999) [arXiv:hep-th/9902129].
70. V. Balasubramanian and P. Kraus, "Spacetime and the holographic renormalization group," Phys. Rev. Lett. **83**, 3605 (1999) [arXiv:hep-th/9903190]; V. Balasubramanian and P. Kraus, "A stress tensor for anti-de Sitter gravity," Commun. Math. Phys. **208**, 413 (1999) [arXiv:hep-th/9902121].
71. G. E. Arutyunov and S. A. Frolov, "On the origin of supergravity boundary terms in the AdS/CFT correspondence," Nucl. Phys. B **544**, 576 (1999) [arXiv:hep-th/9806216].
72. W. Muck and K. S. Viswanathan, "Regular and irregular boundary conditions in the AdS/CFT correspondence," Phys. Rev. D **60**, 081901 (1999) [arXiv:hep-th/9906155].
73. W. Muck and K. S. Viswanathan, "A regularization scheme for the AdS/CFT correspondence," JHEP **9907**, 018 (1999) [arXiv:hep-th/9904039].
74. I. Bena, "On the construction of local fields in the bulk of AdS(5) and other spaces," Phys. Rev. D **62**, 066007 (2000) [arXiv:hep-th/9905186].
75. H. Boschi-Filho and N. R. Braga, "Quantum fields in anti de Sitter spacetime and degrees of freedom in the bulk / boundary correspondence," Phys. Lett. B **505**, 263 (2001) [arXiv:hep-th/0009039].
76. G. Arutyunov and S. Frolov, "On the correspondence between gravity fields and CFT operators," JHEP **0004**, 017 (2000) [hep-th/0003038].

77. P. Heslop and P. S. Howe, "On harmonic superspaces and superconformal fields in four dimensions," Class. Quant. Grav. **17**, 3743 (2000) [arXiv:hep-th/0005135]; H. Ooguri, J. Rahmfeld, H. Robins and J. Tannenhauser, "Holography in superspace," JHEP **0007**, 045 (2000) [arXiv:hep-th/0007104]; P. Heslop and P. S. Howe, "Chiral superfields in IIB supergravity," Phys. Lett. B **502**, 259 (2001) [arXiv:hep-th/0008047]; P. S. Howe and P. C. West, "AdS/SCFT in superspace," Class. Quant. Grav. **18**, 3143 (2001) [arXiv:hep-th/0105218].
78. E. D'Hoker and D. Z. Freedman, "Gauge boson exchange in AdS(d+1)," Nucl. Phys. **B544** (1999) 612, hep-th/9809179.
79. E. D'Hoker, D. Z. Freedman, S. D. Mathur, A. Matusis and L. Rastelli, "Graviton and gauge boson propagators in AdS(d+1)," Nucl. Phys. **B562** (1999) 330, hep-th/9902042.
80. M. Henningson and K. Sfetsos, "Spinors and the AdS/CFT correspondence," Phys. Lett. B **431**, 63 (1998) [arXiv:hep-th/9803251]; G. E. Arutyunov and S. A. Frolov, "Antisymmetric tensor field on AdS(5)," Phys. Lett. B **441**, 173 (1998) [arXiv:hep-th/9807046]; U. H. Danielsson, E. Keski-Vakkuri and M. Kruczenski, "Vacua, propagators, and holographic probes in AdS/CFT," JHEP **9901**, 002 (1999) [arXiv:hep-th/9812007]; A. Polishchuk, "Massive symmetric tensor field on AdS," JHEP **9907**, 007 (1999) [arXiv:hep-th/9905048]; T. Kawano and K. Okuyama, "Spinor exchange in AdS(d+1)," Nucl. Phys. B **565**, 427 (2000) [arXiv:hep-th/9905130].
81. I. Bena, "The antisymmetric tensor propagator in AdS," Phys. Rev. D **62**, 127901 (2000) [arXiv:hep-th/9910059]; I. Bena, "The propagator for a general form field in AdS(d+1)," Phys. Rev. D **62**, 126008 (2000) [arXiv:hep-th/9911073]; A. Naqvi, "Propagators for massive symmetric tensor and p-forms in AdS(d+1)," JHEP **9912**, 025 (1999) [arXiv:hep-th/9911182]; W. Muck, "Spinor parallel propagator and Green's function in maximally symmetric spaces," J. Phys. A **33**, 3021 (2000) [arXiv:hep-th/9912059]; I. Bena, H. Nastase and D. Vaman, "Propagators for p-forms in AdS(2p+1) and correlation functions in the AdS(7)/(2,0) CFT correspondence," Phys. Rev. D **64**, 106009 (2001) [arXiv:hep-th/0008239].
82. M. M. Caldarelli, "Quantum scalar fields on anti-de Sitter spacetime," Nucl. Phys. B **549**, 499 (1999) [arXiv:hep-th/9809144].
83. Z. Chang, C. B. Guan and H. Y. Guo, "AdS dynamics for massive scalar field: Exact solutions vs. bulk boundary propagator," arXiv:hep-th/0104251.
84. C. Fronsdal, Phys. Rev **D10** (1974) 589; C.P. Burgess and C.A. Lutken, "Propagators and effective potentials in anti-de Sitter space", Nucl. Phys. **B272** (1986) 661; T. Inami and H. Ooguri, "One-loop effective potential in anti-de Sitter space", Prog. Theor. Phys. **73** (1985) 1051; C.J. Burges, D.Z. Freedman, S. Davis and G.W. Gibbons, "Supersymmetry in anti-de Sitter space", Ann. Phys. **167** (1986) 285.
85. E. D'Hoker and R. Jackiw, "Space Translation Breaking and Compactification in the Liouville Theory", Phys. Rev. Lett. **50** (1983) 1719; E. D'Hoker D.Z. Freedman and R. Jackiw, "SO(2,1) Invariant Quantization of the Liouville Theory", Phys. Rev. **D28** (1983) 2583.

86. A. Strominger, "AdS(2) quantum gravity and string theory," JHEP **9901**, 007 (1999) [arXiv:hep-th/9809027]; J. Michelson and A. Strominger, "The geometry of (super)conformal quantum mechanics," Commun. Math. Phys. **213**, 1 (2000) [arXiv:hep-th/9907191].
87. E. D'Hoker, "Classical and Quantal Supersymmetric Liouville Theory", Phys. Rev. **D28** (1983) 1346.
88. B. Allen and T. Jacobson, "Vector two-point functions in maximally symmetric spaces", Commun. Math. Phys. **103** (1986) 669.
89. G. Mack and A. Salam, "Finite Component Field Representation of the Conformal Group", Ann. Phys. (NY) **53** (1969) 174; D. Bohm, M. Flato, D. Sternheimer and J.P. Vigier, "Conformal group symmetry of elementary particles", Nuovo Cimento **38** (1966) 1941; M. Flato, J. Simon and D. Sternheimer, "Conformal covariance of field equations", Ann. Phys. (NY) **61** (1970) 78.
90. D. Z. Freedman, S. D. Mathur, A. Matusis and L. Rastelli, "Correlation functions in the CFT(d)/AdS($d+1$) correspondence," Nucl. Phys. B **546**, 96 (1999) [arXiv:hep-th/9804058]; H. Liu and A. A. Tseytlin, "D = 4 super Yang-Mills, D = 5 gauged supergravity, and D = 4 conformal supergravity," Nucl. Phys. B **533**, 88 (1998) [arXiv:hep-th/9804083].
91. S. M. Lee, S. Minwalla, M. Rangamani and N. Seiberg, "Three-point functions of chiral operators in D = 4, N = 4 SYM at large N," Adv. Theor. Math. Phys. **2**, 697 (1998) [arXiv:hep-th/9806074].
92. E. D'Hoker, D. Z. Freedman and W. Skiba, "Field theory tests for correlators in the AdS/CFT correspondence," Phys. Rev. **D59**, 045008 (1999) [hep-th/9807098].
93. W. Skiba, "Correlators of short multi-trace operators in N = 4 supersymmetric Yang-Mills," Phys. Rev. **D60**, 105038 (1999) [hep-th/9907088]; F. Gonzalez-Rey, B. Kulik and I. Y. Park, "Non-renormalization of two point and three point correlators of N = 4 SYM in N = 1 superspace," Phys. Lett. **B455**, 164 (1999) [hep-th/9903094].
94. S. Penati, A. Santambrogio and D. Zanon, "Two-point functions of chiral operators in N = 4 SYM at order g^4," JHEP **9912**, 006 (1999) [hep-th/9910197]; S. Penati, A. Santambrogio and D. Zanon, "More on correlators and contact terms in N = 4 SYM at order g^4," Nucl. Phys. B **593**, 651 (2001) [arXiv:hep-th/0005223].
95. K. Intriligator, "Bonus symmetries of N = 4 super-Yang-Mills correlation functions via AdS duality," Nucl. Phys. **B551**, 575 (1999) [hep-th/9811047]; K. Intriligator and W. Skiba, "Bonus symmetry and the operator product expansion of N = 4 super-Yang-Mills," Nucl. Phys. **B559**, 165 (1999) [hep-th/9905020].
96. P. Howe and P. West, "Superconformal invariants and extended supersymmetry", Phys. Lett. **B400** (1997) 307, hep-th/9611075; P. S. Howe, E. Sokatchev and P. C. West, "3-point functions in N = 4 Yang-Mills," Phys. Lett. **B444**, 341 (1998) [hep-th/9808162]; B. Eden, P. S. Howe and P. C. West, "Nilpotent invariants in N = 4 SYM," Phys. Lett. **B463**, 19 (1999) [hep-th/9905085];

P. S. Howe, C. Schubert, E. Sokatchev and P. C. West, "Explicit construction of nilpotent covariants in N = 4 SYM," Nucl. Phys. **B571**, 71 (2000) [hep-th/9910011].

97. E. D'Hoker, "A New Invariant of Superconformal $OSp(n|1)$", J. Math. Phys. **32** (1991) 832.
98. P. J. Heslop and P. S. Howe, "OPEs and 3-point correlators of protected operators in N = 4 SYM," arXiv:hep-th/0107212; E. D'Hoker and A.V. Ryzhov, "Three Point Functions of Quarter BPS Operators in N=4 SYM", hep-th/0109065; L. Hoffmann, L. Mesref, A. Meziane and W. Ruhl, "Multi-trace quasi-primary fields of N = 4 SYM(4) from AdS n-point functions," arXiv:hep-th/0112191.
99. W. Muck and K. S. Viswanathan, "Conformal field theory correlators from classical scalar field theory on AdS(d+1)," Phys. Rev. D **58**, 041901 (1998) [arXiv:hep-th/9804035].
100. G. Chalmers, H. Nastase, K. Schalm and R. Siebelink, "R-current correlators in N = 4 super Yang-Mills theory from anti-de Sitter supergravity," Nucl. Phys. B **540**, 247 (1999) [arXiv:hep-th/9805105].
101. W. Muck and K. S. Viswanathan, "Conformal field theory correlators from classical field theory on anti-de Sitter space. II: Vector and spinor fields," Phys. Rev. D **58**, 106006 (1998) [arXiv:hep-th/9805145].
102. F. Bastianelli and R. Zucchini, "Three point functions for a class of chiral operators in maximally supersymmetric CFT at large N," Nucl. Phys. B **574**, 107 (2000) [arXiv:hep-th/9909179].
103. F. Bastianelli and R. Zucchini, "3-point functions of universal scalars in maximal SCFTs at large N," JHEP **0005**, 047 (2000) [arXiv:hep-th/0003230].
104. P. Mansfield and D. Nolland, "Order 1/N**2 test of the Maldacena conjecture: Cancellation of the one-loop Weyl anomaly," Phys. Lett. B **495**, 435 (2000) [arXiv:hep-th/0005224].
105. E. D'Hoker, D. Z. Freedman, S. D. Mathur, A. Matusis and L. Rastelli, "Extremal correlators in the AdS/CFT correspondence," in the Yuri Golfand Memorial Volume, *Many Faces of the Superworld*, M. Shifman, Editor, World Scientific (2000), (*Invited Contribution*), hep-th/9908160.
106. J. Erdmenger and M. Perez-Victoria, "Non-renormalization of next-to-extremal correlators in N = 4 SYM and the AdS/CFT correspondence," hep-th/9912250
107. B. Eden, P. S. Howe, C. Schubert, E. Sokatchev and P. C. West, "Extremal correlators in four-dimensional SCFT," Phys. Lett. **B472**, 323 (2000) [hep-th/9910150].
108. B. U. Eden, P. S. Howe, E. Sokatchev and P. C. West, "Extremal and next-to-extremal n-point correlators in four-dimensional SCFT," Phys. Lett. B **494**, 141 (2000) [arXiv:hep-th/0004102].
109. M. Bianchi and S. Kovacs, "Non-renormalization of extremal correlators in N = 4 SYM theory," Phys. Lett. **B468**, 102 (1999) [hep-th/9910016].
110. E. D'Hoker, J. Erdmenger, D. Z. Freedman and M. Perez-Victoria, "Near-extremal correlators and vanishing supergravity couplings in AdS/CFT," Nucl. Phys. **B589** (2000) 3, hep-th/0003218.

111. G. Arutyunov and S. Frolov, "Scalar quartic couplings in type IIB supergravity on $AdS_5 \times S^5$," hep-th/9912210; G. Arutyunov and S. Frolov, "Scalar quartic effective action on AdS_5," hep-th/0002152.
112. M. Gunaydin, L. J. Romans and N. P. Warner, "Gauged N=8 Supergravity In Five-Dimensions," Phys. Lett. **B154**, 268 (1985).
113. M. Pernici, K. Pilch and P. van Nieuwenhuizen, "Gauged N=8 D = 5 Supergravity," Nucl. Phys. **B259**, 460 (1985).
114. H. Nastase and D. Vaman, "The AdS-CFT correspondence, consistent truncations and gauge invariance," arXiv:hep-th/0004123.
115. S. Fubini, A.J. Hanson and R. Jackiw, "New Approach to Field Theory", Phys. Rev **D7** (1973) 1932.
116. E. D'Hoker, S. D. Mathur, A. Matusis and L. Rastelli, "The operator product expansion of $\mathcal{N} = 4$ SYM and the 4-point functions of supergravity", Nucl. Phys. **B589** (2000) 38, hep-th/9911222.
117. L. Hoffmann, T. Leonhardt, L. Mesref and W. Ruhl, "Composite fields, generalized hypergeometric functions and the U(1)Y symmetry in the AdS/CFT correspondence," arXiv:hep-th/0102162.
118. G. Arutyunov, B. Eden and E. Sokatchev, "On non-renormalization and OPE in superconformal field theories," Nucl. Phys. B **619**, 359 (2001) [arXiv:hep-th/0105254].
119. G. Arutyunov, B. Eden, A. C. Petkou and E. Sokatchev, "Exceptional non-renormalization properties and OPE analysis of chiral four-point functions in N = 4 SYM(4)," Nucl. Phys. B **620**, 380 (2002) [arXiv:hep-th/0103230].
120. D.J. Gross, "Applications of the Renormalization Group to High Energy Physics", *Methods in Field Theory*, Session XXVIII, Les Houches 1975, R. Balian and J. Zinn-Justin eds, North Holland/World Scientific 1981
121. E. Witten, "Short Distance Analysis of Weak Interactions", Nucl. Phys. **122** (1977) 109.
122. S. Ferrara, R. Gatto, A.F. Grillo and G. Parisi, Nucl. Phys. **B49** (1972) 77; Nuovo Cimento **26** (1975) 226.
123. J. H. Park, "Superconformal symmetry and correlation functions," Nucl. Phys. B **559**, 455 (1999) [arXiv:hep-th/9903230].
124. L. Hoffmann, L. Mesref and W. Ruhl, "Conformal partial wave analysis of AdS amplitudes for dilaton axion four-point functions," Nucl. Phys. B **608**, 177 (2001) [arXiv:hep-th/0012153]; L. Hoffmann, A. C. Petkou and W. Ruhl, "Aspects of the conformal operator product expansion in AdS/CFT correspondence," arXiv:hep-th/0002154.
125. F. A. Dolan and H. Osborn, "Conformal four point functions and the operator product expansion," Nucl. Phys. B **599**, 459 (2001) [arXiv:hep-th/0011040].
126. S. Ferrara and E. Sokatchev, "Universal properties of superconformal OPEs for 1/2 BPS operators in $3 \leq D \leq 6$," arXiv:hep-th/0110174.
127. B. Eden and E. Sokatchev, "On the OPE of 1/2 BPS short operators in N = 4 SCFT(4)," Nucl. Phys. B **618**, 259 (2001) [arXiv:hep-th/0106249].
128. G. Arutyunov, S. Frolov and A. Petkou, "Perturbative and instanton corrections to the OPE of CPOs in N = 4 SYM(4)," Nucl. Phys. B **602**, 238 (2001) [Erratum-ibid. B **609**, 540 (2001)] [arXiv:hep-th/0010137].

129. M. Bianchi, S. Kovacs, G. Rossi and Y. S. Stanev, "Properties of the Konishi multiplet in N = 4 SYM theory," JHEP **0105**, 042 (2001) [arXiv:hep-th/0104016].
130. H. Liu and A. A. Tseytlin, "On four-point functions in the CFT/AdS correspondence," Phys. Rev. D **59**, 086002 (1999) [arXiv:hep-th/9807097].
131. H. Liu and A. A. Tseytlin, "Dilaton-fixed scalar correlators and AdS(5) x S(5) - SYM correspondence," JHEP **9910**, 003 (1999) [arXiv:hep-th/9906151].
132. G. Arutyunov, S. Frolov and A. C. Petkou, "Operator product expansion of the lowest weight CPOs in N = 4 SYM(4) at strong coupling," Nucl. Phys. B **586**, 547 (2000) [Erratum-ibid. B **609**, 539 (2000)] [arXiv:hep-th/0005182].
133. E. D'Hoker and D. Z. Freedman, "General scalar exchange in AdS(d+1)," Nucl. Phys. **B550**, 261 (1999) hep-th/9811257.
134. E. D'Hoker, D. Z. Freedman, S. D. Mathur, A. Matusis and L. Rastelli, "Graviton exchange and complete 4-point functions in the AdS/CFT correspondence," Nucl. Phys. **B562** (1999) 353, hep-th/9903196.
135. H. Liu, "Scattering in anti-de Sitter space and operator product expansion," Phys. Rev. D **60**, 106005 (1999) [arXiv:hep-th/9811152].
136. E. D'Hoker, D. Z. Freedman and L. Rastelli, "AdS/CFT 4-point functions: How to succeed at z-integrals without really trying," Nucl. Phys. **B562** (1999) 395, hep-th/9905049.
137. M. Bianchi, S. Kovacs, G. Rossi and Y. S. Stanev, "On the logarithmic behavior in N = 4 SYM theory," JHEP **9908**, 020 (1999) [arXiv:hep-th/9906188].
138. S. Sanjay, "On direct and crossed channel asymptotics of four-point functions in AdS/CFT correspondence," Mod. Phys. Lett. A **14**, 1413 (1999) [arXiv:hep-th/9906099].
139. B. Eden, P. S. Howe, C. Schubert, E. Sokatchev and P. C. West, "Simplifications of four-point functions in N = 4 supersymmetric Yang-Mills theory at two loops," Phys. Lett. B **466**, 20 (1999) [arXiv:hep-th/9906051]; B. U. Eden, P. S. Howe, A. Pickering, E. Sokatchev and P. C. West, "Four-point functions in N = 2 superconformal field theories," Nucl. Phys. B **581**, 523 (2000) [arXiv:hep-th/0001138]; B. Eden, C. Schubert and E. Sokatchev, "Four-point functions of chiral primary operators in N = 4 SYM," arXiv:hep-th/0010005; S. Penati and A. Santambrogio, "Superspace approach to anomalous dimensions in N = 4 SYM," Nucl. Phys. B **614**, 367 (2001) [arXiv:hep-th/0107071].
140. M. Bianchi, S. Kovacs, G. Rossi and Y. S. Stanev, "Anomalous dimensions in N = 4 SYM theory at order g**4," Nucl. Phys. B **584**, 216 (2000) [arXiv:hep-th/0003203]; B. Eden, A. C. Petkou, C. Schubert and E. Sokatchev, "Partial non-renormalisation of the stress-tensor four-point function in N = 4 SYM and AdS/CFT," Nucl. Phys. B **607**, 191 (2001) [arXiv:hep-th/0009106]; B. Eden, C. Schubert and E. Sokatchev, "Three-loop four-point correlator in N = 4 SYM," Phys. Lett. B **482**, 309 (2000) [arXiv:hep-th/0003096]; L. Hoffmann, L. Mesref and W. Ruhl, "AdS box graphs, unitarity and operator product expansions," Nucl. Phys. B **589**, 337 (2000) [arXiv:hep-th/0006165].

141. T. Banks and M. B. Green, "Non-perturbative effects in AdS(5) x S**5 string theory and d = 4 SUSY Yang-Mills," JHEP **9805**, 002 (1998) [arXiv:hep-th/9804170].
142. J. H. Brodie and M. Gutperle, "String corrections to four point functions in the AdS/CFT correspondence," Phys. Lett. B **445**, 296 (1999) [arXiv:hep-th/9809067].
143. S. M. Lee, "AdS(5)/CFT(4) four-point functions of chiral primary operators: Cubic vertices," Nucl. Phys. B **563**, 349 (1999) [arXiv:hep-th/9907108].
144. C. P. Herzog, "OPEs and 4-point functions in AdS/CFT correspondence," JHEP **0102**, 038 (2001) [arXiv:hep-th/0002039].
145. 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," Nucl. Phys. B **554**, 215 (1999) [arXiv:hep-th/9810051].
146. G. Arutyunov and S. Frolov, "Four-point functions of lowest weight CPOs in N = 4 SYM(4) in supergravity approximation," Phys. Rev. D **62**, 064016 (2000) [arXiv:hep-th/0002170].
147. W. S. l'Yi, "Correlators of currents corresponding to the massive p-form fields in AdS/CFT correspondence," Phys. Lett. B **448**, 218 (1999) [arXiv:hep-th/9811097].
148. H. Osborn and G. M. Shore, "Correlation functions of the energy momentum tensor on spaces of constant curvature," Nucl. Phys. B **571**, 287 (2000) [arXiv:hep-th/9909043].
149. J. Kalkkinen and D. Martelli, "Holographic renormalization group with fermions and form fields," Nucl. Phys. B **596**, 415 (2001) [arXiv:hep-th/0007234].
150. M. Bertola, J. Bros, U. Moschella and R. Schaeffer, "AdS/CFT correspondence for n-point functions," arXiv:hep-th/9908140.
151. M. A. Vasiliev, "Cubic interactions of bosonic higher spin gauge fields in AdS(5)," Nucl. Phys. B **616**, 106 (2001) [arXiv:hep-th/0106200]; M. A. Vasiliev, "Conformal higher spin symmetries of 4D massless supermultiplets and osp(L,2M) invariant equations in generalized (super)space," arXiv:hep-th/0106149; O. V. Shaynkman and M. A. Vasiliev, "Higher spin conformal symmetry for matter fields in 2+1 dimensions," Theor. Math. Phys. **128**, 1155 (2001) [Teor. Mat. Fiz. **128**, 378 (2001)] [arXiv:hep-th/0103208]; S. E. Konstein, M. A. Vasiliev and V. N. Zaikin, "Conformal higher spin currents in any dimension and AdS/CFT correspondence," JHEP **0012**, 018 (2000) [arXiv:hep-th/0010239]; P. Haggi-Mani and B. Sundborg, "Free large N supersymmetric Yang-Mills theory as a string theory," JHEP **0004**, 031 (2000) [arXiv:hep-th/0002189]; E. Sezgin and P. Sundell, "Doubletons and 5D higher spin gauge theory," JHEP **0109**, 036 (2001) [arXiv:hep-th/0105001]; E. Sezgin and P. Sundell, "Towards massless higher spin extension of D = 5, N = 8 gauged supergravity," JHEP **0109**, 025 (2001) [arXiv:hep-th/0107186]; R. R. Metsaev, "Light cone form of field dynamics in anti-de Sitter spacetime and AdS/CFT correspondence," Nucl. Phys. B **563**, 295 (1999) [arXiv:hep-th/9906217]; R. R. Metsaev, "Massless arbitrary spin fields in AdS(5)," arXiv:hep-th/0201226.

152. M. Bianchi, M. B. Green, S. Kovacs and G. Rossi, "Instantons in supersymmetric Yang-Mills and D-instantons in IIB superstring theory," JHEP **9808**, 013 (1998) [arXiv:hep-th/9807033].
153. N. Dorey, T. J. Hollowood, V. V. Khoze, M. P. Mattis and S. Vandoren, "Multi-instanton calculus and the AdS/CFT correspondence in N = 4 superconformal field theory," Nucl. Phys. B **552**, 88 (1999) [arXiv:hep-th/9901128].
154. R. Gopakumar and M. B. Green, "Instantons and non-renormalisation in AdS/CFT," JHEP **9912**, 015 (1999) [arXiv:hep-th/9908020].
155. M. Bianchi, "(Non-)perturbative tests of the AdS/CFT correspondence," Nucl. Phys. Proc. Suppl. **102**, 56 (2001) [arXiv:hep-th/0103112].
156. N. Dorey, "Instantons, compactification and S-duality in N = 4 SUSY Yang-Mills theory. I," JHEP **0104**, 008 (2001) [arXiv:hep-th/0010115].
157. G. Bonelli, L. Bonora, S. Terna and A. Tomasiello, "Instantons and scattering in N = 4 SYM in 4D," arXiv:hep-th/9912227.
158. G. Chalmers and J. Erdmenger, "Dual expansions of N = 4 super Yang-Mills theory via IIB superstring theory," Nucl. Phys. B **585**, 517 (2000) [arXiv:hep-th/0005192].
159. I. R. Klebanov, "Introduction to the AdS/CFT correspondence," arXiv:hep-th/0009139.
160. J. Maldacena, "Wilson loops in large N field theories," Phys. Rev. Lett. **80**, 4859 (1998) [arXiv:hep-th/9803002].
161. J. K. Erickson, G. W. Semenoff and K. Zarembo, "Wilson loops in N = 4 supersymmetric Yang-Mills theory," Nucl. Phys. B **582**, 155 (2000) [arXiv:hep-th/0003055].
162. N. Drukker and D. J. Gross, "An exact prediction of N = 4 SUSYM theory for string theory," J. Math. Phys. **42**, 2896 (2001) [arXiv:hep-th/0010274].
163. D. Anselmi, M. T. Grisaru and A. Johansen, "A Critical Behaviour of Anomalous Currents, Electric-Magnetic Universality and CFT_4," Nucl. Phys. B **491**, 221 (1997) [arXiv:hep-th/9601023].
164. Coleman, Sidney "ASPECTS OF SYMMETRY: selected Erice lectures of Sidney Coleman," Cambridge Univ. Press, 1985. 402p.
165. W. Heidenreich, "All Linear Unitary Irreducible Representations Of De Sitter Supersymmetry With Positive Energy," Phys. Lett. B **110**, 461 (1982).
166. P. Breitenlohner and D. Z. Freedman, "Stability In Gauged Extended Supergravity," Annals Phys. **144**, 249 (1982).
167. S. Ferrara, C. Fronsdal and A. Zaffaroni, "On N = 8 supergravity on AdS(5) and N = 4 superconformal Yang-Mills theory," Nucl. Phys. B **532**, 153 (1998) [arXiv:hep-th/9802203].
168. Ya.S. Stanev, *Bulg. J. Phys.* **15** (1988) 93.
169. H. Osborn and A. Petkou, *Implications of Conformal Invariance for Quantum Field Theories in* $d > 2$ Ann. Phys. (N.Y.) 231 (1994) 311, hep-th/9307010,

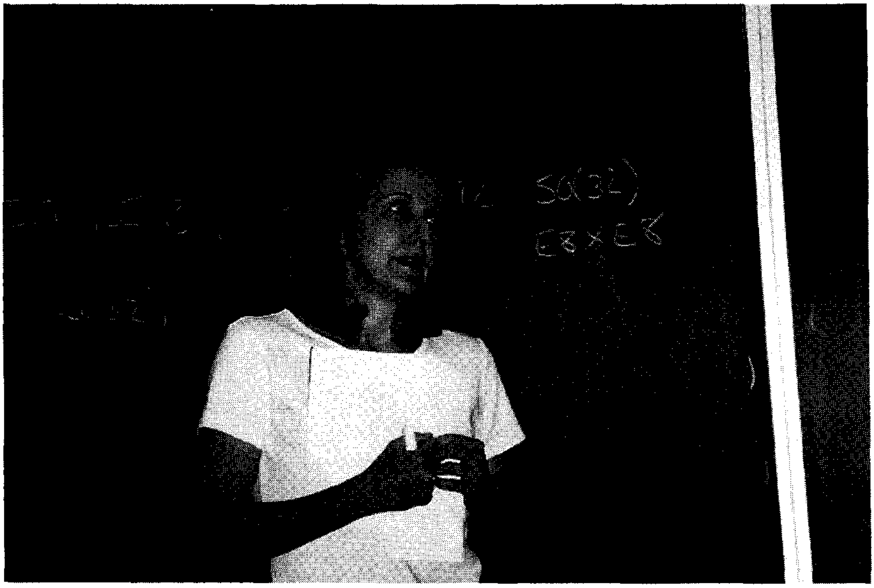
J. Erdmenger and H. Osborn, *Conserved Currents and the Energy Momentum Tensor in Conformally Invariant Theories for General Dimensions*, Nucl. Phys. B483 (1997) 431, hep-th/9605009.

170. I. R. Klebanov and E. Witten, "AdS/CFT correspondence and symmetry breaking," Nucl. Phys. B **556**, 89 (1999) [arXiv:hep-th/9905104].
171. M. Henningson and K. Skenderis, "The holographic Weyl anomaly," JHEP **9807**, 023 (1998) [arXiv:hep-th/9806087]; "Holography and the Weyl anomaly," Fortsch. Phys. **48**, 125 (2000) [arXiv:hep-th/9812032].
172. S. de Haro, S. N. Solodukhin and K. Skenderis, "Holographic reconstruction of spacetime and renormalization in the AdS/CFT correspondence," Commun. Math. Phys. **217**, 595 (2001) [arXiv:hep-th/0002230]
173. I. S. Gradshteyn and I. M. Ryzhik, Table of Integrals, Series, and Products Academic Press, New York (1965).
174. D. Z. Freedman, K. Johnson and J. I. Latorre, "Differential regularization and renormalization: A New method of calculation in quantum field theory," Nucl. Phys. B **371**, 353 (1992).
175. L. Girardello, M. Petrini, M. Porrati and A. Zaffaroni, "Novel local CFT and exact results on perturbations of N = 4 super Yang-Mills from AdS dynamics," JHEP **9812**, 022 (1998) [arXiv:hep-th/9810126].
176. J. Distler and F. Zamora, "Non-supersymmetric conformal field theories from stable anti-de Sitter spaces," Adv. Theor. Math. Phys. **2**, 1405 (1999) [arXiv:hep-th/9810206].
177. M. Bianchi, D. Z. Freedman and K. Skenderis, "How to go with an RG flow," JHEP **0108**, 041 (2001) [arXiv:hep-th/0105276].
178. M. Bianchi, D. Z. Freedman and K. Skenderis, "Holographic renormalization," arXiv:hep-th/0112119.
179. P. Breitenlohner and D. Z. Freedman, "Positive Energy In Anti-De Sitter Backgrounds And Gauged Extended Supergravity," Phys. Lett. B **115**, 197 (1982).
180. L. Mezincescu and P. K. Townsend, "Stability At A Local Maximum In Higher Dimensional Anti-De Sitter Space And Applications To Supergravity," Annals Phys. **160**, 406 (1985).
181. V. Balasubramanian, P. Kraus and A. E. Lawrence, "Bulk vs. boundary dynamics in anti-de Sitter spacetime," Phys. Rev. D **59**, 046003 (1999) [arXiv:hep-th/9805171].
182. V. Balasubramanian, P. Kraus, A. E. Lawrence and S. P. Trivedi, "Holographic probes of anti-de Sitter space-times," Phys. Rev. D **59**, 104021 (1999) [arXiv:hep-th/9808017].
183. P. Kraus, F. Larsen and S. P. Trivedi, "The Coulomb branch of gauge theory from rotating branes," JHEP **9903**, 003 (1999) [arXiv:hep-th/9811120].
184. A. B. Zamolodchikov, "'Irreversibility' Of The Flux Of The Renormalization Group In JETP Lett. **43**, 730 (1986) [Pisma Zh. Eksp. Teor. Fiz. **43**, 565 (1986)].
185. P. K. Townsend, "Positive Energy And The Scalar Potential In Higher Dimensional (Super)Gravity Theories," Phys. Lett. B **148**, 55 (1984).
186. D. Freedman, S. Gubser, K. Pilch and N. Warner, "Renormalization Group Flows from Holography-Supersymmetry and a c-Theorem", Adv. Theor. Math. Phys. **3** (1999) 363, [hep-th/9904017].

187. K. Skenderis and P. K. Townsend, "Gravitational stability and renormalization-group flow," Phys. Lett. B **468**, 46 (1999) [arXiv:hep-th/9909070].
188. O. DeWolfe, D. Z. Freedman, S. S. Gubser and A. Karch, "Modeling the fifth dimension with scalars and gravity," Phys. Rev. D **62**, 046008 (2000) [arXiv:hep-th/9909134].
189. M. Cvetic and H. Soleng, "Supergravity Domain Walls," Phys. Rept. **282** (1997) 159, [hep-th/9604090].
190. J. de Boer, E. Verlinde and H. Verlinde, "On the holographic renormalization group," JHEP **0008**, 003 (2000) [arXiv:hep-th/9912012].
191. D. Martelli and A. Miemiec, "CFT/CFT interpolating RG flows and the holographic c-function," arXiv:hep-th/0112150.
192. M. Gunaydin, L. Romans, and N. Warner, "Compact and Non-Compact Gauged Supergravity Theories in Five Dimensions," Nucl. Phys. **B272** (1986) 598.
193. K. Pilch and N. Warner, "A New Supersymmetric Compactification of Chiral IIB Supergravity." Phys.Lett. B487 (2000) 22-29 [hep-th/0002192]
194. K. Pilch, N. P. Warner, "N=2 Supersymmetric RG Flows and the IIB Dilaton," Nucl.Phys. B594 (2001) 209-228 [hep-th/0004063]
195. A. Khavaev and N. P. Warner, "An N = 1 supersymmetric Coulomb flow in IIB supergravity," Phys. Lett. B **522**, 181 (2001) [arXiv:hep-th/0106032].
196. B. de Wit (Utrecht U.), H. Nicolai C. V. Johnson, K. J. Lovis, D. C. Page "The Kahler Structure of Supersymmetric Holographic RG Flows," JHEP 0110 (2001) 014 [hep-th/0107261].
197. B. de Wit and H. Nicolai, "The Consistency of the S**7 Truncation in D = 11 Supergravity", Nucl.Phys. **B281**(1987)211.
198. H. Nastase, D. Vaman and P. van Nieuwenhuizen, "Consistency of the AdS(7) x S(4) reduction and the origin of self-duality in odd dimensions," Nucl. Phys. B **581** (2000) 179 [hep-th/9911238].
199. J. Polchinski and M. J. Strassler, "The string dual of a confining four-dimensional gauge theory," arXiv:hep-th/0003136.
200. I. R. Klebanov and M. J. Strassler, "Supergravity and a confining gauge theory: Duality cascades and chiSB-resolution of naked singularities," JHEP **0008**, 052 (2000) [arXiv:hep-th/0007191].
201. N. P. Warner, "Some Properties Of The Scalar Potential In Gauged Supergravity Theories," Nucl. Phys. B **231**, 250 (1984), "Some New Extrema Of The Scalar Potential Of Gauged N=8 Supergravity," Phys. Lett. B **128**, 169 (1983).
202. A. Khavaev, K. Pilch and N. P. Warner, "New vacua of gauged N = 8 supergravity in five dimensions," Phys. Lett. B **487**, 14 (2000) [arXiv:hep-th/9812035].
203. K. Pilch (private communication).
204. D.Z. Freedman, S.S. Gubser, K. Pilch and N.P. Warner, "Continuous distributions of D3-branes and gauged supergravity", JHEP **0007** (2000) 038, [hep-th/9906194].

205. A. Brandhuber and K. Sfetsos, "Wilson loops from multicentre and rotating branes, mass gaps and phase structure in gauge theories," Adv. Theor. Math. Phys. **3** (1999) 851 [hep-th/9906201].
206. L. Girardello, M. Petrini, M. Porrati and A. Zaffaroni, "The Supergravity Dual of $N = 1$ Super Yang-Mills Theory", Nucl. Phys. **B569** (2000) 451-469, [hep-th/9909047].
207. G. Arutyunov, S. Frolov and S. Theisen, "A note on gravity-scalar fluctuations in holographic RG flow geometries," Phys. Lett. B **484** (2000) 295 [arXiv:hep-th/0003116].
208. M. Bianchi, O. DeWolfe, D.Z. Freedman and K. Pilch, "Anatomy of two holographic renormalization group flows," JHEP **0101** (2001) 021 [hep-th/0009156].
209. A. Brandhuber and K. Sfetsos, "Current correlators in the Coulomb branch of N = 4 SYM," JHEP **0012** (2000) 014 [hep-th/0010048].
210. W. Mück, "Correlation functions in holographic renormalization group flows," hep-th/0105270.
211. C. Vafa and E. Witten, Nucl. Phys. B **431**, 3 (1994) [arXiv:hep-th/9408074].
212. R. Donagi and E. Witten, Nucl. Phys. B **460**, 299 (1996) [arXiv:hep-th/9510101].
213. R. G. Leigh and M. J. Strassler, "Exactly marginal operators and duality in four-dimensional N=1 supersymmetric gauge theory," Nucl. Phys. B **447**, 95 (1995) [arXiv:hep-th/9503121].
214. M. Henningson and K. Sfetsos, "Spinors and the AdS/CFT correspondence," Phys. Lett. B **431**, 63 (1998) [arXiv:hep-th/9803251].
215. A. Volovich, "Rarita-Schwinger field in the AdS/CFT correspondence," JHEP **9809**, 022 (1998) [arXiv:hep-th/9809009].
216. A. S. Koshelev and O. A. Rytchkov, "Note on the massive Rarita-Schwinger field in the AdS/CFT correspondence," Phys. Lett. B **450**, 368 (1999) [arXiv:hep-th/9812238].
217. A. Polishchuk, "Massive symmetric tensor field on AdS", JHEP **9907**, 007 (1999) [arXiv:hep-th/9905048]
218. D. Anselmi, "Central functions and their physical implications," JHEP **9805**, 005 (1998) [arXiv:hep-th/9702056].
219. D. Anselmi, "A universal flow invariant in quantum field theory," Class. Quant. Grav. **18**, 4417 (2001) [arXiv:hep-th/0101088].
220. M. J. Duff, "Twenty years of the Weyl anomaly," Class. Quant. Grav. **11**, 1387 (1994) [arXiv:hep-th/9308075].
221. A. Cappelli, D. Friedan and J. I. Latorre, "C Theorem And Spectral Representation," Nucl. Phys. B **352**, 616 (1991).
222. D. Anselmi, D. Z. Freedman, M. T. Grisaru and A. A. Johansen, "Nonperturbative formulas for central functions of supersymmetric gauge theories," Nucl. Phys. B **526**, 543 (1998) [arXiv:hep-th/9708042].
223. A. Cappelli, R. Guida and N. Magnoli, "Exact consequences of the trace anomaly in four dimensions," Nucl. Phys. B **618**, 371 (2001) [arXiv:hep-th/0103237].
224. J. L. Cardy, "Is There A C Theorem In Four-Dimensions?," Phys. Lett. B **215**, 749 (1988).

225. S. Forte and J. I. Latorre, "A proof of the irreversibility of renormalization group flows in four dimensions," Nucl. Phys. B **535**, 709 (1998) [arXiv:hep-th/9805015].
226. D. Anselmi, J. Erlich, D. Z. Freedman and A. A. Johansen, "Positivity constraints on anomalies in supersymmetric gauge theories," Phys. Rev. D **57**, 7570 (1998) [arXiv:hep-th/9711035].
227. I. I. Kogan, M. A. Shifman and A. I. Vainshtein, "Matching conditions and duality in N=1 SUSY gauge theories in the conformal window," Phys. Rev. D **53**, 4526 (1996) [Erratum-ibid. D **59**, 109903 (1996)] [arXiv:hep-th/9507170].
228. M. Cvetic, H. Lu and C. N. Pope, "Brane-world Kaluza-Klein reductions and branes on the brane," J. Math. Phys. **42**, 3048 (2001) [arXiv:hep-th/0009183].
229. C. Imbimbo, A. Schwimmer, S. Theisen and S. Yankielowicz, "Diffeomorphisms and Holographic Anomalies", Class. Quant. Grav. **17** (2000) 1129-1138, [hep-th/9910267].
230. C. Fefferman and C. Robin Graham, 'Conformal Invariants', in *Elie Cartan et les Mathématiques d'aujourd'hui* (Astérisque, 1985) 95.
231. A. Karch, D. Lust and A. Miemiec, "New N = 1 superconformal field theories and their supergravity description," Phys. Lett. B **454**, 265 (1999) [arXiv:hep-th/9901041].
232. J. Maldacena and H. Ooguri, "Strings in AdS(3) and SL(2,R) WZW model. I," J. Math. Phys. **42**, 2929 (2001) [arXiv:hep-th/0001053]; J. Maldacena, H. Ooguri and J. Son, "Strings in AdS(3) and the SL(2,R) WZW model. II: Euclidean black hole," J. Math. Phys. **42**, 2961 (2001) [arXiv:hep-th/0005183]; J. Mldacena and H. Ooguri, "Strings in AdS(3) and the SL(2,R) WZW model. III: Correlation functions," arXiv:hep-th/0111180.



LOUISE DOLAN

This page intentionally left blank

PERTURBATIVE STRING THEORY AND RAMOND-RAMOND FLUX

LOUISE DOLAN
Department of Physics, University of North Carolina,
Chapel Hill, North Carolina 27599-3255
E-mail: dolan@physics.unc.edu

1. Introduction

Recently, the promise of solving strong coupling Yang-Mills theory by considering the Type IIB superstring on anti de Sitter space (AdS) via the AdS/CFT correspondence has led to a renewed interest in perturbative string theory and its formulation on background curved spaces. The first lecture reviews the derivation of the physical spectrum and scattering amplitudes in the old covariant quantization for open and closed bosonic string theory, with attention given to the structures that will require modification when the background spacetime is curved. The second lecture reviews various worldsheet formulations for the superstring, including the Berkovits-Vafa-Witten variables which provide a manifestly supersymmetric and covariant quantization in six dimensions. In the third lecture, these worldsheet fields are used to solve the string constraints on the vertex operators for the Type IIB superstring on $AdS_3 \times S^3 \times K3$ with background Ramond flux. A short section on computing correlation functions using these fields has been added in 4.3.

2. Old Covariant Quantization

We review the traditional quantization, Refs. 1-6, of the open and closed bosonic strings and point to the steps that need generalization to accommodate strings with background Ramond fields.

2.1. *Open Bosonic String*

We introduce the Fubini-Veneziano fields:

$$X^\mu(z) = q^\mu - ip^\mu \ln z + i\sum_{n\neq 0} \frac{a_n^\mu}{n} z^{-n} \tag{1}$$

where

$$[a_m^\mu, a_n^\nu] = \eta^{\mu\nu} m\delta_{m,-n}; \quad [q^\mu, p^\nu] = i\eta^{\mu\nu}; \quad [q^\mu, a_n^\nu] = 0, n \neq 0; \quad n \in \mathbf{Z}\,. \tag{2}$$

These commutators are the Lorentz covariant quantization conditions. Here $p^\mu \equiv a_0^\mu$. The fields $X^\mu(z)$ are restricted to appear in an exponential or as a derivative, since they do not exist rigorously as quantum fields which have a well-defined scaling dimension. The metric is space-like, $\eta_{\mu\nu} = \text{diag}(-1, 1, ...1)$, $\mu = 0, 1, ..., d-1$; and the a_n^μ satisfy the hermiticity relations $a_n^{\mu\dagger} = a_{-n}^\mu$. In flat spacetime, momentum is conserved and it is convenient when quantizing to use a basis of momentum eigenstates, $|k\rangle = e^{ik\cdot q}|0\rangle$, $p^\mu|k\rangle = k|k\rangle$, to represent $[q^\mu, p^\nu] = i\eta^{\mu\nu}$, where the vacuum state $|0\rangle$ satisfies $a_n^\mu|0\rangle = 0, n \geq 0$. When the spacetime contains AdS, the isometry group no longer contains translations, so we lose momentum conservation and will just work in position space.

A *string* is a one-dimensional object that moves through spacetime and is governed by an action that describes the area of the worldsheet. Its trajectory $x^\mu(\sigma, \tau)$ describes the position of the string in space and time. τ is the evolution parameter $-\infty < \tau < \infty$, and σ is the spatial coordinate labelling points along the string; $0 \leq \sigma \leq \pi$ for the open string, which is topologically an infinite strip; and $0 \leq \sigma \leq 2\pi$ for the closed string which is topologically an infinite cylinder with periodicity condition $x^\mu(\sigma, \tau) = x^\mu(\sigma + 2\pi, \tau)$. $g^{\alpha\beta}(\sigma, \tau)$ is the two-dimensional metric, $\alpha, \beta = 0, 1$. The action for the bosonic string is

$$\begin{aligned} S_2 = -\tfrac{1}{4\pi\alpha'} \int d\sigma d\tau [\, & \sqrt{|g|} g^{\alpha\beta}\, \partial_\alpha x^\mu \partial_\beta\, x^\nu\, G_{\mu\nu}(x) \\ & + \epsilon^{\alpha\beta} \partial_\alpha x^\mu \partial_\beta\, x^\nu\, B_{\mu\nu}(x) \\ & + \alpha' \sqrt{|g|}\, R\, \phi(x)\,]\,. \end{aligned} \tag{3}$$

The Regge slope α' has dimensions of length squared $[L]^2$; the string tension is defined as $T = \frac{1}{2\pi\alpha'}$. The low energy limit of the string theory is an effective point field theory. This corresponds to the leading term in an expansion in the external momentum times $\sqrt{\alpha'}$. In the zero Regge slope limit,i.e. the infinite string tension limit, the interactions of the vectors and tensors are those of Yang-Mills bosons and gravitons. In the other limit, the zero tension limit, one expects an infinite number of massless particles.

It is conjectured for strings on AdS that both limits may be simple field theories, since interacting massless particles of spin higher than 2 appear to be consistent in an Einstein spacetime, whereas they are not in flat space.

For general $G_{\mu\nu}(x)$, from a $2d$ point of view, S_2 is a nontrivial interacting field theory: a conformally invariant non-linear sigma model. To quantize in flat spacetime, we choose $G_{\mu\nu}(x) = \eta_{\mu\nu}$, and the other background fields to vanish: the two form field potential $B_{\mu\nu}$ and the dilaton ϕ. This choice reduces S_2 to a free worldsheet theory. As a result, the correlation functions even at tree level in the string loop expansion are exact in α'. The Type IIB superstring on $AdS_5 \times S^5$ requires additional background fields besides the curved metric $G_{\mu\nu}(x)$. In this latter case, S_2 is a nontrivival worldsheet theory where the string tree level amplitudes will appear as an expansion in α'.

The string has two parameters: the length scale $\sqrt{2\alpha'}$ and the dimensionless string coupling constant g. These are both related to the dilaton vacuum expectation value $\langle 0|\phi|0\rangle$, and thus the string has no free dimensionless (nor dimensionful) parameters. The gravitational coupling is $\kappa \equiv \sqrt{8\pi G}$, Newton's constant is $G = \frac{\hbar c}{m^2_{PLANCK}}$; the Planck mass is $m_{\text{PLANCK}} = (\frac{\hbar c}{G})^{\frac{1}{2}} \sim 2.2 \times 10^{-5}\text{gm} \sim 1.2 \times 10^{19} GeV$; the Planck length is $l_{\text{PLANCK}} = (\frac{\hbar G}{c^3})^{\frac{1}{2}} \sim 1.6 \times 10^{-33}\text{cm}$. The Planck time is $t_{\text{PLANCK}} = (\frac{\hbar G}{c^5})^{\frac{1}{2}} \sim 10^{-43}$ sec.
At these scales, the effects of stringiness will be important, whereas at larger distance scales or lower energies, an ordinary point quantum field theory QFT can be used as an effective theory. In flat space and for $\langle 0|\phi|0\rangle = 0$, from identification of the graviton vertices we find $\kappa = \frac{1}{2} g\sqrt{2\alpha'}$, and from the Yang-Mill vertices, $g_{ym} = g$. Then $\alpha' \sim \frac{1}{m^2_{PLANCK}}$ i.e. $\alpha' \sim \kappa^2$, specifically $\alpha' = 2\kappa^2/g^2 = 16\pi G/g^2$. Thus the value of the universal Regge slope parameter α' is given in terms of Newton's constant G and the Yang-Mills coupling g. In particular since g is of order 1, α' is of order the Planck length (squared). We see that the scale of the entire unified string theory is set by the Planck mass. This scale does not appear to be associated with the secret of any symmetry breaking, as does the scale $\sqrt{\lambda/3}\, a = \sqrt{-2m^2} \equiv m_H$ given by the Higgs mass $m_H \sim 250\, Gev$. Discussions of the gauge hierarchy problem, *i.e.* why is m/m_{PLANCK} so small, what sets the ratio of m to m_{PLANCK}? and what is the origin of mass, *i.e.* how is $m \neq 0$? are beyond the scope of these lectures.

The worldsheet action S_2 has two-dimensional general coordinate invariance $g^{\alpha\beta} \to \partial_\gamma \Lambda^\alpha \partial_\delta \Lambda^\beta g^{\gamma\delta} \sim g^{\alpha\beta} + f^\gamma \partial_\gamma g^{\alpha\beta} + \partial_\gamma f^\alpha g^{\beta\gamma} + \partial_\gamma f^\beta g^{\alpha\gamma}$

and $x^\mu \longrightarrow x^\mu + f^\alpha \partial_\alpha x^\mu$; and local Weyl rescaling $g^{\alpha\beta} \longrightarrow \Lambda(\sigma,\tau) g^{\alpha\beta}$ and $x^\mu \longrightarrow x^\mu$. These symmetries allow one to make the *covariant gauge choice* $g^{\alpha\beta} = \eta^{\alpha\beta} = \mathrm{diag}(-1,1)$. This results in the remnant constraint equations $(\partial_\sigma x^\mu \pm \partial_\tau x^\mu)^2 = 0$ which are equivalent to $\partial_\sigma x \cdot \partial_\tau x = 0$ and $\partial_\sigma x \cdot \partial_\sigma x + \partial_\tau x \cdot \partial_\tau x = 0$.

The equations of motion are $\partial_\sigma^2 x^\mu - \partial_\tau^2 x^\mu = 0$. The open string boundary conditions are $\partial_\sigma x^\mu(\sigma,\tau) = 0$ at $\sigma = 0, \pi$. The general solution of the equations of motion with these string boundary conditions is

$$\begin{aligned} x^\mu(\sigma,\tau) &= q^\mu + p^\mu \tau + i \textstyle\sum_{n\neq 0} \frac{a_n^\mu}{n} e^{-in\tau} \cos n\sigma \\ &= q^\mu + p^\mu \tau + \frac{i}{2} \textstyle\sum_{n\neq 0} \frac{a_n^\mu}{n} (e^{-in(\tau+\sigma)} + e^{-in(\tau-\sigma)}) \\ &= \frac{1}{2} X^\mu(e^{i(\tau+\sigma)}) + \frac{1}{2} X^\mu(e^{i(\tau-\sigma)}) . \end{aligned} \tag{4}$$

In covariant gauge, we implement the constraint equations by observing that

$$(\partial_\sigma x^\mu \pm \partial_\tau x^\mu)^2 = z^2 L(z) , \qquad z = e^{i(\tau\pm\sigma)} \tag{5}$$

where

$$L(z) = \tfrac{1}{2} : a(z) \cdot a(z) := \sum_n L_n z^{-n-2} \tag{6}$$

and $a^\mu(z) \equiv i \frac{dX^\mu(z)}{dz} = \sum_n a_n z^{-n-1}$. So the expectation value of the constraints vanish (as $\hbar \longrightarrow 0$) for the *physical state conditions* in covariant gauge in the old covariant quantization:

$$L_0 |\psi\rangle = |\psi\rangle \tag{7}$$

$$L_n |\psi\rangle = 0 \quad \text{for}\, n > 0 . \tag{8}$$

Furthermore (7) is the mass shell condition $p^2 = -m^2$, where $\frac{1}{2} m^2 \equiv N - 1$ so that $L_0 = \frac{1}{2} p^2 + N = 1$. Here $N \doteq \sum_{n=1}^\infty a_{-n} \cdot a_n$. That is to say a physical state $|\psi\rangle$ has momentum k which takes on a specific value corresponding to the N^{th} excited level, $\alpha' k^2 = 1 - N$. The physical state conditions (7,8) can be shown to eliminate ghosts, *i.e.* negative norm states when $d = 26$. This result is known as the No-Ghost theorem.[2,3]

If instead, we had considered light-cone gauge, we would have observed that in the constraints $(\partial_\sigma x^\mu \pm \partial_\tau x^\mu)^2 = 0$ there is still residual gauge invariance to choose the light-cone gauge where $x^+(\sigma,\tau) = q^+ + p^+ \tau$, and the equations of motion are $(\partial_\sigma^2 - \partial_\tau^2)\, x^i = 0 ,\, 1 \leq i \leq d-2$.

2.2. *Locality of Worldsheet Fields*

A conformal field theory $\mathcal{H}$ is a Hilbert space of states H, such as the space of finite occupation number states in a Fock space, together with a set of vertex operators $V(\psi, z)$, i.e. conformal fields which are in one to one correspondence with the states $\psi \in \mathcal{F}(H)$, where $\mathcal{F}(H)$ is a dense subspace of the Hilbert space H of states.

$$\mathcal{H} = (H, \{V(\psi, z) : \psi \in \mathcal{F}(H)\}) . \tag{9}$$

The conformal field theory requires that the vertex operators $V(\psi, z)$ form a system of mutually local fields, where $\lim_{z \to 0} V(\psi, z)|0\rangle = \psi$ for each field. That is to say the conformal fields $V(\psi, z)$ acting on the vacuum create asymptotic "in" states $\psi = V(\psi, 0)|0\rangle$ with conformal weight h_ψ, $L_0\psi = h_\psi \psi$ (recall that $z = 0$ is $i\tau = t = -\infty$) on the cylinder. There is a one to one correspondence between the fields and the states in the Hilbert space they create at $z = 0$. Locality implies the s-t duality relation $V(\psi, z)V(\phi, \zeta) = V(V(\psi, z - \zeta)\phi, \zeta)$ which provides a precise version of the operator product expansion.[9] In particular locality requires

$$V(\psi, z)V(\phi, \zeta) \sim V(\phi, \zeta)V(\psi, z) \tag{10}$$

where the left side is defined for $|z| > |\zeta|$, the right side for $|\zeta| > |z| >$ and $\sim$ denotes analytic continuation. Locality ensures well defined scattering amplitudes. We shall also assume that the theory has a hermitian structure, in the sense that there is a definition of conjugation on the states, $\psi \mapsto \overline{\psi}$, an antilinear map with $V(\overline{\psi}, z) = z^{-2h} V(e^{z^* L_1}\psi, 1/z^*)^\dagger$.

2.3. *Virasoro Algebra*

One of the conformal fields is the Virasoro current

$$L(z) = V(\psi = \frac{1}{2} a_{-1} \cdot a_{-1}|0\rangle, z) = \sum_n L_n z^{-n-2} . \tag{11}$$

It satisfies the *operator product expansion*

$$L(z)L(\zeta) = \frac{c}{2}(z - \zeta)^{-4} + 2L(\zeta)(z - \zeta)^{-2} + \frac{dL(\zeta)}{d\zeta}(z - \zeta)^{-1} \tag{12}$$

which can be reexpressed as either of

$$\begin{aligned} [L_n, L(\zeta)] &= 2(n+1)\zeta^n L(\zeta) + \zeta^{n+1}\tfrac{dL(\zeta)}{d\zeta} + \zeta^{n-2}\tfrac{c}{12}(n^3 - n) \\ [L_n, L_m] &= (n - m)L_{n+m} + \tfrac{c}{12}(n^3 - n)\delta_{n,-m} . \end{aligned} \tag{13}$$

Vertex operators for *physical states* are primary fields. In this case ϕ is a highest weight state for the Virasoro algebra, or *primary state*, $V(\phi,\zeta)$ is a *primary field* and

$$L(z)V(\phi,\zeta) = h_\phi(z-\zeta)^{-2}V(\phi,\zeta) + (z-\zeta)^{-1}\frac{dV}{d\zeta}(\phi,\zeta)\,. \tag{14}$$

Therefore in covariant gauge, the vertex operators for physical states have to satisfy,[7]

$$[L_n,\, V(\psi,z)] = z^{n+1}\frac{d}{dz}V(\psi,z) + (n+1)z^n V(\psi,z) \tag{15}$$

for all n, and

$$\lim_{z\to 0} V(\psi,z)|0> = \psi\,. \tag{16}$$

This follows from $V(z) = \sum_r V_r z^{-r-h}$; so that $[L_n, V_r] = (-r + n(h-1))V_{n+r}$. Since $V_r|0>=0$ for $r>-h$, and $\psi = V(0)|0> = V_{-h}|0>$, then $L_n\psi = [L_n, V_{-h}]|0> = 0$ for $n \geq 1$, and physical fields are primary fields of conformal dimension $h=1$. (Note that if (15) holds only for $n = 0, \pm 1$ then $V(\phi,\zeta)$ is a *quasi-primary field.*)

2.4. *Mass Spectrum and Tree Level Amplitudes*

We now consider the first few mass levels. For $N=0$, there is one state $\psi = |k\rangle$ with $k^2 = 2$. The vertex operator for this state is

$$V(k,z) =: \exp\{ik{\cdot}X(z)\} := \exp\{ik{\cdot}X_<(z)\}e^{ik\cdot q}z^{k\cdot p}\exp\{ik{\cdot}X_>(z)\} \tag{17}$$

where $X^\mu_{\gtrless}(z) = i\sum_{n\gtrless 0}\frac{\alpha_n^\mu}{n}z^{-n}$. Since $z^{L_0}V(k,1)z^{-L_0} = zV(k,z)$, the four point open string tree amplitude for these tachyonic scalars is

$$\begin{aligned}
A_4 &= \alpha'\int_0^1 dz\, \langle 0;-k_1|V(k_2,1)V(k_3,z)|0;k_4\rangle \\
&= \alpha'(2\pi)^{26}\delta^{26}(k_1+k_2+k_3+k_4)\int_0^1 dz\, z^{k_3\cdot k_4}(1-z)^{k_2\cdot k_3} \\
&= \alpha'(2\pi)^{26}\delta^{26}(k_1+k_2+k_3+k_4)B(-1-\tfrac{1}{2}s,-1-\tfrac{1}{2}t) \\
&= \alpha'(2\pi)^{26}\delta^{26}(k_1+k_2+k_3+k_4)B(-1-\alpha' s,-1-\alpha' t)
\end{aligned} \tag{18}$$

Here the Mandelstam variables are $s = -(k_1+k_2)^2$, $t = -(k_2+k_3)^2$, and $u = -(k_1+k_3)^2$; the overall factor of α' is due to the propagator. As mentioned earlier, this string amplitude is exact in α', reflecting the fact the worldsheet theory has free operator products $a^\mu(z)a^\nu(\zeta) = \eta^{\mu\nu}(z-\zeta)^{-2}$. Here $a^\mu(z) = \sum_n a_n z^{-n-1}$. The three point open string tree amplitude for these tachyonic scalars is $A_3 = \langle 0; -k_1|V(k_2,1)|0;k_3\rangle = (2\pi)^{26}\delta^{26}(k_1 + k_2 + k_3 + k_4)$. The first excited level, $N = 1$, contains the vector states $\psi = \epsilon \cdot a_{-1}|k\rangle$ with $k^2 = 0$. In order to satisfy $L_n\psi = 0$ for $n > 0$, (in particular $L_1\psi = 0$) we must have $\epsilon \cdot k = 0$. The vertex operator for this state is

$$V(k,\epsilon,z) = \epsilon \cdot a(z)e^{ik\cdot X(z)} . \tag{19}$$

The three point open string tree amplitude for these massless vectors is

$$\begin{aligned} A_3 &= \langle -k_1|\epsilon_1 \cdot a_1 V(k_2,\epsilon_2,1)\epsilon_3 \cdot a_{-1}|k_3\rangle \\ &= (2\pi)^{26}\delta^{26}(k_1+k_2+k_3)\,\sqrt{2\alpha'}(\epsilon_1^\alpha \epsilon_2^\mu \epsilon_3^\lambda t_{\alpha\mu\lambda}(k_i) \\ &\quad + 2\alpha'(\epsilon_1 \cdot k_2 \epsilon_2 \cdot k_3 \epsilon_3 \cdot k_1))\,, \end{aligned} \tag{20}$$

where

$$t_{\alpha\mu\lambda}(k_i) \equiv k_{2\alpha}\eta_{\mu\lambda} + k_{3\mu}\eta_{\lambda\alpha} + k_{1\lambda}\eta_{\alpha\mu} \tag{21}$$

and we have recovered the dependence on the dimensional parameter α' by dimensional analysis, *i.e.* $k \to \sqrt{2\alpha'}k$. The four point open string tree amplitude for these massless vectors has tachyon poles and is given by Ref. 4 as

$$A_4 = \int_0^1 dz\,\langle -k_1|\epsilon_1 \cdot a_1 V(k_2,\epsilon_2,1)V(k_3,\epsilon_3,z)\epsilon_4 \cdot a_{-1}|k_4\rangle$$

$$= \int_0^1 dz\, z^{-\alpha' s}\,\langle -k_1 - k_2|\epsilon_1 \cdot a_1 \epsilon_2 \cdot a(1) e^{-k_2\cdot\sum_{n<0}\frac{\alpha_n^\mu}{n}z^{-n}} e^{-k_2\cdot\sum_{n>0}\frac{\alpha_n}{n}z^{-n}}$$

$$\cdot e^{-k_3\cdot\sum_{n<0}\frac{\alpha_n}{n}z^{-n}} e^{-k_3\cdot\sum_{n>0}\frac{\alpha_n}{n}z^{-n}}\epsilon_3 \cdot a(z)\ \epsilon_4 \cdot a_{-1}|k_4\rangle$$

$$\begin{aligned} &= (2\pi)^{26}\delta^{26}(k_1+k_2+k_3+k_4)\int_0^1 dz\, z^{-\alpha' s - 2}\,(1-z)^{-\alpha' t} \\ &\quad \cdot[\epsilon_1 \cdot \epsilon_2 \epsilon_3 \cdot \epsilon_4 \\ &\quad -\epsilon_1 \cdot \epsilon_2 \epsilon_4 \cdot k_3\{\epsilon_3 \cdot k_4 - \epsilon_3 \cdot k_2 z(1-z)^{-1}\} \\ &\quad -\epsilon_3 \cdot \epsilon_4 \epsilon_1 \cdot k_2\{-\epsilon_2 \cdot k_1 + \epsilon_2 \cdot k_3 z(1-z)^{-1}\} \\ &\quad -\epsilon_1 \cdot k_2 \epsilon_4 \cdot k_3\{\epsilon_2 \cdot \epsilon_3 z(1-z)^{-2} - \epsilon_2 \cdot k_1 \epsilon_3 \cdot k_4 \\ &\quad +(\epsilon_3 \cdot k_4 \epsilon_2 \cdot k_3 + \epsilon_2 \cdot k_1 \epsilon_3 \cdot k_2)z(1-z)^{-1} - \epsilon_3 \cdot k_2 \epsilon_2 \cdot k_3 z^2(1-z)^{-2}\}]\,. \end{aligned} \tag{22}$$

2.5. *Closed Bosonic String*

The closed string satifies the same equations of motion $\partial_\sigma^2 x^\mu - \partial_\tau^2 x^\mu = 0$ but is toplogically a cylinder with boundary condition $x^\mu(\sigma,\tau) = x^\mu(\sigma + 2\pi, \tau)$. The general solution is

$$x^\mu(\sigma,\tau) = q^\mu + p^\mu \tau + \frac{i}{2}\sum_{n\neq 0}\{\frac{a_n^{L\mu}}{n}e^{-in(\tau+\sigma)} + \frac{a_n^{R\mu}}{n}e^{-in(\tau-\sigma)}\}$$

$$= \frac{1}{2}X_L^\mu(e^{i(\tau+\sigma)}) + \frac{1}{2}X_R^\mu(e^{i(\tau-\sigma)}) \tag{23}$$

where $X_L^\mu(z) = q^\mu - ip_L^\mu \ln z + i\sum_{n\neq 0}\frac{a_n^{L\mu}}{n}z^{-n}$, $X_R^\mu(z) = q^\mu - ip_R^\mu \ln z + i\sum_{n\neq 0}\frac{a_n^{R\mu}}{n}z^{-n}$, and $p_L = p_R = p$. The covariant quantization conditions are

$$[a_m^{L\mu}, a_n^{L\nu}] = [a_m^{R\mu}, a_n^{R\nu}] = \eta^{\mu\nu} m\delta_{m,-n}; \quad [q^\mu, p^\nu] = \frac{i}{2}\eta^{\mu\nu} \tag{24}$$

$$[q^\mu, a_n^{L\nu}] = [q^\mu, a_n^{R\nu}] = 0, \quad n \neq 0 \tag{25}$$

$$[a_m^{L\mu}, a_n^{R\nu}] = 0; \quad a_0^{L\mu} \equiv a_0^{R\mu} \equiv p^\mu\,. \tag{26}$$

The physical state conditions are

$$L_0^L|\psi\rangle = |\psi\rangle\,, \quad L_n^L|\psi\rangle = 0\,, \quad \text{for}\, n > 0\,, \tag{27}$$

$$L_0^R|\psi\rangle = |\psi\rangle\,, \quad L_n^R|\psi\rangle = 0\,, \quad \text{for}\, n > 0\,. \tag{28}$$

The mass shell condition is $p^2 = -m^2$ where $m^2 = N_L + N_R - 2$ and $N_L = N_R$. L_n^R, L_N^L are two commuting Virasoro algebras, both with $c = d = 26$. In this quantization of the closed string in flat spacetime, the theory is seen to be a tensor product of left and right copies of the open string case. One can form vertices for the closed string by taking the tensor products of open string conformal fields for the left- and right movers, and using the variable z for the left vertices and $\bar{z}$ with the right vertices. Here we have defined a euclidean world sheet metric $i\tau \equiv t$, so that $z = e^t e^{i\sigma}$, $\bar{z} = e^t e^{-i\sigma}$, which maps the cylinder traced by the moving string onto the complex plane.[7] Time ordering is radial ordering; t = constant hypersurfaces are circles concentric about the origin of the $z - plane$. In fact the local operator product relations will extend naturally to arbitrary Riemann surfaces with local conformal coordinates z,$\bar{z}$. Since for closed strings, $(L_0^L + L_0^R)|\psi\rangle = 2|\psi\rangle$, the tachyon $|k\rangle \equiv e^{i2k\cdot q}|0\rangle$, $k^2 = 2$ has vertex operator

$$V(k,z,\bar{z}) =: \exp\{ik\cdot X_L(z)\} :: \exp\{ik\cdot X_R(\bar{z})\} : \tag{29}$$

When the spacetime is not flat, the conformal fields in general will not factor into left times right, although the two copies of the Virasoro algebra will remain holomorphic (and antiholomorphic).[12]

The four point closed string tree amplitude for the tachyonic scalars in flat spacetime is again exact in α':

$$\begin{aligned} A_4 &= \tfrac{\alpha'}{2\pi} \int d^2z \, \langle -k_1 | V(k_2,1,1) z\bar{z} V(k_3,z,\bar{z}) | k_4 \rangle \\ &= \tfrac{\alpha'}{2\pi} (2\pi)^{26} \delta^{26}(k_1+k_2+k_3+k_4) \int d^2z \, |z|^{2k_3\cdot k_4} |1-z|^{2k_2\cdot k_3} \\ &= \tfrac{\alpha'}{2\pi} (2\pi)^{26} \delta^{26}(k_1+k_2+k_3+k_4) \tfrac{\Gamma(-1-\alpha' s)\Gamma(-1-\alpha' t)\Gamma(-1-\alpha' u)}{\Gamma(2+\alpha' s)\Gamma(2+\alpha' t)\Gamma(2+\alpha' u)} \end{aligned} \tag{30}$$

which is totally symmetric under the exchange of s,t,u. The massless level $N_L = N_R = 1$ contains the states $\psi = \epsilon^L \cdot a^L_{-1} \epsilon^R \cdot a^R_{-1} |k\rangle$ with $k^2 = 0$, and $\epsilon^L \cdot k = \epsilon^R \cdot k = 0$, forming the spin two graviton, the 2-form antisymmetric tensor, and the dilaton.

The preceeding analysis of physical state conditions can be reexpressed in an equivalent formulation using BRST cohomology,[7] which includes both the "matter" sector described above and a BRST "ghost" sector with equal and opposite central charge. The worldsheet variables used to formulate strings in curved spacetime with Ramond flux, do not exhibit such a 'matter times ghost" factorization,[10,12,20] although they still define physical state conditions as a version of cohomology.

3. Various Formulations of Superstring Worldsheet Fields

3.1. *Ramond-Neveu-Schwarz (RNS)*

This a Lorentz covariant but not *manifestly* supersymmetric quantization. The Neveu-Schwarz (NS) fields $b^\mu(z) = \sum_s b^\mu_s z^{-s-\frac{1}{2}}$ have $\{b^\mu_r, b^\nu_s\} = \eta^{\mu\nu}\delta_{r,-s}$; where $r,s \in \mathbf{Z} + \frac{1}{2}$ and $b^{\mu\dagger}_s = b^\mu_{-s}$. They are used to construct the super Virasoro generators $G(z) = a(z)\cdot b(z)$ and $L(z) = \frac{1}{2} : a(z)\cdot a(z) : + \frac{1}{2} : \frac{db(z)}{dz} \cdot b(z) :$ which satisfy the $N=1$ super Virasoro algebra

$$\begin{aligned} [L_m, L_n] &= (m-n)L_{n+m} + \tfrac{c}{12}(m^3 - m)\delta_{m,-n} \\ \{G_r, G_s\} &= 2L_{r+s} + \tfrac{c}{3}(r^2 - \tfrac{1}{4})\delta_{r,-s} \\ [L_n, G_s] &= (\tfrac{n}{2} - s)G_{n+s} \,. \end{aligned} \tag{31}$$

The No-Ghost theorem selects $c = \frac{3}{2}d = 15$. The *physical state* conditions in the $\mathcal{F}_2$-picture are $L_0|\psi\rangle = \frac{1}{2}|\psi\rangle$, $L_n|\psi\rangle = 0$ for $n > 0$, $G_s|\psi\rangle = 0$ for $s > 0$. In a superconformal field theory the states are in one-to-one correspondence with a conformal superfield $V(\psi,z,\vartheta) = V_0(\psi,z) + \vartheta V_1(\psi,z)$, where ϑ is a fermionic coordinate, the supersymmetry partner of z. $V_0(\psi,z)$

and $V_1(\psi, z)$ are called the lower and upper components of the superfield respectively.

$$\lim_{z\to 0} V_0(\psi, z)|0\rangle = |\psi\rangle\,, \quad \lim_{z\to 0} V_1(\psi, z)|0\rangle = G_{-\frac{1}{2}}|\psi\rangle\,. \tag{32}$$

In covariant gauge, the vertex operators for physical states have to satisfy

$$\begin{aligned}
[L_n, V_0(\psi, z)] &= z^{n+1}\tfrac{d}{dz}V_0(\psi, z) + h(n+1)z^n V_0(\psi, z)\\
[L_n, V_1(\psi, z)] &= z^{n+1}\tfrac{d}{dz}V_1(\psi, z) + (h+\tfrac{1}{2})(n+1)z^n V_1(\psi, z)\\
\{G_s, V_0(\psi, z)\}_{\pm} &= z^{s+\frac{1}{2}}V_1(\psi, z)\\
[G_s, V_1(\psi, z)]_{\mp} &= z^{s+\frac{1}{2}}\tfrac{dV_0(\psi,z)}{dz} + 2h(s+\tfrac{1}{2})z^{s-\frac{1}{2}}V_0(\psi, z)
\end{aligned} \tag{33}$$

for all n and s, for $h = \frac{1}{2}$. At level $N = 0$, there is one state $\psi = |k\rangle$ with $k^2 = 1$. Its vertex operator has components

$$V_0(k, z) =: \exp\{ik\cdot X(z)\} :\,, \quad V_1(k, z) = \sqrt{2\alpha'}\,k\cdot b(z) : \exp\{ik\cdot X(z)\} : \,. \tag{34}$$

Unlike the bosonic case, the three point amplitude for the Neveu-Schwarz tachyon vanishes: $A_3 = \langle -k_1|V_1(k_2, 1)|k_3\rangle = 0\,.$

The massless vector is at level $N = \frac{1}{2}$, $\psi = \epsilon\cdot b_{-\frac{1}{2}}|k\rangle$ with $k^2 = 0$. To satisfy $L_n\psi = 0$ for $n > 0$, $G_s\psi = 0$ for $s > 0$ (in particular $G_{\frac{1}{2}}\psi = 0$) we must have $\epsilon\cdot k = 0$. Its vertex operator has components

$$\begin{aligned}
V_0(k, \epsilon, z) &= \epsilon\cdot b(z)e^{ik\cdot X(z)}\,,\\
V_1(k, \epsilon, z) &= \left\{\sqrt{2\alpha'}k\cdot b(z)\,\epsilon\cdot b(z) + \epsilon\cdot a(z)\right\}\exp\{ik\cdot X(z)\}\,.
\end{aligned} \tag{35}$$

For this state, $G_{-\frac{1}{2}}\psi = (k\cdot b_{-\frac{1}{2}}\epsilon\cdot b_{-\frac{1}{2}} + \epsilon\cdot a_{-1})|k\rangle$. The three point open string tree amplitude for these massless vectors is

$$\begin{aligned}
A_3 &= \langle -k_1|\epsilon_1\cdot b_{\frac{1}{2}}V_1(k_2, \epsilon_2, 1)\epsilon_3\cdot b_{-\frac{1}{2}}|k_3\rangle\\
&= (2\pi)^{10}\delta^{10}(k_1 + k_2 + k_3)\,\sqrt{2\alpha'}\,\epsilon_1^\alpha\epsilon_2^\mu\epsilon_3^\lambda\,t_{\alpha\mu\lambda}(k_i)\,.
\end{aligned} \tag{36}$$

We see that the Neveu-Schwarz computation of A_3 has no α' corrections, in contrast with (20). Non-renormalization theorems for the superstring often prevent α' corrections to the tree level three point functions, although not to four point functions. For both expressions, the zero slope limit reduces to the conventional three gluon field theory coupling: $\lim_{\alpha'\to 0} A_3\frac{1}{\sqrt{2\alpha'}} = (2\pi)^d\delta^{10}(\sum_i k_i)\epsilon_1^\alpha\epsilon_2^\mu\epsilon_3^\lambda\,t_{\alpha\mu\lambda}(k_i)\,.$

The space of states for the open *superstring*, $\tilde{\mathcal{H}}$, is obtained by starting with the states of the untwisted Neveu-Schwarz theory, $\mathcal{H}$, introduced above, then adding in keeping only the subspace of each defined by $\theta = 1$, with $\theta^2 = 1$. The states of the untwisted theory are generated by the action of d infinite sets of half-integrally moded oscillators,

b_s^μ, $0 \le \mu \le d-1$ (together with the integrally moded oscillators, a_n^μ), on the vacuum state, $|0\rangle$. The twisted sector is obtained from the action of d infinite sets of integrally moded oscillators, d_n^μ, (together with the integrally moded oscillators, a_n^μ), on the twisted ground states which form a $2^{d/2}$ irreducible representation, $\mathcal{X}$, of the gamma matrix Clifford algebra, $\{\gamma^\mu\}$. The involution θ is defined on the untwisted space $\mathcal{H}$ by $\theta|0\rangle = (-1)|0\rangle$, $\theta b_s^\mu \theta^{-1} = -b_s^\mu$, and on the twisted space, $\mathcal{H}_T$, by $\theta|0\rangle_R^\pm = \pm|0\rangle_R^\pm$, $\theta d_n^\mu \theta^{-1} = -d_n^\mu$, where $\mathcal{X} = |0\rangle_R^+ + |0\rangle_R^-$ Whenever d is even we can define $\gamma^{d+1} \equiv \gamma^1\gamma^2\ldots\gamma^d$ which satisfies $\{\gamma^{d+1},\gamma^\mu\} = 0$, $(\gamma^{d+1})^2 = 1$. The operators $\frac{1}{2}(1 \pm \gamma^{d+1}(-1)^{\sum_{n>0} d_{-n}\cdot d_n})$ are chirality projection operators; they project onto spinors of definite chirality. A spinor of definite chirality is called a Weyl spinor; and the restriction to spinors of one chirality or the other is called a Weyl condition. In the Neveu-Schwarz sector, $\theta \equiv (-1)(-1)^{\sum_{s>0} b_{-s}\cdot b_s}$, and in the Ramond sector, $\theta \equiv \gamma^{d+1}(-1)^{\sum_{n>0} d_{-n}\cdot d_n}$. The Ramond Fock space splits into two $SO(d)$ invariant subspaces, according to the eigenvalue of θ, the ground states being denoted by $|0\rangle_R^\pm$. The $\theta = 1$ subspace is thus a projection onto the odd b sector, and onto chiral fermions in the Ramond sector, and is known as the *Gliozzi-Scherk-Olive* (GSO) projection. The worldsheet fermion fields are $\psi^\mu(z)$ in either the Neveu-Schwarz $\psi^\mu(z) = b^\mu(z)$ or Ramond representation $\psi^\mu(z) = d^\mu(z) = \sum_n d_n^\mu z^{-n-\frac{1}{2}}$, representing the vertices for emitting the massless vector state from a Neveu-Schwarz or Ramond line, respectively. The ground state $|a\rangle$ of the Ramond sector of the superstring is a spacetime fermion which is in one-to-one correspondence with a worldsheet field $S_a(z)$ called a *spin field*:

$$|a\rangle = \lim_{z\to 0} S_a(z)|0\rangle\,. \tag{37}$$

The spin fields are non-local with respect to the ordinary superconformal fields $\psi^\mu(z)$:

$$\begin{gathered}\psi^\mu(z)S^a(\zeta) \sim (z-\zeta)^{-\frac{1}{2}}\gamma^{\mu\,a}{}_{\dot b}S^{\dot b}(\zeta)\\ \psi^\mu(z)\psi^\nu(\zeta) \sim (z-\zeta)^{-1}\\ a^\mu(z)a^\nu(\zeta) \sim (z-\zeta)^{-2}\,, \quad a^\mu(z)\psi^\nu(\zeta) \sim 0\,, \quad a^\mu(z)S^a(\zeta) \sim 0\end{gathered} \tag{38}$$

due to the non-meromorphic structure of the operator product of ψ with S. It follows that the worldsheet supercurrents $G(z)$ are not local with respect to the spin fields. It follows that in the presence of *background Ramond fields*, the superconformal invariance of the worldsheet action S_2 would appear to be violated: the supersymmetrization of the bosonic Polyakov action (3) contains terms of the form

$$S_2 = -\frac{1}{4\pi\alpha'} \int d\sigma d\tau \sqrt{|g|}(\partial_\alpha x^\mu \partial_\beta x^\nu + i\bar{\psi}^\mu \gamma^\beta \partial_\beta \psi^\nu)\, G_{\mu\nu}(x) \tag{39}$$

and is superconformally invariant, called the guiding principle of the RNS description of perturbative superstrings. When a background Ramond field $\mathcal{B}_{\mu\nu}$ is required, one might try to add to S_2 a term such as

$$\int d\sigma d\tau S^L_a S^R_b [\gamma^\mu, \gamma^\nu]^{ab} B_{\mu\nu}(x) \tag{40}$$

but the presence of spin fields jeopardizes the *super*conformal invariance. Several efforts to answer this problem have been made, some of which we cite here.[30,31,32,33,12,20] One formulation[12,20] which survives quantization dispenses with spin fields altogether. It is discussed in sections 3.3 and 4.

An extension of the RNS superstring to include BRST ghost fields recasts the physical state conditions as cohomology. In this reformulation, in addition to the left and right moving "matter" fields $X^\mu(z,\bar{z})$, $\psi^\mu_L(z), \psi^\mu_R(\bar{z})$ combining to give central charge $c = 15$, there are "ghost" fields $b_L(z), c_L(z), \beta_L(z), \gamma_L(z)$ and $b_R(\bar{z}), c_R(\bar{z}), \beta_R(\bar{z}), \gamma_R(\bar{z})$ contributing $c = -15$. Left and right-moving BRST charges each have the structure $Q \sim c(L_m + \frac{1}{2}L_g) + \gamma(G_m + \frac{1}{2}G_g)$ and the left and right Virasoro generators have zero central charge $L \sim L_m + L_g$.

3.2. *Green-Schwarz (GS)*

This is a supersymmetric but not Lorentz covariant quantization. The open string worldsheet fields are $S_a(z)$, $a^\mu(z)$ and satisfy meromorphic operator products

$$S^a(z)\bar{S}^b(\zeta) \sim (z-\zeta)^{-1}c^{ab},\; a^i(z)a^j(\zeta) \sim (z-\zeta)^{-2}\delta^{ij},\; a^i(z)S^a(\zeta) \sim 0 \tag{41}$$

since the coordinates are limited to the light-cone $1 \leq i,j \leq 8$, and $1 \leq a,b \leq 8$. Here the RNS spin field $S^a(z)$ has been promoted to a fundamental worldsheet variable, whereas in the RNS case in fact it is expressible as a combination of the ψ fields, $S \sim e^{b\dots d}$. The Green-Schwarz formulation of the superstring dispenses with the need to sum over different spin stuctures (related to the NS and R sectors) in the one-loop string amplitudes.

3.3. *Berkovits-Vafa-Witten (BVW)*

This is a covariant and supersymmetric quantization in six spacetime dimensions. It has been applied primarily to compactifications of the

Type IIB string either in the "flat" case $R^6 \times K3$, or the "curved" case $AdS_3 \times S^3 \times K3$. BVW provides a partially covariant quantization of the Green-Schwarz superstring. Eight of the sixteen supersymmetries are manifest, in the sense that they act geometrically on the target space of the worldsheet sigma model. In addition, there are no worldsheet spin fields and so can more easily incorporate Ramond-Ramond background fields.

The BVW worldsheet fields are $X^m, \theta^a, \bar{\theta}^a$, the conjugate fermions $p^a, \bar{p}^a$ for $1 \leq m \leq 6; 1 \leq a \leq 4$, and two additional worldsheet bosons $\rho, \sigma, \bar{\rho}, \bar{\sigma}$. These describe the $d = 6$ part of the Type IIB string. The $K3$ part is described by the standard RNS description of a T^4/Z_2 orbifold. The θ^a's are ordinary conformal fields, not spin fields. In flat space, the worldsheet variables are holomorphic and satisfy free operator products relations including

$$p_a(z)\theta^b(\zeta) \sim (z-\zeta)^{-1}\delta_a^b \,. \tag{42}$$

In curved space, the worldsheet fields are no longer holomorphic, and the worldsheet action becomes a sigma model with the supergroup $PSU(2|2)$ as target, which is no longer a free conformal field theory nor a Wess-Zumino-Witten (WZW) model.

A ten dimensional version of these variables has appeared recently.[20–25] In flat spacetime, field redefinitions give back the RNS formalism. In $AdS_5 \times S^5$, vertex operator constraint equations have been considered.[23] The Berkovits variables are given by the ten-dimensional superspace variables $X^\mu(z,\bar{z})$, $\theta^\alpha_L(z,\bar{z})$, $\theta^\alpha_R(z,\bar{z})$, for $0 \leq \mu \leq 9$, $1 \leq \alpha \leq 16$; and the conjugate fermionic worldsheet fields $p^\alpha_L(z,\bar{z})$, $p^\alpha_R(z,\bar{z})$. There are additional worldsheet bosons that are spacetime spinors $\lambda^\alpha_L(z,\bar{z}), \lambda^\alpha_R(z,\bar{z})$, which separately satisfy $\lambda^\alpha \gamma^\mu_{\alpha\beta} \lambda^\beta = 0$ and carry 22 degrees of freedom. The construction includes left and right-moving BRST charge operators and Virasoro generators. The contribution to the central charge is $10 + 22$ from the worldsheet bosons, and -32 from the worldsheet fermions. The variables do not exhibit the "matter times ghost" structure of conventional the BRST formalism. For $AdS_5 \times S^5$ spacetime, the worldsheet fields are not holomorphic.

4. Type IIB Superstrings on $AdS_3 \times S^3 \times K3$

Compatification of the Type IIB superstring on either $R^6 \times K3$ or $AdS_3 \times S^3 \times K3$ yields a $d = 6$, $N = (2,0)$ theory which has sixteen supercharges. The $6d$ massless particle content is a supergravity multiplet and 21 tensor multiplets. In flatspace the multiplets are representations of the light-cone little group $SO(4)$: sg $(3,3)+5(3,1)+4(3,2)$, tensor $(1,3)+5(1,1)+4(1,2)$. In curved space the number of physical degrees of freedom in the multiplets remains the same. The "compactification independent" 6d fields make up the supergravity multiplet and one of the tensor multiplets, they are the graviton $(3,3)$, an antisymmetric tensor $(3,1)+(1,3)$, and a scalar $(1,1)$: $g_{mn}(x), b_{mn}(x), \phi(x)$; four self-dual tensors $(3,1)$ and four scalars contained in $V_{aa}^{--}(x), F^{++}(x), A_a^{-+\bar{a}}(x), A_{\bar{a}}^{+-a}$; and four gravitinos $(3,2)$ and four spinors $(1,2)$ contained in $\xi_{ma}^{-}(x), \bar{\xi}_{ma}^{-}(x), \chi_m^a(x), \bar{\chi}_m^a(x)$.

The vertex operator which describes these states is given in terms of the (compactification independent) worldsheet fields $X^m, \theta^a, \bar{\theta}^a$ by the superfield

$$
\begin{aligned}
V_{1,1} = {} & \theta^a \bar{\theta}^{\bar{a}} V_{a\bar{a}}^{--} + \theta^a \theta^b \bar{\theta}^{\bar{a}} \sigma_{ab}^m \bar{\xi}_{m\,\bar{a}}^{-} + \theta^a \bar{\theta}^{\bar{a}} \bar{\theta}^{\bar{b}} \sigma_{\bar{a}\bar{b}}^m \xi_{m\,a}^{-} \\
& + \theta^a \theta^b \bar{\theta}^{\bar{a}} \bar{\theta}^{\bar{b}} \sigma_{ab}^m \sigma_{\bar{a}\bar{b}}^n (g_{mn} + b_{mn} + \bar{g}_{mn}\phi) + \theta^a (\bar{\theta}^3)_{\bar{a}} A_a^{-+\,\bar{a}} + (\theta^3)_a \bar{\theta}^{\bar{a}} A_{\bar{a}}^{+-\,a} \\
& + \theta^a \theta^b (\bar{\theta}^3)_{\bar{a}} \sigma_{ab}^m \bar{\chi}_m^{+\,\bar{a}} + (\theta^3)^a \bar{\theta}^{\bar{a}} \bar{\theta}^{\bar{b}} \sigma_{\bar{a}\bar{b}}^m \chi_m^{+\,a} + (\theta^3)_a (\bar{\theta}^3)_{\bar{a}} F^{++\,a\bar{a}} \,. \qquad (43)
\end{aligned}
$$

The string constraint equations which select the physical states are generated in this formalism by a topological $N=4$ superVirasoro algebra we will discuss in the next section. For flat spacetime, the constraints will result in that all the above 6d fields satisfy $\partial^m \partial_m \phi = 0$ and

$$
\begin{aligned}
& \partial^m g_{mn} = -\partial_n \phi \,, \quad \partial^m b_{mn} = 0 \,, \quad \partial^m \chi_m^{\pm b} = \partial^m \bar{\chi}_m^{\pm \bar{b}} = 0 \\
& \partial_{ab} \chi_m^{\pm b} = \partial_{\bar{a}\bar{b}} \chi_m^{\pm \bar{b}} = 0 \,, \quad \partial_{cb} F^{\pm\pm b\bar{a}} = \partial_{\bar{c}\bar{b}} F^{\pm\pm \bar{b}a} = 0 \,, \qquad (44)
\end{aligned}
$$

where

$$
\begin{aligned}
F^{+-a\bar{a}} = \partial^{\bar{a}\bar{b}} A_{\bar{b}}^{+-a} \quad , \quad & F^{-+a\bar{a}} = \partial^{ab} A_b^{-+\bar{a}} \,, \quad F^{--a\bar{a}} = \partial^{ab} \partial^{\bar{a}\bar{b}} V_{b\bar{b}}^{--} \\
\chi_m^{-a} = \partial^{ab} \xi_{mb}^{-} \,, \quad & \bar{\chi}_m^{-\bar{a}} = \partial^{\bar{a}\bar{b}} \bar{\xi}_{m\bar{b}}^{-} \,. \qquad (45)
\end{aligned}
$$

These are equivalent to the equations of motion for $D=6$, $N=(2,0)$ supergravity[15] with one tensor multiplet expanded around the six-dimensional Minkowski metric.

In the curved case $AdS_3 \times S^3$, the constraints will result in a different set of equations of motion for the 6d fields. We give the answer here for the bosonic 6d fields, and show the derivation in section 4.2. The six-dimensional metric field g_{rs}, the dilaton ϕ, and the two-form b_{rs} satisfy

$$
\begin{aligned}
\tfrac{1}{2} D^p D_p b_{rs} = {} & -\tfrac{1}{2} (\sigma_r \sigma^p \sigma^q)_{ab} \delta^{ab} D_p \left[g_{qs} + \bar{g}_{qs} \phi \right] \\
& + \tfrac{1}{2} (\sigma_s \sigma^p \sigma^q)_{ab} \delta^{ab} D_p \left[g_{qr} + \bar{g}_{qr} \phi \right] \\
& - \bar{R}_{\tau r s \lambda} b^{\tau\lambda} - \tfrac{1}{2} \bar{R}_r{}^{\tau} b_{\tau s} - \tfrac{1}{2} \bar{R}_s{}^{\tau} b_{r\tau} \\
& + \tfrac{1}{4} F_{\text{asy}}^{++gh} \sigma_r^{ab} \sigma_s^{ef} \delta_{ah} \delta_{be} \delta_{gf} \qquad (46)
\end{aligned}
$$

$$\tfrac{1}{2}D^pD_p\,(g_{rs}+\bar{g}_{rs}\phi) = -\tfrac{1}{2}(\sigma_r\sigma^p\sigma^q)_{ab}\delta^{ab}\,D_pb_{qs} + \tfrac{1}{2}(\sigma_s\sigma^p\sigma^q)_{ab}\delta^{ab}\,D_pb_{rq}$$
$$- \bar{R}_{\tau rs\lambda}\,(g^{\tau\lambda}+\bar{g}^{\tau\lambda}\phi) - \tfrac{1}{2}\bar{R}_r{}^{\tau}\,(g_{\tau s}+\bar{g}_{\tau s}\phi)$$
$$- \tfrac{1}{2}\bar{R}_s{}^{\tau}\,(g_{r\tau}+\bar{g}_{r\tau}\phi) + \tfrac{1}{4}F_{\rm sym}^{++gh}\,\sigma_{rga}\sigma_{shb}\,\delta^{ab}\,. \quad (47)$$

This is the curved space version of the flat space zero Laplacian condition $\partial^p\partial_pb_{rs}=\partial^p\partial_pg_{rs}=\partial^p\partial_p\phi=0$.

Four self-dual tensor and scalar pairs come from the string bispinor fields F^{++ab}, V_{ab}^{--}, A_b^{+-a}, A_a^{-+b}. From the string constraint equations they satisfy

$$\sigma^p_{da}\,D_p\,F_{\rm asy}^{++ab}=0 \quad (48)$$

$$\tfrac{1}{4}\,[\,\delta^{Ba}\,\sigma^r_{ga}\,D_r\,F_{\rm sym}^{++gH} - \delta^{Ha}\,\sigma^r_{ga}\,D_r\,F_{\rm sym}^{++gB}\,] = -\tfrac{1}{4}\,\epsilon^{BH}{}_{cd}\,F_{\rm asy}^{++cd} \quad (49)$$

These can be shown[13] to be equivalent to the linearized supergravity equations[16] for the supergravity multiplet and one tensor multiplet of the $d=6$, $N=(2,0)$ theory expanded around the $AdS_3\times S^3$ metric and a self-dual three-form, by using the following field identifications: the vertex operator components in terms of the supergravity fields $g_{prs}, g^6_{prs}, h_{rs}, \phi^i$, $1\le i\le 5$, (and $2\le I\le 5$) are

$$H_{prs} \equiv g^6_{prs} + 2\,g^1_{prs} + B^I\,g^I_{prs} \quad (50)$$
$$g_{rs} \equiv h_{rs} - \tfrac{1}{6}\bar{g}_{rs}\,h^\lambda_{\ \lambda} \quad (51)$$
$$\phi = -\tfrac{1}{3}\,h^\lambda_{\ \lambda} \quad (52)$$
$$F_{\rm sym}^{++ab} = \tfrac{2}{3}(\sigma_p\sigma_r\sigma_s)^{ab}\,B^I\,g^I_{prs} + \delta^{ab}\,\phi^{++}$$
$$F_{\rm asy}^{++ab} = \sigma^{p\,ab}\,D_p\,\phi^{++}$$
$$\phi^{++} = 4C^I\,\phi^I \quad (53)$$

which follows from choosing the graviton trace $h^\lambda_{\ \lambda}$ to satisfy $\phi^1 - h^\lambda_{\ \lambda} \equiv -2\,C^I\phi^I$. Here $H_{prs}\equiv\partial_pb_{rs}+\partial_rb_{sp}+\partial_sb_{pr}$. The combinations $C^I\phi^I$ and $B^Ig^I_{prs}$ reflect the $SO(4)_{\rm R}$ symmetry of the $D=6, N=(2,0)$ theory on $AdS_3\times S^3$. We relabel $C^I=C^I_{++}$, $B^I=B^I_{++}$. To define the remaining string components in terms of supergravity fields, we consider linearly independent quantities $C^I_\ell \phi^I$, $B^I_\ell g^I_{prs}$, $\ell = ++, +-, -+, --$.

$$\begin{aligned}
F_{\text{sym}}^{+-ab} &= \tfrac{2}{3}(\sigma_p\sigma_r\sigma_s)^{ab}\, B^I_{+-}\; g^I_{prs} + \delta^{ab}\,\phi^{+-} \\
F_{\text{asy}}^{+-ab} &= \sigma^{p\,ab}\, D_p\,\phi^{+-} \\
\phi^{+-} &= 4C^I_{+-}\;\phi^I \\
F_{\text{sym}}^{-+ab} &= \tfrac{2}{3}(\sigma_p\sigma_r\sigma_s)^{ab}\, B^I_{-+}\; g^I_{prs} + \delta^{ab}\,\phi^{-+} \\
F_{\text{asy}}^{-+ab} &= \sigma^{p\,ab}\, D_p\,\phi^{-+} \\
\phi^{-+} &= 4C^I_{-+}\;\phi^I
\end{aligned} \tag{54}$$

V^{--}_{ab} is given in terms of the fourth tensor/scalar pair $C^I_{--}\;\phi^I$, $B^I_{--}g^I_{mnp}$ through

$$\begin{aligned}
D^pD_pV^{--}_{cd} - \delta^{gh}\sigma^p_{ch}\,D_pV^{--}_{gd} + \delta^{gh}\sigma^p_{dh}\,D_pV^{--}_{cg} + \tfrac{1}{2}\epsilon_{cd}{}^{gh}\,V^{--}_{gh}& \\
= -8\,\sigma^m_{ce}\,\sigma^n_{df}\,\delta^{ef}\,G_{mn}\,.&
\end{aligned} \tag{55}$$

4.1. *Topological Strings*

The origin of the constraints is an $N = 4$ twisted superconformal algebra. In this section, we review how the superstring can be reformulated as an $N = 4$ topological string theory, and show how this formalism gives a description of the superstring with manifest $d = 6$ spacetime supersymmetry. That is to say, the spectrum of the superstring can be identified with the states surviving a set of $N = 4$ constraints. We begin this subsection by remembering how the bosonic string can be reorganized as an $N = 2$ topological string.[10,12] An $N = 2$ topological string has a twisted $N = 2$ superconformal algebra

$$\begin{aligned}
\tilde{T}(z)\tilde{T}(\zeta) &= (z-\zeta)^{-2}2\tilde{T}(\zeta) + (z-\zeta)^{-1}\partial\tilde{T}(\zeta)\,, \\
\tilde{T}(z)G^+(\zeta) &= (z-\zeta)^{-2}G^+(\zeta) + (z-\zeta)^{-1}\partial G^+(\zeta)\,, \\
\tilde{T}(z)G^-(\zeta) &= (z-\zeta)^{-2}2G^-(\zeta) + (z-\zeta)^{-1}\partial G^-(\zeta)\,, \\
G^+(z)G^-(\zeta) &= (z-\zeta)^{-3}\frac{c}{3} + (z-\zeta)^{-2}J(\zeta) + (z-\zeta)^{-1}\tilde{T}(\zeta)\,, \\
G^+(z)G^+(\zeta) &= 0\,, && * \\
G^-(z)G^-(\zeta) &= 0\,, \\
\tilde{T}(z)J(\zeta) &= (z-\zeta)^{-3}(-\frac{c}{3}) + (z-\zeta)^{-2}J(\zeta) + (z-\zeta)^{-1}\partial J(\zeta)\,, \\
J(z)J(\zeta) &= (z-\zeta)^{-2}\frac{c}{3}\,, && * \\
J(z)G^+(\zeta) &= (z-\zeta)^{-1}G^+(\zeta)\,, && * \\
J(z)G^-(\zeta) &= -(z-\zeta)^{-1}G^-(\zeta)\,.
\end{aligned} \tag{56}$$

Equations (56) is related to the generators of the (untwisted) $N = 2$ superconformal algebra:

$$
\begin{aligned}
L(z)L(\zeta) &= (z-\zeta)^{-4}\frac{c}{2} + (z-\zeta)^{-2}2L(\zeta) + (z-\zeta)^{-1}\partial L(\zeta)\,,\\
L(z)G^{\pm}(\zeta) &= (z-\zeta)^{-2}\frac{2}{3}G^{\pm}(\zeta) + (z-\zeta)^{-1}\partial G^{\pm}(\zeta)\,,\\
G^{+}(z)G^{-}(\zeta) &= (z-\zeta)^{-3}\frac{c}{3} + (z-\zeta)^{-2}J(\zeta) + (z-\zeta)^{-1}(L(\zeta) + \frac{1}{2}\partial J(\zeta))\,,\\
G^{+}(z)G^{+}(\zeta) &= 0\,,\\
G^{-}(z)G^{-}(\zeta) &= 0\,,\\
L(z)J(\zeta) &= (z-\zeta)^{-2}J(\zeta) + (z-\zeta)^{-1}\partial J(\zeta)\,,\\
J(z)J(\zeta) &= (z-\zeta)^{-2}\frac{c}{3}\,,\\
J(z)G^{\pm}(\zeta) &= \pm(z-\zeta)^{-1}G^{\pm}(\zeta)\,,
\end{aligned}
\tag{57}
$$

where the generators differ only by the twisted Virasoro generator $\tilde{T}(z) \equiv L(z) + \frac{1}{2}\partial J(z)$. Since the OPE's (56) resemble somewhat those of the bosonic string (58), (*i.e.* only the starred equations (*) in (56) differ), we can define physical fields relative to Q_0 cohomology where $G^{+}(z) = \sum_n Q_n z^{-n-1}$. Physical fields correspond to chiral primary fields $\Phi^{+}(z)$ with ghost charge $+1$ and dimension 0 (they arise from operators that have ghost charge $+1$ and dimension $\frac{1}{2}$ before the algebra is twisted), so that $\{Q_0, \Phi^{+}(\zeta)\} = 0$.

Bosonic string theory can be viewed as a two-dimensional conformal field theory with certain additional features (we concentrate on left-movers and will denote right-movers with barred notation):

$$
\begin{aligned}
T_{\rm tot}(z)T_{\rm tot}(\zeta) &= (z-\zeta)^{-2}\,2T_{\rm tot}(\zeta) + (z-\zeta)^{-1}\partial T_{\rm tot}(\zeta)\,,\\
T_{\rm tot}(z)j_{\rm BRST}(\zeta) &= (z-\zeta)^{-2}j_{\rm BRST}(\zeta) + (z-\zeta)^{-1}\partial j_{\rm BRST}(\zeta)\,,\\
T_{\rm tot}(z)b(\zeta) &= (z-\zeta)^{-2}2b(\zeta) + (z-\zeta)^{-1}\partial b(\zeta)\,,\\
j_{\rm BRST}(z)b(\zeta) &= (z-\zeta)^{-3}\frac{(c=9)}{3} + (z-\zeta)^{-2}j_{\rm ghost}(\zeta) + (z-\zeta)^{-1}T_{\rm tot}(\zeta)\,,\\
j_{\rm BRST}(z)j_{\rm BRST}(\zeta) &= -j_{\rm BRST}(\zeta)j_{\rm BRST}(z) = 2\frac{\partial}{\partial\zeta}[(z-\zeta)^{-2}\,{}^{\times}_{\times}\,\partial c(\zeta)\,c(\zeta)\,{}^{\times}_{\times}] \neq 0
\end{aligned}
$$

$$
\begin{aligned}
b(z)b(\zeta) &= 0\,, \\
T_{\rm tot}(z)j_{\rm ghost}(\zeta) &= (z-\zeta)^{-3}\frac{(c=9)}{(-3)} + (z-\zeta)^{-2}j_{\rm ghost}(\zeta) + (z-\zeta)^{-1}\partial j_{\rm ghost}(\zeta) \\
j_{\rm ghost}(z)j_{\rm ghost}(\zeta) &= (z-\zeta)^{-2}\frac{(c=3)}{3}\,, \\
j_{\rm ghost}(z)j_{\rm BRST}(\zeta) &= (z-\zeta)^{-3}4c(\zeta) + (z-\zeta)^{-2}2\partial c(\zeta) + (z-\zeta)^{-1}j_{\rm BRST}(\zeta)\,, \\
j_{\rm ghost}(z)b(\zeta) &= -(z-\zeta)^{-1}b(\zeta)\,, \qquad (58)
\end{aligned}
$$

where the generators are

$$
\begin{aligned}
T_{\rm tot} = T_m^{N=0} + T_g^{N=0} &= T_m^{N=0} - 2\,{}^\times_\times b\partial c\,{}^\times_\times - {}^\times_\times \partial b\, c\,{}^\times_\times \\
j_{\rm BRST}(z) = cT_m + \tfrac{1}{2}\,{}^\times_\times cT_g\,{}^\times_\times + \tfrac{3}{2}\partial^2 c &= cT_m - {}^\times_\times cb\partial c\,{}^\times_\times + \tfrac{3}{2}\partial^2 c \\
&b(z) \\
j_{\rm ghost}(z) &= {}^\times_\times cb\,{}^\times_\times \qquad (59)
\end{aligned}
$$

and $c(z)b(\zeta) = -b(\zeta)c(z) = (z-\zeta)^{-1} + {}^\times_\times c(z)b(\zeta)\,{}^\times_\times$. The normal ordering has been defined putting the annihilation operators to the right of the creation operators, where $b_n|0\rangle^{bc} = 0$ for $n \geq -1$; $c_n|0\rangle^{bc} = 0$ for $n \geq 2$. The BRST charge is $Q \equiv \frac{1}{2\pi i}\oint dz j_{BRST}(z) = \sum_m c_m L_m^X - \frac{1}{2}\sum_{m,n}(m-n)\,{}^\times_\times c_{-m}c_{-n}b_{n+m}\,{}^\times_\times$. Then $Q\left(|0\rangle^{bc}\otimes|0\rangle_X\right) = 0$. The ghost charge is $J_0 \equiv \sum_n {}^\times_\times c_n b_{-n}\,{}^\times_\times$. Then $(c_1|0\rangle^{bc}\otimes|\phi\rangle_X)$ has ghost charge eigenvalue, *i.e.* ghost number, equal to one. The physical state conditions in the "old covariant" formalism are $(L_0^X - 1)|\phi\rangle_X = 0$, $L_n^X|\phi\rangle_X = 0$ for $n > 0$. Since $Q|\psi\rangle = 0$ implies $\left(c_0(L_0^X - 1) + \sum_n c_{-n}L_n^X\right)|\psi\rangle = 0$ when $|\psi\rangle = c_1|0\rangle^{bc}\otimes|\phi\rangle_X$, then in the BRST formalism the physical vertex operators are defined by ghost number one fields $\Phi^+(z)$ that obey $\{Q, \Phi^+(z)\} = 0$, *i.e.* the OPE of $j_{BRST}(z)\Phi^+(\zeta)$ has no single pole. So here, every physical state is in one-to-one correspondence with a primary field of the Virasoro algebra of dimension 0, *i.e.* $\Phi^+(z) = c(z)\phi_X(z)$. *Thus we have used the twisted $N=2$ superVirasoro algebra to define physical fields relative to Q_0 cohomology where $G^+(z) = \sum_n Q_n z^{-n-1}$. Physical states correspond to chiral primary fields $\Phi^+(z)$ with ghost charge $+1$ and dimension* 0.

For $N=2$ topological strings, $c=9$. $N=4$ topological strings are used when $c=6$. From an $N=2$ superconformal algebra with $c=6$, we construct a topological $N=4$ string by defining the remaining generators and twisting the $N=4$ superconformal algebra:

$$
\begin{aligned}
\tilde{T}(z)\tilde{T}(\zeta) &= (z-\zeta)^{-2}2\tilde{T}(\zeta) + (z-\zeta)^{-1}\partial\tilde{T}(\zeta)\,, \\
\tilde{T}(z)G^+(\zeta) &= (z-\zeta)^{-2}G^+(\zeta) + (z-\zeta)^{-1}\partial G^+(\zeta)\,,
\end{aligned}
$$

$$\tilde{T}(z)G^-(\zeta) = (z-\zeta)^{-2}2G^-(\zeta) + (z-\zeta)^{-1}\partial G^-(\zeta)\,,$$

$$G^+(z)G^-(\zeta) = (z-\zeta)^{-3}\frac{(c=6)}{3} + (z-\zeta)^{-2}J(\zeta) + (z-\zeta)^{-1}\tilde{T}(\zeta)\,,$$

$$G^+(z)G^+(\zeta) = 0\,,$$

$$G^-(z)G^-(\zeta) = 0\,,$$

$$\tilde{T}(z)J(\zeta) = (z-\zeta)^{-3}(-\frac{(c=6)}{3}) + (z-\zeta)^{-2}J(\zeta) + (z-\zeta)^{-1}\partial J(\zeta)\,,$$

$$J(z)J(\zeta) = (z-\zeta)^{-2}\frac{(c=6)}{3}\,,$$

$$J(z)G^\pm(\zeta) = \pm(z-\zeta)^{-1}G^\pm(\zeta)\,,$$

$$J(z)J^\pm(\zeta) = (z-\zeta)^{-1}(\pm 2)J^\pm(\zeta)\,,$$

$$J^+(z)J^-(\zeta) = -(z-\zeta)^{-2}\frac{(c=6)}{6} - (z-\zeta)^{-1}J(\zeta)\,,$$

$$\tilde{T}(z)J^+(\zeta) = (z-\zeta)^{-1}\partial J^+(\zeta)\,,$$

$$\tilde{T}(z)J^-(\zeta) = (z-\zeta)^{-2}2J^-(\zeta) + (z-\zeta)^{-1}\partial J^-(\zeta)\,,$$

$$J^+(z)G^-(\zeta) = -(z-\zeta)^{-1}\tilde{G}^+(\zeta)\,,$$

$$J^-(z)G^-(\zeta) = 0\,,$$

$$J^-(z)G^+(\zeta) = -(z-\zeta)^{-1}\tilde{G}^-(\zeta)\,,$$

$$J^+(z)G^+(\zeta) = 0\,,$$

$$J^-(z)\tilde{G}^+(\zeta) = (z-\zeta)^{-1}G^-(\zeta)\,,$$

$$J^-(z)\tilde{G}^-(\zeta) = 0\,,$$

$$J^+(z)\tilde{G}^-(\zeta) = (z-\zeta)^{-1}G^+(\zeta)\,,$$

$$J^+(z)\tilde{G}^+(\zeta) = 0\,,$$

$$G^-(z)\tilde{G}^+(\zeta) = 0\,,$$

$$G^+(z)\tilde{G}^-(\zeta) = 0\,,$$

$$\tilde{G}^+(z)\tilde{G}^-(\zeta) = (z-\zeta)^{-3}\frac{(c=6)}{3} + (z-\zeta)^{-2}J(\zeta) + (z-\zeta)^{-1}\tilde{T}(\zeta)\,,$$

$$G^+(z)\tilde{G}^+(\zeta) = (z-\zeta)^{-2}2J^+(\zeta) + (z-\zeta)^{-1}\partial J^+(\zeta)\,,$$

$$G^-(z)\tilde{G}^-(\zeta) = (z-\zeta)^{-2}2J^-(\zeta) + (z-\zeta)^{-1}\partial J^-(\zeta)\,,$$

$$\tilde{G}^+(z)\tilde{G}^+(\zeta) = 0\,,$$

$$\tilde{G}^-(z)\tilde{G}^-(\zeta) = 0\,,$$

$$\tilde{T}(z)\tilde{G}^+(\zeta) = (z-\zeta)^{-2}\tilde{G}^+(\zeta) + (z-\zeta)^{-1}\partial\tilde{G}^+(\zeta)\,,$$

$$\tilde{T}(z)\tilde{G}^-(\zeta) = (z-\zeta)^{-2}2\tilde{G}^-(\zeta) + (z-\zeta)^{-1}\partial\tilde{G}^-(\zeta)\,,$$
$$J(z)\tilde{G}^\pm(\zeta) = \pm(z-\zeta)^{-1}\tilde{G}^\pm(\zeta)\,. \tag{60}$$

Since the superstring can be written as an $N = 2$ super Virasoro algebra with $c = 6$, it's necessary to find additional generators making up an $N = 4$ topological string. In RNS variables they are

$$\begin{aligned}
\tilde{T}(z) &= T_m^{N=1} + T_g^{N=1} \\
&= T_m^{N=1} - 2 \,{}^\times_\times b\partial c\,{}^\times_\times - {}^\times_\times \partial b c\,{}^\times_\times - \tfrac{3}{2}\,{}^\times_\times \beta\partial\gamma\,{}^\times_\times - \frac{1}{2}\,{}^\times_\times \partial\beta\gamma\,{}^\times_\times \\
G^+(z) &= \gamma G_m + c(T_m - \tfrac{3}{2}\beta\partial\gamma - \tfrac{1}{2}\partial\beta\gamma - b\partial c) - \gamma^2 b + \partial^2 c + \partial(c\xi\eta) \\
G^-(z) &= b \qquad J(z) = cb + \eta\xi \qquad J^+(z) = c\eta \qquad J^-(z) = b\xi \\
\tilde{G}^+(z) &= \eta \\
\tilde{G}^-(z) &= b(ie^\phi G_m + \eta e^{2\phi}\partial b - c\partial\xi) \\
&\quad + \xi(T_m - \tfrac{3}{2}\beta\partial\gamma - \tfrac{1}{2}\gamma\partial\beta - 2b\partial c + c\partial b) + \partial^2\xi\,,
\end{aligned} \tag{61}$$

with $c(z)b(\zeta) = -b(\zeta)c(z) = (z-\zeta)^{-1} + {}^\times_\times c(z)b(\zeta)\,{}^\times_\times$. Also the super-reparametrization ghosts with $\gamma(z)\beta(\zeta) = \beta(\zeta)\gamma(z) = (z-\zeta)^{-1} + {}^\times_\times \gamma(z)\beta(\zeta)\,{}^\times_\times$ have been bosonized as ($\beta = ie^{-\phi}\partial\xi, \gamma = -i\eta e^\phi$) with $\xi(z)\eta(\zeta) = -\eta(\zeta)\xi(z) = (z-\zeta)^{-1} + {}^\times_\times \xi(z)\eta(\zeta)\,{}^\times_\times$, and $\phi(z)\phi(\zeta) = -\ln(z-\zeta) + {}^\times_\times \phi(z)\phi(\zeta)\,{}^\times_\times$ so that $e^{-\phi(z)}e^{\phi(\zeta)} =: e^{-\phi(z)+\phi(\zeta)} : (z-\zeta)$.

The generators (61) satisfy the ($c = 6$) twisted $N = 4$ superconformal algebra given in (60). Since the algebra is twisted, the Virasoro generators close with no anomaly (*i.e.* $c = 0$) but c still appears in the rest of the algebra, such as the anomaly of the $U(1)$ current J. For the IIB superstring we have both the holomorphic $N = 4$ superconformal algebra (60) and another anti-holomorphic one. The holomorphic generators, when specialized for IIB compactified to 6d, and rewritten in terms of BVW worldsheet variables which display manifest 6d spacetime supersymmetry and eschew spin fields, become

$$\begin{aligned}
T &= -\frac{1}{2}\partial x^m\partial x_m - p_a\partial\theta^a - \frac{1}{2}\partial\rho\partial\rho - \frac{1}{2}\partial\sigma\partial\sigma + \partial^2(\rho + i\sigma) + T_C \\
G^+ &= -e^{-2\rho-i\sigma}(p)^4 + \tfrac{i}{2}e^{-\rho}p_a p_b\partial x^{ab} \\
&\quad + e^{i\sigma}\left(-\frac{1}{2}\partial x^m\partial x_m - p_a\partial\theta^a - \frac{1}{2}\partial(\rho+i\sigma)\partial(\rho+i\sigma)\right. \\
&\quad \left. + \frac{1}{2}\partial^2(\rho+i\sigma)\right) + G_C^+
\end{aligned}$$

$$
\begin{aligned}
G^- &= e^{-i\sigma} + G_C^- \\
J &= \partial(\rho + i\sigma) + J_C \\
\tilde{G}^+ &= e^{iH_C}(-e^{-3\rho-2i\sigma}(p)^4 + \tfrac{i}{2}e^{-2\rho-i\sigma}p_a p_b \partial x^{ab} \\
&\quad + e^{-\rho}(-\frac{1}{2}\partial x^m \partial x_m - p_a \partial\theta^a - \frac{1}{2}\partial(\rho+i\sigma)\partial(\rho+i\sigma) \\
&\quad + \frac{1}{2}\partial^2(\rho+i\sigma)) + e^{-\rho-i\sigma}\tilde{G}_C^- \\
J^+ &= e^{\rho+i\sigma}J_C^+ \\
J^- &= e^{-\rho-i\sigma}J_C^- \,. \qquad (62)
\end{aligned}
$$

These currents are given in terms of the left-moving bosons $\partial x^m, \rho, \sigma$, and the left-moving fermionic worldsheet fields p^a, θ^a, where $1 \le m \le 6, 1 \le a \le 4$. The conformal weights of p_a, θ^a are 1 and 0, respectively. We define $p^4 \equiv \frac{1}{24}\epsilon^{abcd}p_a p_b p_c p_d = p_1 p_2 p_3 p_4$; and $\partial x^{ab} = \partial x^m \sigma_m^{ab}$ where $\sigma_m^{ab}\,\sigma_{nac} + \sigma_n^{ab}\,\sigma_{mac} = \eta_{mn}\,\delta_c^b$. Here lowered indices mean $\sigma_{mab} \equiv \frac{1}{2}\epsilon_{abcd}\sigma_m^{cd}$. Note that e^ρ and $e^{i\sigma}$ are worldsheet fermions. Also $e^{\rho+i\sigma} \equiv e^\rho e^{i\sigma} = -e^{i\sigma}e^\rho$. Here $J_C \equiv i\partial H_C, J_C^+ \equiv -e^{iH_C}, J_C^- \equiv e^{-iH_C}$. Both $\tilde{T}, G^\pm, J, J^\pm, \tilde{G}^\pm$ and the generators describing the compactification $\tilde{T}_C, G_C^\pm, J_C, J_C^\pm, \tilde{G}_C^\pm$ satisfy the twisted $N=4$, $c=6$, superconformal algebra (60), *i.e.* both $\tilde{T}$ and $\tilde{T}_C$ have $c=0$. However, as seen in (refeq:n4tw) and (56), c still appears in the twisted $N=4$ and $N=2$ algebras; and the N=2 generators in (62) $\tilde{T}, G^\pm, J$ decompose into a $c=0$ six-dimensional part and a $c=6$ compactification-dependent piece. (That is to say, the uncompactified piece of the twisted $N=2$ generators in (62) satisfies (56) with $c=0$, not just for the Virasoro generator but also whereever c appears in (56).

The other non-vanishing OPE's are $x^m(z,\bar{z})x^n(\zeta,\bar{\zeta}) = -\eta^{mn}\ln|z-\zeta|$; for the left-moving worldsheet fermion fields $p_a(z)\theta^b(\zeta) = (z-\zeta)^{-1}\delta_a^b$; and for the left-moving worldsheet bosons $\rho(z)\rho(\zeta) = -\ln(z-\zeta)$; $\sigma(z)\sigma(\zeta) = -\ln(z-\zeta)$. Right-movers are denoted by barred notation and have similar OPE's.

Both holomorphic and anti-holomorphic sets of generators are used to implement the physical state conditions on the vertex operators, a procedure[10,11,12] which results in a set of string constraint equations for flat spacetime. The notation $O_n\Phi$ denotes the pole of order $d+n$ in the OPE of O with Φ, when O is a dimension d operator. For the generators (62) since G^+ and $\tilde{G}^+$ are dimension one, and nilpotent, in analogy with the bosonic string, the physical $N=4$ topological vertex operators $\Phi^+(z)$ are defined by the conditions:

$$G_0^+\Phi^+ = 0\,; \qquad \tilde{G}_0^+\Phi^+ = 0\,; \qquad (J_0 - 1)\Phi^+ = 0\,. \qquad (63)$$

These are the physical conditions for a BRST-invariant vertex operator in the standard RNS formalism for the superstring. Since $\tilde{G}_0^+ = \eta_0$, the $\tilde{G}_0^+$ cohomology is trivial, *i.e.* $\tilde{G}_0^+\Phi^+ = 0$ implies $\Phi^+ = \tilde{G}_0^+ V$. So it is always possible to define a V satisfying

$$\Phi^+ = \tilde{G}_0^+ V\,; \qquad G_0^+\tilde{G}_0^+ V = J_0 V = 0\,. \tag{64}$$

Note that $\Phi^{++}(z)$ is a worldsheet fermion and $V(z)$ is a worldsheet boson. To describe the massless compactification independent states we introduce the $U(1)$-neutral vertex operator $V(z)$, but it is straightforward to go from $V(z)$ to the $U(1)$ charge equal to one vertex operator $\Phi^(z)$, using the relationship described above. In addition to (64), one can use the gauge invariance $V \sim V + G_0^+\Lambda + \tilde{G}_0^+\tilde{\Lambda}$ to further require

$$G_0^- V = \tilde{G}_0^- V = T_0 V = 0\,. \tag{65}$$

In this gauge the physical $U(1)$-charged vertex operators $\Phi^{++}(z)$ of the *closed* $N = 4$ toplogical string must satisfy

$$\begin{aligned} G_0^+\Phi^{++} &= \tilde{G}_0^+\Phi^{++} = \bar{G}_0^+\Phi^{++} = \bar{\tilde{G}}_0^+\Phi^{++} = 0\,,\\ G_0^-\Phi^{++} &= \tilde{G}_0^-\Phi^{++} = T_0\Phi^{++} = (J_0-1)\,\Phi^{++} = 0\,,\\ \bar{G}_0^-\Phi^{++} &= \bar{\tilde{G}}_0^-\Phi^{++} = \bar{T}_0\Phi^{++} = (\bar{J}_0-1)\,\Phi^{++} = 0\,. \end{aligned} \tag{66}$$

Similarly the $U(1)$-neutral vertex operator V defined by $\Phi^{++} = \tilde{G}_0^+\,\bar{\tilde{G}}_0^+ V$ must satisfy the conditions

$$\begin{aligned} G_0^+\tilde{G}_0^+ V &= \bar{G}_0^+\bar{\tilde{G}}_0^+ V = 0\,,\\ G_0^- V = \tilde{G}_0^- V = \bar{G}_0^- V &= \bar{\tilde{G}}_0^- V = T_0 V = \bar{T}_0 V = J_0 V = \bar{J}_0 V = 0\,. \end{aligned} \tag{67}$$

The integrated form of the *closed* superstring vertex operator $\Phi^{++}(z)$ is $\int d^2z G_{-1}^-\Phi^{++}$. In terms of V it will be defined as

$$U = \int d^2z G_{-1}^-\bar{G}_{-1}^- G_0^+\bar{G}_0^+ V\,. \tag{68}$$

The $N = 4$ topological prescription[12] for calculating superstring tree-level amplitudes is

$$< V_1(z_1)\,(\tilde{G}_0^+ V_2(z_2))(G_0^+ V_3(z_3))\prod_{r=4}^{n}\int dz_r G_{-1}^- G_0^+ V(z_r) > \,. \tag{69}$$

Note that since V are $U(1)$-neutral, the amplitude (69) has operators with a total $U(1)$ charge equal to 2. This is related to the RNS requirement that non-vanishing tree scattering amplitudes must have total superconformal ghost charge -2 and total conformal ghost charge 3. The *closed* $N = 4$ topological tree-level amplitudes are given by

$$< V_1(z_1,\bar{z}_1)(\tilde{G}_0^+\bar{\tilde{G}}_0^+ V_2(z_2,\bar{z}_2))(G_0^+\bar{G}_0^+ V_3(z_3,\bar{z}_3)) \\ \times \prod_{r=4}^n \int d^2 z_r G_{-1}^- \bar{G}_{-1}^- G_0^+ \bar{G}_0^+ V(z_r,\bar{z}_r) > \qquad (70)$$

4.2. *String Constraint Equations*

Using (67) on the general massless vertex operator

$$V = \sum_{m,n=-\infty}^{\infty} e^{m(i\sigma+\rho)+n(i\bar{\sigma}+\bar{\rho})} V_{m,n}(x,\theta,\bar{\theta}), \qquad (71)$$

we find in flat spacetime the constraints from the left and right-moving worldsheet super Virasoro algebras to be

$$\begin{aligned} (\nabla)^4 V_{1,n} &= \nabla_a \nabla_b \partial^{ab} V_{1,n} = 0 \\ \tfrac{1}{6}\epsilon^{abcd}\nabla_b \nabla_c \nabla_d V_{1,n} &= -i\nabla_b \partial^{ab} V_{0,n} \\ \nabla_a \nabla_b V_{0,n} - \tfrac{i}{2}\epsilon_{abcd}\partial^{cd} V_{-1,n} &= 0\,, \qquad \nabla_a V_{-1,n} = 0\,; \end{aligned} \qquad (72)$$

$$\begin{aligned} \bar{\nabla}^4 V_{n,1} &= \bar{\nabla}_{\bar{a}} \bar{\nabla}_{\bar{b}} \bar{\partial}^{\bar{a}\bar{b}} V_{n,1} = 0 \\ \tfrac{1}{6}\epsilon^{\bar{a}\bar{b}\bar{c}\bar{d}}\bar{\nabla}_{\bar{b}} \bar{\nabla}_{\bar{c}} \bar{\nabla}_{\bar{d}} V_{n,1} &= -i\bar{\nabla}_{\bar{b}} \bar{\partial}^{\bar{a}\bar{b}} V_{n,0} \\ \bar{\nabla}_{\bar{a}} \bar{\nabla}_{\bar{b}} V_{n,0} - \tfrac{i}{2}\bar{\epsilon}_{\bar{a}\bar{b}\bar{c}\bar{d}}\bar{\partial}^{\bar{c}\bar{d}} V_{n,-1} &= 0\,, \qquad \bar{\nabla}_{\bar{a}} V_{n,-1} = 0 \end{aligned} \qquad (73)$$

$$\partial^p \partial_p V_{m,n} = 0 \qquad (74)$$

for $-1 \le m,n \le 1$, with the notation $\nabla_a = d/d\theta^a$, $\bar{\nabla}_{\bar{a}} = d/d\bar{\theta}^{\bar{a}}$, $\partial^{ab} = -\sigma^{pab}\,\partial_p$. These conditions further imply $V_{m,n} = 0$ for $m > 1$ or $n > 1$ or $m < 1$ or $n < 1$, leaving nine non-zero components. In fact, the independent degrees of freedom can be shown to reside in V_{11}, and the surviving constraints yield (44).

In $AdS_3 \times S^3$ space, we generalize[13] the flat space string constraint equations (72-74) as follows:

$$\begin{aligned} F^4 V_{1,n} &= F_a F_b K^{ab} V_{1,n} = 0 \\ \tfrac{1}{6}\epsilon^{abcd} F_b F_c F_d V_{1,n} &= -iF_b K^{ab} V_{0,n} + 2iF^a V_{0,n} - E^a V_{-1,n} \\ F_a F_b V_{0,n} - \tfrac{i}{2}\epsilon_{abcd} K^{cd} V_{-1,n} &= 0\,, \qquad F_a V_{-1,n} = 0\,; \end{aligned} \qquad (75)$$

$$\begin{aligned} \bar{F}^4 V_{n,1} &= \bar{F}_{\bar{a}} \bar{F}_{\bar{b}} \bar{K}^{\bar{a}\bar{b}} V_{n,1} = 0 \\ \tfrac{1}{6}\epsilon^{\bar{a}\bar{b}\bar{c}\bar{d}} \bar{F}_{\bar{b}} \bar{F}_{\bar{c}} \bar{F}_{\bar{d}} V_{n,1} &= -i\bar{F}_{\bar{b}} \bar{K}^{\bar{a}\bar{b}} V_{n,0} + 2i\bar{F}^{\bar{a}} V_{n,0} - \bar{E}^{\bar{a}} V_{n,-1} \\ \bar{F}_{\bar{a}} \bar{F}_{\bar{b}} V_{n,0} - \tfrac{i}{2}\bar{\epsilon}_{\bar{a}\bar{b}\bar{c}\bar{d}} \bar{K}^{\bar{c}\bar{d}} V_{n,-1} &= 0\,, \qquad \bar{F}_{\bar{a}} V_{n,-1} = 0. \end{aligned} \qquad (76)$$

There is also a spin zero condition constructed from the Laplacian

$$(\,F_a\,E_a\,+\tfrac{1}{8}\epsilon_{abcd}\,K^{ab}\,K^{cd}\,)\,V_{n,m}=(\,\bar{F}_{\bar{a}}\,\bar{E}_{\bar{a}}\,+\tfrac{1}{8}\bar{\epsilon}_{\bar{a}\bar{b}\bar{c}\bar{d}}\,\bar{K}^{\bar{a}\bar{b}}\,K^{\bar{c}\bar{d}}\,)\,V_{n,m}=0\,. \tag{77}$$

We derived the curved space equations (75-77) by deforming the flat space equations by requiring invariance under the $PSU(2|2)$ transformations (78) that replace the $d=6$ super Poincare transformations of flat space. The Lie algebra of the supergroup $PSU(2|2)$ contains six even elements $K_{ab}\in SO(4)$ and eight odd E_a, F_a. They generate the infinitesimal symmetry transformations of the constraint equations:

$$\begin{aligned}
\Delta_a^-\,V_{m,n}&=F_a\,V_{m,n}\,,\\
\Delta_{ab}\,V_{m,n}&=K_{ab}\,V_{m,n}\,,\\
\Delta_a^+\,V_{1,n}&=E_a\,V_{1,n}\,,\\
\Delta_a^+\,V_{0,n}&=E_a\,V_{0,n}+iF_aV_{1,n}\,,\\
\Delta_a^+\,V_{-1,n}&=E_a\,V_{-1,n}-iF_aV_{0,n}\,.
\end{aligned} \tag{78}$$

We write E_a, F_a, and K_{ab} for the operators that represent the left action of e_a, f_a, and t_{ab} on g. In the above coordinates,

$$\begin{aligned}
F_a&=\frac{d}{d\theta^a}\,,\qquad K_{ab}=-\theta_a\frac{d}{d\theta^b}+\theta_b\frac{d}{d\theta^a}+t_{Lab}\\
E_a&=\tfrac{1}{2}\epsilon_{abcd}\,\theta^b\,(t_L^{cd}-\theta^c\frac{d}{d\theta_d})+h_{a\bar{b}}\frac{d}{d\bar{\theta}_{\bar{b}}}\,,
\end{aligned} \tag{79}$$

where we have introduced an operator t_L that generates the left action of $SU(2)\times SU(2)$ on h alone, without acting on the θ's. Here

$$g=g(x,\theta\,,\bar{\theta})\;=e^{\theta^a f_a}\,e^{\frac{1}{2}\sigma^{pcd}x_p t_{cd}}\,e^{\bar{\theta}^{\bar{a}}e_{\bar{a}}}=e^{\theta^a f_a}\,h(x)\,e^{\bar{\theta}^{\bar{a}}e_{\bar{a}}}\,, \tag{80}$$

$$t_{Lab}\,g\;=\;e^{\theta^a f_a}\,(-t_{ab})\,h(x)\,e^{\bar{\theta}^{\bar{a}}e_{\bar{a}}}\,, \tag{81}$$

and we found (79) by requiring $F_a g\;=\;f_a g$, $E_a g\;=\;e_a g$, $K_{ab}g\;=\;-t_{ab}g$. Similar expressions hold for the right-acting generators $\bar{K}_{\bar{a}\bar{b}}$, $\bar{E}_{\bar{a}}$, and $\bar{F}_{\bar{a}}$.

The operators $t_L^{ab}, t_R^{\bar{a}\bar{b}}$ describe invariant derivatives on the $SO(4)$ group manifold. These can be related to covariant derivatives $\mathcal{T}_L^{cd}\equiv-\sigma^{p\,cd}\,D_p\,,\,\mathcal{T}_R^{\bar{c}\bar{d}}\equiv\sigma^{p\,\bar{c}\bar{d}}\,D_p$, where for example, acting on a function, $\mathcal{T}_L=t_L$ and $\mathcal{T}_R=t_R$. But when acting on fields that carry vector or spinor indices, they differ so that for example on spinor indices $t_L^{ab}V_e=\mathcal{T}_L^{ab}\,V_e+\frac{1}{2}\delta_e^a\,\delta^{bc}V_c-\frac{1}{2}\delta_e^b\,\delta^{ac}V_c\,$.

In fact, for the Type IIB superstring on $AdS_3\times S^3\times K3$ with background Ramond flux, a sigma model[12] with conventional local interactions (no spin fields in the action) was found using the supergroup $PSU(2|2)$ as target, coupled to ghost fields ρ and σ. The spacetime symmetry group is $PSU(2|2) \times PSU(2|2)$, acting by left and right multiplication on the group manifold, *i.e.* by $g \to agb^{-1}$ where g is a $PSU(2|2)$-valued field, and $a, b \in PSU(2|2)$ are the symmetry group's Lie algebra elements. The supergroup is generated by the super Lie algebra with 12 bosonic generators forming a subalgebra $SO(4)^2$ together with 16 odd generators. Hence our model has non-maximal supersymmetry with 16 supercharges.

The $PSU(2|2)$-valued field g is given in terms of x, θ, and $\bar{\theta}$, which are identified as coordinates on the supergroup manifold. In addition, the Type IIB on $AdS_3 \times S^3 \times M$ has worldsheet variables describing the compactification degrees of freedom on the four-dimensional space M. The vertex operators $V_{mn}(x, \theta, \bar{\theta})$ are examples of the field g.

To interpret the generators E_a, F_a, K_{ab}, we recall that in flat space, the $d = 6$ supersymmetry algebra for the left-movers is given by

$$\begin{aligned} \{q_a^+, q_c^-\} &= \frac{1}{2}\epsilon_{abcd}P^{cd} \\ [P_{ab}, P_{cd}] &= 0 = [P_{ab}, q_c^\pm] = \{q_a^+, q_b^+\} = \{q_a^-, q_b^-\} \end{aligned} \tag{82}$$

where $P_{ab} \equiv \delta_{ac}\delta_{bd}P^{cd}$ and

$$\begin{aligned} q_a^- &= \oint F_a(z) \\ q_a^+ &= \oint (e^{-\rho - i\sigma}F_a(z) + iE_a(z)) \\ P^{ab} &= \oint \partial x_m(z)\sigma^{mab} . \end{aligned} \tag{83}$$

In flat space we have $F_a(z) = p_a(z)$ and $E_a(z) = \frac{1}{2}\epsilon_{abcd}\theta^b(z)\partial x_m(z)\sigma^{mcd}$. We distinguish between the currents and their zero moments E_a, F_a which together with P_{ab} also generate the flat space supersymmetry algebra

$$\begin{aligned} &[P_{ab}, P_{cd}] = 0 = [P_{ab}, F_c] = [P_{ab}, E_c] , \\ &\{E_a, F_b\} = \frac{1}{2}\epsilon_{abcd}P^{cd} , \qquad \{E_a, E_b\} = \{F_a, F_b\} = 0 . \end{aligned} \tag{84}$$

On $AdS_3 \times S^3$, the Poincare supersymmetry algebra (84) is replaced by the $PSU(2|2)$ superalgebra

$$\begin{aligned} [K_{ab}, K_{cd}] &= \delta_{ac}K_{bd} - \delta_{ad}K_{bc} - \delta_{bc}K_{ad} + \delta_{bd}K_{ac} \\ [K_{ab}, E_c] &= \delta_{ac}E_b - \delta_{bc}E_a , \\ [K_{ab}, F_c] &= \delta_{ac}F_b - \delta_{bc}F_a , \end{aligned}$$

$$\{E_a, F_b\} = \frac{1}{2} e_{abcd} K^{cd} \,,$$
$$\{E_a, E_b\} = 0 = \{F_a, F_b\} \tag{85}$$

The generators $q_a^{\pm}$, which generate the AdS tranformations (78), still have a form similar to (83) but $E_a(z,\bar{z}), F_a(z,\bar{z})$ are no longer holomorphic and their zero moments with respect to z satisfy (85).

For the bosonic field components of the vertex operator the AdS constraint equations (75-77) result in

$$\Box h^g{}_{\bar{a}}\, V_{ag}^{--} = -4\, \sigma^m_{ab}\, \sigma^n_{gh}\, \delta^{bh}\, h^g{}_{\bar{a}}\, G_{mn} \tag{86}$$

$$\Box h^g{}_{\bar{a}}\, h^h{}_{\bar{b}}\, \sigma^m_{ab}\, \sigma^n_{gh}\, G_{mn} = \tfrac{1}{4} \epsilon_{abce} \epsilon_{fghk}\, \delta^{ch}\, h^f{}_{\bar{a}}\, h^g{}_{\bar{b}}\; F^{++ek} \tag{87}$$

$$\Box h_g{}^{\bar{a}}\, F^{++ag} = 0\,, \quad \Box h_g{}^{\bar{a}}\, A_a^{-+g} = 0\,, \quad \Box h^g{}_{\bar{a}}\, A_g^{+-a} = 0 \tag{88}$$

$$\epsilon_{eacd}\, t_L^{cd}\, h^b{}_{\bar{a}}\, A_b^{+-a} = 0\,, \qquad \epsilon_{\bar{e}\bar{b}\bar{c}\bar{d}}\, t_R^{\bar{c}\bar{d}}\, h_a{}^{\bar{a}}\, A_{\bar{a}}^{-+\bar{b}} = 0 \tag{89}$$

$$\epsilon_{eacd}\, t_L^{cd}\, h_b{}^{\bar{a}}\, F^{++ab} = 0\,, \qquad \epsilon_{\bar{e}\bar{b}\bar{c}\bar{d}}\, t_R^{\bar{c}\bar{d}}\, h^a{}_{\bar{a}}\, F^{++\bar{a}\bar{b}} = 0 \tag{90}$$

$$t_L^{ab}\, h^g{}_{\bar{a}}\, h^h{}_{\bar{b}}\, \sigma^m_{ab}\, \sigma^n_{gh}\, G_{mn} = 0\,, \quad t_R^{\bar{a}\bar{b}}\, h_a^{\bar{g}}\, h_b^{\bar{h}}\, \sigma^m_{\bar{g}\bar{h}}\, \sigma^n_{\bar{a}\bar{b}}\, G_{mn} = 0\,. \tag{91}$$

We have expanded $G_{mn} = g_{mn} + b_{mn} + \bar{g}_{mn}\phi$. The $SO(4)$ Laplacian is $\Box \equiv \frac{1}{8} \epsilon_{abcd}\, t_L^{ab}\, t_L^{cd} = \frac{1}{8} \epsilon_{\bar{a}\bar{b}\bar{c}\bar{d}}\, t_R^{\bar{a}\bar{b}}\, t_R^{\bar{c}\bar{d}}$. In order to compare this with the supergravity field theory, we can use our expressions for the group manifold invarinat derivatives terms of covariant derivatives. We will also use the fact that on $AdS_3 \times S^3$ we can write the Riemann tensor and the metric tensor as

$$\bar{R}_{mnp\tau} = \tfrac{1}{4} \left(\bar{g}_{m\tau} \bar{R}_{np} + \bar{g}_{np} \bar{R}_{m\tau} - \bar{g}_{n\tau} \bar{R}_{mp} - \bar{g}_{mp} \bar{R}_{n\tau} \right)$$
$$\bar{g}_{mn} = \tfrac{1}{2}\, \sigma_m^{ab}\, \sigma_{n\,ab}\,. \tag{92}$$

The sigma matrices σ^{mab} satisfy the algebra $\sigma^{mab}\sigma^n_{ac} + \sigma^{nab}\sigma^m_{ac} = \eta^{mn}\delta^b_c$ in flat space, where η^{mn} is the six-dimensional Minkowski metric. Sigma matrices with lowered indices are defined by $\sigma^m_{ab} = \frac{1}{2}\epsilon_{abcd}\sigma^{mcd}$, although for other quantities indices are raised and lowered with δ^{ab}, so we distinguish σ^m_{ab} from $\delta_{ac}\,\delta_{bd}\,\sigma^{mcd}$. In curved space, η_{mn} is replaced by the $AdS_3 \times S^3$ metric $\bar{g}_{mn}$. We then find from the string constraints that the six-dimensional string field components g_{mn}, b_{mn}, ϕ, etc. satisfy (46,47).

4.3. *Correlation Functions*

We will review[17] the six-dimensional three-graviton tree level amplitude in (6d) flat space, for Type IIB superstrings on $\mathbf{R}^6 \times K3$ in the BVW formalism. It is contained in the closed string three-point function

$$< V(z_1,\bar{z}_1)\,(G_0^+\bar{G}_0^+ V(z_2,\bar{z}_2))(\tilde{G}_0^+\bar{\tilde{G}}_0^+ V(z_3,\bar{z}_3)) > \tag{93}$$

where the vertex operators are given by

$$\begin{aligned} V(z,\bar{z}) = e^{i\sigma(z)+\rho(z)}\, e^{i\bar{\sigma}(\bar{z})+\bar{\rho}(\bar{z})}\, \theta^a(z)\theta^b(z)\,\bar{\theta}^{\bar{a}}(\bar{z}) \\ \times \bar{\theta}^{\bar{b}}(\bar{z})\,\sigma^m_{ab}\,\sigma^n_{\bar{a}\bar{b}}\,\phi_{mn}(X(z,\bar{z}))\,, \end{aligned} \tag{94}$$

when the field

$$\phi_{mn} = g_{mn} + b_{mn} + \bar{g}_{mn}\,\phi$$

satisfies the constraints we found previously $\partial^m\phi_{mn} = 0$, and $\Box\phi_{mn} = 0$. These constraints imply the gauge conditions $\partial^m b_{mn} = 0$ for the two-form, and $\partial^m g_{mn} = -\partial_n\phi$ for the traceless graviton g_{mn} and dilaton ϕ. There is a residual gauge symmetry

$$g_{mn} \to g_{mn} + \partial_m\xi_n + \partial_n\xi_m\,, \qquad \phi \to \phi\,, \qquad b_{mn} \to b_{mn} \tag{95}$$

with $\Box\xi_n = 0$, $\partial\cdot\xi = 0$. To evaluate (93), we extract the simple poles as

$$\begin{aligned} G_0^+\bar{G}_0^+\,V(z,\bar{z}) = e^{i\sigma}\,e^{i\bar{\sigma}}\,(-4)\,[\phi_{mn}(X)\,\partial X^m\bar{\partial}X^n \\ - p_a\theta^b\,\sigma^m_{cb}\,\sigma^{pca}\bar{\partial}X^n\,\partial_p\,\phi_{mn}(X) \\ - \bar{p}_{\bar{a}}\bar{\theta}^{\bar{b}}\,\sigma^n_{\bar{c}\bar{b}}\,\sigma^{p\bar{c}\bar{a}}\partial X^m\,\partial_p\phi_{mn}(X) \\ + p_a\theta^b\,\bar{p}_{\bar{a}}\bar{\theta}^{\bar{b}}\,\sigma^m_{cb}\,\sigma^{pca}\,\sigma^n_{\bar{c}\bar{b}}\,\sigma^{q\bar{c}\bar{a}}\,\partial_p\partial_q\,\phi_{mn}(X)\,] \end{aligned} \tag{96}$$

$$\tilde{G}_0^+\bar{\tilde{G}}_0^+\,V(z,\bar{z}) = e^{iH_C+2\rho+i\sigma}\,e^{i\bar{H}_C+2\bar{\rho}+i\bar{\sigma}}\,\theta^a\theta^b\,\bar{\theta}^{\bar{a}}\,\bar{\theta}^{\bar{b}}\,\sigma^m_{ab}\,\sigma^n_{\bar{a}\bar{b}}\,\phi_{mn}(X)\,. \tag{97}$$

Using the OPE's for the ghost fields and H_C, we partially compute (93) by evaluating the leading singularities to find

$$\begin{aligned} &< V_1(z_1,\bar{z}_1)\,(G_0^+\bar{G}_0^+ V_2(z_2,\bar{z}_2))(\tilde{G}_0^+\bar{\tilde{G}}_0^+ V_3(z_3,\bar{z}_3)) > \\ &\quad = (z_1-z_2)(z_2-z_3)(z_1-z_3)^{-1}(\bar{z}_1-\bar{z}_2)(\bar{z}_2-\bar{z}_3)(\bar{z}_1-\bar{z}_3)^{-1} \\ &\quad\quad \times 4 < e^{iH_C(z_3)}e^{\rho(z_1)+2\rho(z_3)}e^{i\sigma(z_1)+i\sigma(z_2)+i\sigma(z_3)} \\ &\quad\quad \times e^{i\bar{H}_C(\bar{z}_3)}e^{\bar{\rho}(\bar{z}_1)+2\bar{\rho}(\bar{z}_3)}e^{i\bar{\sigma}(\bar{z}_1)+i\bar{\sigma}(\bar{z}_2)+i\bar{\sigma}(\bar{z}_3)} \\ &\quad\quad \times \theta^a(z_1)\theta^b(z_1)\bar{\theta}^{\bar{a}}(\bar{z}_1)\bar{\theta}^{\bar{b}}(\bar{z}_1)\,\sigma^m_{ab}\,\sigma^n_{\bar{a}\bar{b}}\,\phi_{mn}(X(z_1,\bar{z}_1)) \\ &\quad\quad \times [\,\phi_{jk}(X(z_2,\bar{z}_2))\,\partial X^j(z_2)\bar{\partial}X^k(\bar{z}_2) \end{aligned}$$

$$
\begin{aligned}
&- p_e(z_2)\theta^f(z_2)\,\sigma^j_{uf}\,\sigma^{pue}\bar\partial X^k(\bar z_2)\,\partial_p\,\phi_{jk}(X(z_2,\bar z_2))\\
&- \bar p_{\bar e}(\bar z_2)\bar\theta^{\bar f}(\bar z_2)\,\sigma^k_{\bar u\bar f}\,\sigma^{p\bar u\bar e}\partial X^j(z_2)\,\partial_p\phi_{jk}(X(z_2,\bar z_2))\\
&+ p_e(z_2)\theta^f(z_2)\,\bar p_{\bar e}(\bar z_2)\bar\theta^{\bar f}(\bar z_2)\,\sigma^j_{uf}\,\sigma^{pue}\,\sigma^k_{\bar u\bar f}\,\sigma^{q\bar u\bar e}\,\partial_p\partial_q\,\phi_{jk}(X(z_2,\bar z_2))\,]\\
&\times\theta^c(z_3)\theta^d(z_3)\bar\theta^{\bar c}(\bar z_3)\bar\theta^{\bar d}(\bar z_3)\,\sigma^g_{cd}\,\sigma^h_{\bar c\bar d}\,\phi_{gh}(X(z_3,\bar z_3))\ > \ . \qquad (98)
\end{aligned}
$$

Evaluating the remaining z_2, z_3 operators products, and using the $SL(2,C)$ invariance of the amplitude to take the three points to constants $z_1 \to \infty$, $\bar z_1 \to \infty$, $z_2 \to 1, \bar z_2 \to 1$, $z_3 \to 0, \bar z_3 \to 0$, we find

$$
\begin{aligned}
&< V_1(z_1,\bar z_1)\,(G_0^+\bar G_0^+ V_2(z_2,\bar z_2))(\tilde G_0^+\bar{\tilde G}_0^+ V_3(z_3,\bar z_3))\ > \qquad (99)\\
&\quad = (z_2-z_3)(\bar z_2-\bar z_3)(z_2-z_3)^{-1}(\bar z_2-\bar z_3)^{-1}\,\cdot\\
&\quad\times 4 < e^{iH_C(0)+3\rho(0)+3i\sigma(0)}e^{i\bar H_C(0)+3\bar\rho(0)+3i\bar\sigma(0)}\theta_0^a\theta_0^b\theta_0^c\theta_0^d\,\bar\theta_0^{\bar a}\bar\theta_0^{\bar b}\bar\theta_0^{\bar c}\bar\theta_0^{\bar d}\ >\\
&\quad\times[\sigma^m_{ab}\,\sigma^g_{cd}\,\sigma^n_{\bar a\bar b}\,\sigma^h_{\bar c\bar d}\ < \phi_{mn}(X(\infty))\,\phi_{jk}(X(1))\,\partial^j\partial^k\phi_{gh}(X(0)) >\\
&\quad + 2\sigma^m_{ab}\,(\sigma^j\sigma^p\sigma^g)_{cd}\,\sigma^n_{\bar a\bar b}\,\sigma^h_{\bar c\bar d}\ < \phi_{mn}(X(\infty))\,\partial_p\phi_{jk}(X(1))\,\partial^k\phi_{gh}(X(0)) >\\
&\quad + 2\sigma^m_{ab}\,\sigma^g_{cd}\,\sigma^n_{\bar a\bar b}\,(\sigma^k\sigma^p\sigma^h)_{\bar c\bar d}\ < \phi_{mn}(X(\infty))\,\partial_p\phi_{jk}(X(1))\,\partial^j\phi_{gh}(X(0)) >\\
&\quad + 4\sigma^m_{ab}\,(\sigma^j\sigma^p\sigma^g)_{cd}\,\sigma^n_{\bar a\bar b}\,(\sigma^k\sigma^q\sigma^h)_{\bar c\bar d}\ <\phi_{mn}(X(\infty))\,\partial_p\partial_q\phi_{jk}(X(1))\\
&\quad \times\phi_{gh}(X(0))>]
\end{aligned}
$$

which results in

$$
\begin{aligned}
&= 4\,[\ \bar g^{mg}\bar g^{nh}\ < \phi_{mn}(x_0)\,\phi_{jk}(x_0)\,\partial^j\partial^k\phi_{gh}(x_0) > \qquad (100)\\
&\quad - \bar g^{nh}\,(\sigma^m\sigma^j\sigma^p\sigma^g)^d_{\ d}\ < \phi_{mn}(x_0)\,\partial_p\phi_{jk}(x_0)\,\partial^k\phi_{gh}(x_0) >\\
&\quad - \bar g^{mg}\,(\sigma^n\sigma^k\sigma^p\sigma^h)^{\bar d}_{\ \bar d}\ < \phi_{mn}(x_0)\,\partial_p\phi_{jk}(x_0)\,\partial^j\phi_{gh}(x_0) >\\
&\quad + (\sigma^m\sigma^j\sigma^p\sigma^g)^d_{\ d}(\sigma^n\sigma^k\sigma^q\sigma^h)^{\bar d}_{\ \bar d}\ < \phi_{mn}(x_0)\,\partial_p\partial_q\phi_{jk}(x_0)\,\phi_{gh}(x_0) >]\,.
\end{aligned}
$$

The second equality follows from the vacuum expectation value of the ghost fields, H_C and eight fermion zero modes

$$
\begin{aligned}
&< e^{iH_C(0)+3\rho(0)+3i\sigma(0)}e^{i\bar H_C(0)+3\bar\rho(0)+3i\bar\sigma(0)}\theta_0^a\theta_0^b\theta_0^c\theta_0^d\,\bar\theta_0^{\bar a}\bar\theta_0^{\bar b}\bar\theta_0^{\bar c}\bar\theta_0^{\bar d} >\\
&\qquad = \tfrac{1}{16}\epsilon^{abcd}\,\epsilon^{\bar a\bar b\bar c\bar d}. \qquad (101)
\end{aligned}
$$

We have also used various sigma matrix identities. Since $(\sigma^m\sigma^n\sigma^p\sigma^q)^d_{\ d} = \bar g^{mn}\bar g^{pq} + \bar g^{mq}\bar g^{np} - \bar g^{mp}\bar g^{nq}$ where in flat space $\bar g_{mn} = \eta_{mn}$, and using the gauge condition $\partial^m \phi_{mn} = 0$ once more, we find

$$
\begin{aligned}
&< V_1(z_1,\bar{z}_1)\,(G_0^+\bar{G}_0^+ V_2(z_2,\bar{z}_2))(\tilde{G}_0^+\bar{\tilde{G}}_0^+ V_3(z_3,\bar{z}_3)) > \qquad (102)\\
&\quad = 4\,[\,\bar{g}^{mg}\bar{g}^{nh} \; < \phi_{mn}(x_0)\,\phi_{jk}(x_0)\,\partial^j\partial^k\phi_{gh}(x_0) >\\
&\qquad - \bar{g}^{nh}(\bar{g}^{mj}\bar{g}^{pg} - \bar{g}^{mp}\bar{g}^{jg}) \; < \phi_{mn}(x_0)\,\partial_p\phi_{jk}(x_0)\,\partial^k\phi_{gh}(x_0) >\\
&\qquad - \bar{g}^{mg}\,(\bar{g}^{nk}\bar{g}^{ph} - \bar{g}^{np}\bar{g}^{kh}) \; < \phi_{mn}(x_0)\,\partial_p\phi_{jk}(x_0)\,\partial^j\phi_{gh}(x_0) >\\
&\qquad + (\bar{g}^{mj}\bar{g}^{pg} - \bar{g}^{mp}\bar{g}^{jg})\,(\bar{g}^{nk}\bar{g}^{qh} - \bar{g}^{nq}\bar{g}^{kh}))\\
&\qquad \times \; < \phi_{mn}(x_0)\,\partial_p\partial_q\phi_{jk}(x_0)\,\phi_{gh}(x_0) >]\\
&\quad = 12[\, < \phi^{mn}(x_0)\,\phi^{jk}(x_0)\,\partial_m\partial_n\phi_{jk}(x_0) >\\
&\qquad + 2 < \phi^{mn}(x_0)\,\partial_m\phi^{jk}(x_0)\,\partial_j\phi_{nk}(x_0) >]\,.
\end{aligned}
$$

To compare this with the supergravity field theory, we consider the Einstein-Hilbert action

$$
I = \int d^dx\sqrt{|g|}\{-\frac{R}{2\kappa^2}\}\,. \qquad (103)
$$

Expanding to third order in κ using $g_{\mu\nu} = \eta_{\mu\nu} + 2\kappa h_{\mu\nu}$, we find the three-point interaction I_3. In harmonic gauge, *i.e.* when $\partial^\mu h_{\mu\nu} - \frac{1}{2}\partial_\nu h^\rho_\rho = 0$, and on shell $\Box h_{\mu\nu} = 0$, the cubic coupling is given by

$$
I_3 = -\kappa \int d^dx[h^{\mu\nu}h^{\rho\sigma}\partial_\mu\partial_\nu h_{\rho\sigma} + 2h^{\mu\nu}\partial_\mu h^{\rho\sigma}\partial_\rho h_{\nu\sigma}]\,. \qquad (104)
$$

The gauge transformations

$$
h_{\mu\nu} \to h_{\mu\nu} + \partial_\mu\xi_\nu + \partial_\nu\xi_\mu \qquad (105)
$$

leave invariant the harmonic gauge condition and I_3, given in (104) , when $\Box\xi_\mu = 0$. With this gauge symmetry, we could further choose $h^\rho_\rho = 0$, $\partial^\mu h_{\mu\nu} = 0$. Then I_3 is the three-graviton amplitude, and it is invariant under residual gauge transformations that have $\partial\cdot\xi = 0$.

To identify the string theory three-graviton amplitude from (103), we set b_{mn} to zero, and use the field identifiations that relate the string fields g_{mn}, ϕ to the supergravity field h_{mn} via $\phi \equiv -\frac{1}{3}h^\rho_\rho$ and $g_{mn} \equiv h_{mn} - \frac{1}{6}\bar{g}_{mn}h^\rho_\rho$, where h_{mn} is in harmonic gauge. Then $\phi_{mn} = h_{mn} - \frac{1}{2}\bar{g}_{mn}h^\rho_\rho$, and from (103) the on shell string tree amplitude is

$$
\begin{aligned}
&-\tfrac{\kappa}{12} < V_1(z_1,\bar z_1)\,(G_0^+\bar G_0^+ V_2(z_2,\bar z_2))(\tilde G_0^+\bar{\tilde G}_0^+ V_3(z_3,\bar z_3)) > \\
&= -\kappa \int d^dx\,[\,\phi^{mn}(x)\,\phi^{jk}(x)\,\partial_m\partial_n\phi_{jk}(x) + 2\,\phi^{mn}(x)\,\partial_m\phi^{jk}(x)\,\partial_j\phi_{nk}(x)\,] \\
&= -\kappa \int d^dx\,[h^{mn}h^{jk}\,\partial_m\partial_n h_{jk} \;+ 2\,h^{mn}\partial_m h^{jk}\partial_j h_{nk}\,] \\
&\qquad + \kappa \int d^dx\, h^{mn}\,\partial_m h^k_k \partial_n h^p_p \\
&= I_3 + I_3'
\end{aligned}
\tag{106}
$$

where I_3' is the one graviton - two dilaton amplitude, I_3 is the three graviton interaction in harmonic gauge, and $d = 6$. I_3 and I_3' are invariant separately under the gauge transformation (105) with $\Box\xi_n = 0$ and $\partial \cdot \xi = 0$, which corresponds to the gauge symmetry of the string field $\phi_{mn} \to \phi_{mn} + \partial_m\xi_n + \partial_n\xi_m$. I_3 by itself is also invariant under gauge transformations for which $\partial \cdot \xi \neq 0$, and these can be used to eliminate the trace of h_{mn} in I_3. In the string gauge, the trace of ϕ_{mn} is related to the dilaton $\phi^m_m = 6\phi$, so even when $b_{mn} = 0$, (103) contains both the three graviton amplitude and the one graviton – two dilaton interaction. So it turns out we could have extracted I_3 from (103) simply by setting both $b_{mn} = 0$ and $\phi = 0$, since then $\phi_{mn} = g_{mn}$ and $\partial^m g_{mn} = 0$.

Correlation functions on $AdS_3 \times S^3$ have also been studied.[17]

5. Concluding Remarks

Type IIB superstrings on $AdS_3 \times S^3 \times K3$ can have either Neveu-Schwarz *or* Ramond background flux to ensure the background metric is a solution to the equations of motion. The Neveu-Schwarz case corresponds to a WZW model and has been extensively studied.[40–52] Since these two cases are S-dual to each other, the massless spectrum is the same, but the perturbative massive spectrum will be different. For the Type IIB superstring on $AdS_5 \times S^5$, the flux supporting the metric can only be Ramond.[26–29] Thus the $AdS_3 \times S^3$ analysis discussed in these lectures is meant as a step towards the AdS_5 case.

The conjectured duality between M-theory or Type IIB string theory on anti-de Sitter (AdS) space and the conformal field theory on the boundary of AdS space[34–39] may be useful in giving a controlled systematic approximation for strongly coupled gauge theories. The formulation of vertex operators and string theory tree amplitudes for the IIB superstring on $AdS_5 \times S^5$ will allow access to the dual conformal $SU(N)$ gauge field theory CFT_4 at large N, but *small* fixed 't Hooft coupling $x = g_{YM}^2 N$ in the dual correspondence, as $(g_{YM}^2 N)^{\frac{1}{2}}(4\pi)^{\frac{1}{2}} = R_{sph}^2/\alpha'$. Presently only the large N, and *large* fixed 't Hooft coupling x limit is accessible in the CFT, since only the supergravity limit ($\alpha' \to 0$) of the correlation functions of the AdS theory is known.

Tree level n-point correlation functions for $n \geq 4$ presumably have α' corrections, since the worldsheet theory is not a free conformal field theory. However, there may be sufficiently many symmetry currents to determine the tree level correlation functions exactly in α' as well. This might be possible via integrable methods for sigma models which have a supergroup manifold target space[18,19] such as the $AdS_3 \times S^3$ theory.

Acknowledgments

I would like to thank the organizers of TASI '01 for the invitation to lecture on these topics and for providing a very lively and hospitable atmosphere. This work is partially supported by the U.S. Department of Energy under Grant No. DE-FG02-97ER-41036/Task A.

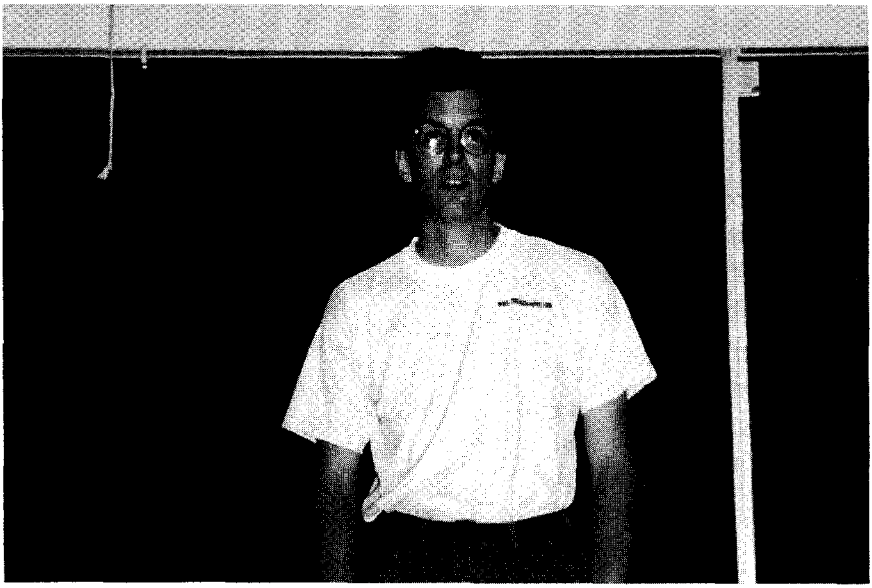
References

1. "Dual Theory", ed. M. Jacob, Physics Reports Reprint Book Series, Vol. 1, North-Holland Elsevier, 1974.
2. C. Thorn and P. Goddard, "Compatibility of the Dual Pomeron with Unitarity and the Absence of Ghosts in the Dual Resonance Model",*Phys. Lett.*B**40**, 235 (1972).
3. R. Brower, "Spectrum Generating Algebra and No Ghost Theorem for the Dual Model", *Phys.Rev.*D**6**, 1655 (1972).
4. J. H. Schwarz, "Superstring Theory",*Phys.Reports***89**, 223 (1982).
5. *Superstrings, Vol. 1&2*, ed. J. H. Schwarz, World Scientific, 1985.
6. M.B. Green, J.H. Schwarz and E. Witten, *Superstring theory, Vol. 1&2*, Cambridge University Press, 1988.
7. D. Friedan, E. Martinec, and S. Shenker, "Conformal Invariance, Supersymmetry and String Theory" *Nucl.Phys.* B**271**, 93 (1986).
8. D. Lust and S. Theisen, *Lectures on String Theory.* New York: Springer-Verlag, 1989.
9. L. Dolan, P. Goddard, and P. Montague, "Conformal Field Theory of Twisted Vertex Operators," *Nucl.Phys.* B**338**, 529 (1990).
10. N. Berkovits and C. Vafa, "$N = 4$ Topological Strings", *Nucl.Phys.* B**433**, 123 (1995), hep-th/9407190.
11. N. Berkovits, "A New Description of the Superstring", Jorge Swieca Summer School 1995, p. 490, hep-th/9604123.

12. N. Berkovits, C. Vafa, and E. Witten, "Conformal Field Theory of AdS Background With Ramond-Ramond Flux", JHEP 9903: 018, 1999; hep-th/9902098.
13. L. Dolan, and E. Witten, "Vertex Operators for AdS3 Background With Ramond-Ramond Flux", JHEP 9911:003, 1999; hep-th/9910205.
14. N. Berkovits, "Quantization of the Superstring in Ramond-Ramond Backgrounds", Presented at Strings '99, Potsdam, Germany, Class. Quant. Grav. 17: 971, 2000; hep-th/9910251.
15. L. Romans, "Self-Duality for Interacting Fields", *Nucl.Phys.* B**276**, 71 (1986).
16. S. Deger, A. Kaya, E. Sezgin and P. Sundell, "Spectrum of $D = 6, N = 4b$ Supergravity on $AdS_3 \times S^3$", *Nucl.Phys.* B**536**, 110 (1988); hepth/9804166.
17. K. Bobkov and L. Dolan,"Three-Graviton Amplitude in Berkovits-Vafa-Witten Variables", hep-th/0201027.
18. M. Bershadsky, S. Zhukov, and A. Vaintrob, "PSL(N/N) Sigma Model as a Conformal Field Theory", *Nucl.Phys.* B**559**, 205 (1999),hep-th/9902180.
19. N. Read and H. Saleur, "Exact Spectra of Conformal Supersymmetric Non-linear Sigma Models in Two Dimensions", *Nucl.Phys.* B**613**, 409 (2001); hep-th/0106124.
20. N. Berkovits, "Super-Poincare Covariant Quantization of the Superstring", JHEP 0004: 018, 2000; hep-th/0001035.
21. N. Berkovits, B.C. Vallilo, "Consistency of Super-Poincare Covariant Superstring Tree Amplitudes", JHEP 0007: 015, 2000; hep-th/0004171.
22. N. Berkovits, "Cohomology in the Pure Spinor Formalism for the Superstring", JHEP 0009: 046, 2000; hep-th/0006003.
23. Berkovits and O. Chandia, "Superstring Vertex Operators in an AdS(5)xS5 Background", *Nucl.Phys.* B**596**, 185 (2001); hep-th/0009168.
24. N. Berkovits, "Relating the RNS and Pure Spinor Formalisms for the Superstring", JHEP 0108: 026, 2001; hep-th/0104247.
25. N. Berkovits and O. Chandia, "Lorentz Invariance of the Pure Spinor BRST Cohomology for the Superstring", *Phys. Lett.*B**514**, 394 (2001); hep-th/0105149.
26. H.Kim, L. Romans, and P. van Nieuwenhuizen, "The Mass Spectrum of Chiral $N = 2$ $D = 10$ Supergravity on S^5, *Phys.Rev.*D**32**, 389 (1985).
27. M. Gunaydin and N. Marcus, "The Spectrum of the S^5 Compactification of the Chiral N=2 D=10 Supergravity and the Unitary Supermultiplets of $U(2,2/4)$," Class. Quant. Grav. **2** (1985) L11.
28. M. Gunaydin, L. Romans and N. Warner, "Gauged $N = 8$ Supergravity in Five Dimensions," Phys. Lett. **B154** (1985) 268; M. Gunaydin, L. Romans and N. Warner, "Compact and Non-Compact Gauged Supergravity Theory Theories in Five Dimensions," Nucl. Phys. **272** (1986) 598.
29. I. Bars and M. Günaydin, "Unitary Representations of Noncompact Supergroups," *Comm.Math.Phys.***91**, 31 (1983).
30. R. Metsaev and A. Tseytlin, "Type IIB superstring action in $AdS_5 \times S^5$ Background," hep-th/9805028;

31. A.A. Tseytlin, "Superstrings in AdS in Light-cone Gauge", hep-th/0009226;
32. R. Kallosh and A.A. Tseytlin, "Simplifying Superstring Action on AdS5 x S5", JHEP 9810:016,1998, hep-th/9808088;
33. N. Drukker and D. Gross, "Green-Schwarz String in $AdS5 \times S5$: Semiclassical Partition Function", JHEP 0004:021,2000, hep-th/0001204.
34. S. Ferrara, C. Fronsdal, and A. Zaffaroni, "On $N = 8$ Supergravity on AdS_5 and $N = 4$ Superconformal Yang-Mills Theory," hep-th/9802203.
35. D. Freedman, S. Mathur, A. Matusis, and L. Rastelli, "Correlation Functions in the CFT/AdS Correspondence", hep-th/9804058.
36. J. Maldacena, "The Large N Limit of Superconformal Field Theories and Supergravity," *Adv. Theor. Math. Phys.***2**, 231 (1998); hep-th/9711200.
37. S. Gubser, I. Klebanov, and A. Polyakov, "Gauge Theory Correlators from Non-critical String Theory," *Phys. Lett.*B**428**, 105 (1998); hep-th/9802109.
38. E. Witten, "Anti-de Sitter Space and Holography," *Adv. Theor. Math. Phys.***2**, 253 (1998); hep-th/9802150.
39. O. Aharony, S. Gubser, J. Maldacena, H. Ooguri, and Y. Oz, "Large N Field Theories, String Theory, and Gravity", *Phys.Reports***323**, 183 (2000); hep-th/9905111.
40. J. Teschner, "On Structure Constants and Fusion Rules in the $SL(2,C)/SU(2)$ WZNW Model", *Nucl.Phys.* B**546**, 390 (1999); hep-th/9712256.
41. J. Teschner, "Crossing Symmetry in the H_3^+ WZNW Model", hep-th/0108121.
42. A. Giveon, D. Kutasov, and N. Seiberg, "Comments on String Theory on ADS(3)," *Adv. Theor. Math. Phys.***2**, 733 (1998); hep-th/9806194.
43. D. Kutasov and N. Seiberg, "More Comments on String Theory on AdS_3", JHEP 9904:008,1999; hep-th/9903219.
44. A. Giveon and D. Kutasov, "Notes on AdS_3", hep-th/0106004.
45. J.M. Evans, M.R. Gaberdiel and M.I. Perry, "The No-ghost Theorem for AdS_3 and the Stringy Exclusion Principle," *Nucl.Phys.* B**535**, 152 (1998); hep-th/9806024.
46. F. Larsen, "The Perturbative Spectrum of Black Holes in $N = 8$ Supergravity," *Nucl.Phys.* B**536**, 258 (1998); hep-th/9805208.
47. J. de Boer, "Six-dimensional Supergravity on $S^3 \times AdS_3$ a 2d Conformal Field Theory," *Nucl.Phys.* B**548**, 139 (1999); hep-th/9806104.
48. D. Kutasov, F. Larsen and R.G. Leigh, "String Theory in Magnetic Monopole Backgrounds," *Nucl.Phys.* B**550**, 183 (1999); hep-th/9812027.
49. J. de Boer, H. Ooguri, H. Robins and J. Tannenhauser, "String Theory on AdS_3," JHEP 9812:026,1998 hep-th/9812046.
50. J. Maldacena and H. Ooguri, "Strings in $AdS(3)$ and $SL(2,R)$ WZW Model 1.: the Spectrum " *Jour.Math.Phys.***42**, 2929 (2001); hep-th/0001053.
51. J. Maldacena and H. Ooguri, "Strings in $AdS(3)$ and $SL(2,R)$ WZW Model. Part 2. Euclidean Black Hole", *Jour.Math.Phys.***42**, 2961 (2001).
52. J. Maldacena and H. Ooguri, "Strings in $AdS(3)$ and $SL(2,R)$ WZW Model. Part 3. Correlation Functions", hep-th/0111180.

This page intentionally left blank



STEVEN S. GUBSER

This page intentionally left blank

SPECIAL HOLONOMY IN STRING THEORY AND M-THEORY

STEVEN S. GUBSER
Lauritsen Laboratory of Physics
452-48 Caltech
Pasadena, CA 91125
and
Joseph Henry Laboratories
Princeton University
Princeton, NJ 08544
E-mail: ssgubser@Princeton.EDU

A brief, example-oriented introduction is given to special holonomy and its uses in string theory and M-theory. We discuss A_k singularities and their resolution; the construction of a K3 surface by resolving $T^4/\mathbf{Z}_2$; holomorphic cycles, calibrations, and worldsheet instantons; aspects of the low-energy effective action for string compactifications; the significance of the standard embedding of the spin connection in the gauge group for heterotic string compactifications; G_2 holonomy and its relation to $\mathcal{N} = 1$ supersymmetric compactifications of M-theory; certain isolated G_2 singularities and their resolution; the Joyce construction of compact manifolds of G_2 holonomy; the relation of D6-branes to M-theory on special holonomy manifolds; gauge symmetry enhancement from light wrapped M2-branes; and chiral fermions from intersecting branes. These notes are based on lectures given at TASI '01.

1. Introduction

Special holonomy plays a prominent role in string theory and M-theory primarily because the simplest vacua preserving some fraction of supersymmetry are compactifications on manifolds of special holonomy. The case that has received the most intensive study is Calabi-Yau three-folds (CY_3), first because heterotic string compactifications on such manifolds provided the first semi-realistic models of particle phenomenology, and second because type II strings on Calabi-Yau three-folds exhibit the seemingly miraculous property of "mirror symmetry." Recently, seven-manifolds with G_2 holonomy have received considerable attention, both because they provide the simplest way to compactify M-theory to four dimensions with $\mathcal{N} = 1$ supersymmetry, and because of some unexpected connections with strongly coupled gauge theory.

The purpose of these two lectures, delivered at TASI '01, is to introduce special holonomy in a way that will make minimal demands on the reader's mathematical erudition,[a] but nevertheless get to the point of appreciating a few deep facts about perturbative and non-perturbative string theory. Some disclaimers are in order: these lectures do not aspire to mathematical rigor, nor to completeness. I have made a perhaps idiosyncratic selection of material that will hopefully serve as a comprehensible invitation to the wider literature. To enhance the appeal of mathematical concepts that may seem abstruse or dreary to the theoretical physicist, I have tried to introduce such concepts either in the context of the simplest possible examples, or in the context of a piece of well-known or important piece of string theory lore. A possible downside of this approach is an occasional loss of clarity.

These lectures were constructed in with the help of some rather standard references: the survey of differential geometry by Eguchi, Gilkey, and Hansen;[1] some of the later chapters of the text by Green, Schwarz, and Witten;[2] appendix B of Polchinski's text;[3] and the original papers by D. Joyce on compact manifolds of G_2 holonomy.[4,5] The student of string theory wishing to go beyond these lectures will find Refs. 1-5 excellent jumping-off points. Also, a set of lectures on special holonomy from a pedagogical but more mathematical point of view has appeared.[6]

2. Lecture 1: on Calabi-Yau manifolds

2.1. A_k *spaces*

The simplest non-trivial Calabi-Yau manifolds are four-dimensional, even though the ones of primary interest in string model building are six-dimensional. To begin our acquaintance with four-dimensional Calabi-Yau's, let's first consider some non-compact orbifolds. In particular, regard four-dimensional flat space as $\mathbf{C}^2$ (that is, the Cartesian product of the complex plane with itself). There is a natural $SU(2)$ action on $\mathbf{C}^2$, where the two complex coordinates form a doublet. Let Γ be a discrete subgroup of $SU(2)$: for example, Γ could be the group of transformations acting on

[a]It is my hope that a graduate student who has learned General Relativity, knows the basic facts about Lie groups and their representations, and has at least a nodding acquaintance with string theory, will be able to follow the gist of this presentation. Some of the more advanced topics will require more erudition or background reading.

$\mathbf{C}^2$ like this:

$$\mathbf{Z}_{n+1}: \quad \begin{pmatrix} a \\ b \end{pmatrix} \to \begin{pmatrix} \omega & 0 \\ 0 & \omega^{-1} \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix}, \tag{1}$$

where (a, b) are coordinates on $\mathbf{C}^2$, and ω is any of the $n+1$ complex numbers satisfying $\omega^{n+1} = 1$. The simplest case would be $n = 1$, so that $\Gamma = \mathbf{Z}_2$, and then the only non-trivial transformation just changes the sign of a and b—that is, it reflects us through the origin of $\mathbf{R}^4 = \mathbf{C}^2$. Now form the orbifold $\mathbf{C}^2/\Gamma$. Overlooking the singular point at the origin, this is a manifold of holonomy Γ. More properly, we should call it an orbifold of holonomy Γ.

I haven't even defined holonomy yet, so how can we make such a statement? Consider a two-dimensional analogy: $\mathbf{R}^2$ admits a natural $SO(2)$ action, and we could also embed $\Gamma = \mathbf{Z}_{n+1} \subset SO(2) = U(1)$ in a natural way. The orbifold $\mathbf{R}^2/\Gamma$ is a cone of holonomy Γ. This claim we can understand just with pictures, and the complex case is only a slight extension. Suppose, as in figure 1, we take a vector at some point away from the tip of the cone, and parallel translate it around a loop. This is easy to do in the original Cartesian coordinates on $\mathbf{R}^2$: the vector doesn't change directions. For the loop that I drew, and for $\Gamma = \mathbf{Z}_6$, the vector comes back to itself rotated by an angle $\varphi = \pi/3$. This is what holonomy is all about: when

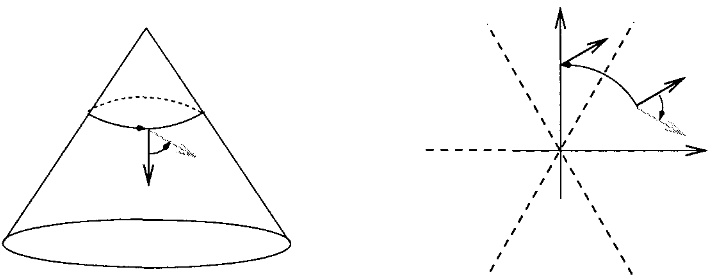


Figure 1. Left: parallel transport of a vector around the tip of a cone changes its direction. Right: the same parallel transport, where the cone is thought of as a plane modded out by a discrete group.

vectors get parallel-transported around some closed loop, their lengths remain constant but their direction can change, and the holonomy group of an n-dimensional real manifold is the subgroup of $O(n)$ that includes all possible changes of direction for a vector so transported. It is a property of the manifold as a whole, not of any special point or closed loop. So for the example in figure 1, the holonomy group is $\mathbf{Z}_6$, acting on the tangent plane of the orbifold in the obvious way. (We can define holonomy in the presence of an orbifold singularity—or any other isolated singularity—just by restricting to paths that avoid the singularity). A generic, smooth, orientable manifold has holonomy $SO(n)$. The smaller the holonomy group, the more special the manifold. If the holonomy group is trivial, the manifold is flat. A non-vanishing Riemann tensor is a local measure of non-vanishing holonomy, but we don't need to know details of this yet.

The argument around figure 1 can be repeated to show that $\mathbf{C}^2/\mathbf{Z}_{n+1}$ has holonomy $\mathbf{Z}_{n+1}$. This orbifold is called an A_n singularity. It's a singular limit of smooth Calabi-Yau manifolds, as we'll see next.

The origin of $\mathbf{C}^2/\mathbf{Z}_{n+1}$ is a curvature singularity. A persistent theme in string theory is the resolution of singularities. Singularity resolution is relatively easy work for Calabi-Yau manifolds because we often have an algebraic description of them. To see how such descriptions arise, note that a and b are double-valued on $\mathbf{C}^2/\mathbf{Z}_2$, but

$$z_1 = a^2\,, \quad z_2 = b^2\,, \quad z_3 = ab \tag{2}$$

are single-valued. We can pick any two of these as good local coordinates for $\mathbf{C}^2/\mathbf{Z}_2$. They are related by the equation

$$z_3^2 = z_1 z_2\,. \tag{3}$$

This is an *equation* for $\mathbf{C}^2/\mathbf{Z}_2$ in $\mathbf{C}^3$ (and the complex structure is correctly inherited from $\mathbf{C}^3$, though the Kahler structure is not—if you don't know what this means, ignore it for now). A *nearby* submanifold of $\mathbf{C}^3$, which is completely smooth, is

$$z_3^2 - \epsilon^2 = z_1 z_2\,, \tag{4}$$

or, after a linear complex change of variables

$$z_1^2 + z_2^2 + z_3^2 = \epsilon^2 \tag{5}$$

where we can, without loss of generality, assume $\epsilon^2 \geq 0$. Clearly, if $\epsilon = 0$, we recover our original $\mathbf{C}^2/\mathbf{Z}_2$ orbifold.

Writing $z_j = x_j + iy_j$, we can recast (5) as

$$\vec{x}^2 - \vec{y}^2 = \epsilon^2\,, \quad \vec{x}\cdot\vec{y} = 0\,. \tag{6}$$

Now define $r^2 = \vec{x}^2 + \vec{y}^2 = \sum_{i=1}^3 |z_i|^2$. For large r, our deformed manifold, (4) or (5), asymptotically approaches the original "manifold," (3). Furthermore, we can easily see that $r^2 \geq \epsilon^2$, and that for $r^2 = \epsilon^2$, we have to have $\vec{y} = 0$ and $\vec{x}^2 = \epsilon^2$: this is a sphere of radius ϵ. In fact, the manifold defined by (5) can also be described as the cotangent bundle over S^2, denoted T^*S^2. To understand this, parametrize S^2 using a real vector $\vec{w}$ with $\vec{w}^2 = \epsilon^2$. Any 1-form on S^2 can be expressed as $\vec{y} \cdot d\vec{w}$, where $\vec{y} \cdot \vec{w} = 0$. The space of all possible 1-forms over a point on S^2 is $\mathbf{R}^2$. The total space of 1-forms over S^2, which we have called T^*S^2, is thus some fibration of $\mathbf{R}^2$ over S^2. And we've just learned that this total space is parametrized by $(\vec{w}, \vec{y})$ with $\vec{w}^2 = \epsilon^2$ and $\vec{y} \cdot \vec{w} = 0$. Now if we change variables from $\vec{w}$ to $\vec{x} = \vec{w}\sqrt{1 + \vec{y}^2/\epsilon^2}$, we reproduce (6).

Let's review what's happened so far. The original orbifold, $\mathbf{C}^2/\mathbf{Z}_2$, is a cone over $S^3/\mathbf{Z}_2$. Note that $S^3/\mathbf{Z}_2$ is smooth, because the $\mathbf{Z}_2$ action on S^3 induced from (1) has no fixed points. (It's the identification of antipodal points). In fact, $S^3/\mathbf{Z}_2$ is the $SO(3)$ group manifold. The higher $S^3/\mathbf{Z}_{n+1}$ are also smooth because the $\mathbf{Z}_{n+1}$ action has no fixed point on the $U(1)$ Hopf fiber. Our algebraic resolution of the singularity led us to a smooth manifold which was asymptotic to the cone over $S^3/\mathbf{Z}_2$, but had a S^2 of radius ϵ at its "tip" rather than a singularity. This is illustrated schematically in figure 2.

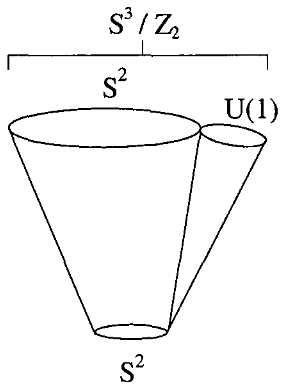


Figure 2. $S^3/\mathbf{Z}_2$ is a $U(1)$ fibration over S^2, and in the interior, the $U(1)$ shrinks but the S^2 doesn't.

This was just the beginning, because we have yet to really specify the metric on the manifolds specified by (5). We should *not* simply suppose that the metric naturally inherited from $\mathbf{C}^3$ is the one we want. In fact, the beautiful truth for these manifolds is that there is a one-parameter family of Ricci-flat Kahler metrics respecting the obvious $SO(3)$ symmetry of the equation (5) (explanation of the word "Kahler" will be forthcoming). These metrics have $SU(2)$ holonomy. This means, precisely, that the spin connection, $\omega_\mu{}^a{}_b$, generically an $SO(4)$ gauge field, lies entirely in one $SU(2)$ subgroup of $SO(4) = SU(2)_L \times SU(2)_R$. By convention we could say that the holonomy group is $SU(2)_L$. Then a *constant* right-handed spinor field ϵ_R obviously satisfies

$$\nabla_\mu \epsilon = \partial_\mu \epsilon_R + \frac{1}{4}\omega_{\mu ab}\gamma^{ab}\epsilon_R = 0\,, \tag{7}$$

just because the second term is a linear combination of the generators of rotation in $SU(2)_L$, under which ϵ_R is invariant. The integrability condition of the equation (7) is

$$[\nabla_\mu, \nabla_\nu]\epsilon_R = \frac{1}{4}R_{\mu\nu ab}\gamma^{ab}\epsilon_R = 0\,, \tag{8}$$

for any ϵ_R such that $\gamma_5\epsilon_R = -\epsilon_R$. (That's an equivalent way of saying that a spinor is right-handed). Thus, for *any* spinor ϵ (right-handed or not),

$$\begin{aligned} R_{\mu\nu ab}\gamma^{ab}(1-\gamma_5)\epsilon &= R_{\mu\nu ab}\gamma^{ab}(1-\gamma^1\gamma^2\gamma^3\gamma^4)\epsilon \\ &= R_{\mu\nu ab}\left(\gamma^{ab} - \frac{1}{2}\epsilon^{abcd}\gamma_{cd}\right)\epsilon \\ &= \left(R_{\mu\nu ab} - \frac{1}{2}\epsilon_{abcd}R_{\mu\nu}{}^{cd}\right)\gamma^{ab}\epsilon = 0\,, \end{aligned} \tag{9}$$

and, evidently, this can be true if and only if the Riemann tensor is self-dual:

$$R_{\mu\nu ab} = \frac{1}{2}\epsilon_{abcd}R_{\mu\nu}{}^{ab}\,. \tag{10}$$

Because (10) looks a lot like the equations for an instanton in non-abelian gauge theory, the metric of $SU(2)$ holonomy on (5) is known as a "gravitational instanton." This metric is known explicitly, and is called the Eguchi Hansen space, or EH_2:

$$ds^2 = \frac{dr^2}{1-(\epsilon/r)^4} + r^2\left(\sigma_x^2 + \sigma_y^2 + (1-(\epsilon/r)^4)\sigma_z^2\right)\,, \tag{11}$$

where

$$\begin{gathered} \sigma_x = \cos\psi d\theta + \sin\psi\sin\theta d\phi \quad \sigma_y = -\sin\psi d\theta + \cos\psi\sin\theta d\phi \\ \sigma_z = d\psi + \cos\theta d\phi\,. \end{gathered} \tag{12}$$

It's worth noting that the metric on S^3 can be written as

$$ds^2_{S^3} = \sigma_x^2 + \sigma_y^2 + \sigma_z^2 = d\theta^2 + \sin^2\theta d\phi^2 + (d\psi + \cos\theta d\phi)^2\,, \tag{13}$$

and the 1-forms σ_i are invariant under the left action of $SU(2)$ on $S^3 = SU(2)$. To cover S^3 once, we should let $0 \leq \theta \leq \pi$, $0 \leq \phi < 2\pi$, and $0 \leq \psi < 4\pi$. On the other hand, in the expression (11) for the Eguchi-Hansen metric, ψ is restricted to range over $[0, 2\pi)$. Thus the metric for large r is indeed a cone over $SO(3) = S^3/\mathbf{Z}_2$: the $\mathbf{Z}_2$ action on S^3 is just $\psi \to \psi + 2\pi$.

Clearly, (11) is the promised one-parameter family of metrics on the resolved A_1 singularity. The parameter is ϵ, and one can verify that the S^2 at $r = \epsilon$ indeed has radius ϵ in the metric (11). The $SO(3)$ symmetry of (5) is included in the $SU(2)$ invariance of the σ_i.

Having thoroughly disposed of this simplest example of a special holonomy metric, it's worth saying that a Calabi-Yau n-fold is, in general, a manifold of $2n$ real dimensions whose holonomy group is $SU(n)$ (or a subgroup thereof—but usually we mean that the holonomy group is precisely $SU(n)$). Any particle physicist will have encountered the embedding of $SU(2)$ in $SO(4)$ as one of the "chiral" subgroups. The inclusion of $SU(n)$ in $SO(2n)$ can be described by saying that the $2n$ real-dimensional vector representation of $SO(2n)$, which we could write as $(x_1, y_1, x_2, y_2, \ldots, x_n, y_n)$, becomes the n-dimensional complex representation of $SU(n)$, which we could write as $(z_1, z_2, \ldots, z_n)$ where $z_j = x_j + iy_j$. Having a holonomy group $SU(n)$ necessarily means that the Calabi-Yau n-fold is Ricci-flat: this is a frequently observed property of special holonomy manifolds. But not always: for instance, Kahler manifolds are $2n$ real-dimensional manifolds with holonomy group $U(n)$ (or a subgroup thereof), and these aren't Ricci-flat unless the holonomy group is contained in $SU(n)$.

The results described so far for the A_1 singularity admit interesting generalizations in several directions:

- A_n singularity: Here the natural, single-valued coordinates are $z_1 = a^{n+1}$, $z_2 = b^{n+1}$, and $z_3 = ab$, and they are related by the equation $z_3^{n+1} = z_1 z_2$, which can be deformed to $\prod_{k=1}^{n+1}(z_3 - \xi_k) = z_1 z_2$. If the constants ξ_k are all distinct, the deformed equation defines a smooth manifold in $\mathbf{C}^3$. All such manifolds admit Ricci-flat metrics. The "tip of the resolved cone" is a rather more complicated geometry now: there are $n(n+1)/2$ holomorphic embeddings of S^2 into a resolved A_n singularity, but only n are distinct in homology. Thus $b_2 = n$ for these manifolds.
- D_n and E_6, E_7, E_8 are the other finite subgroups of $SU(2)$. One can find algebraic descriptions and resolutions of $\mathbf{C}^2/\Gamma$ for these

cases as well, in a manner similar to the A_n cases.

- Another important class of $SU(2)$ holonomy metrics is the multi-center Taub-NUT solutions. They are $U(1)$ fibrations over $\mathbf{R}^3$, with metric

$$ds^2_{TN} = H d\vec{r}^2 + H^{-1}(dx^{11} + \vec{C}\cdot d\vec{r})^2 \qquad \text{where}$$
$$\nabla \times \vec{C} = -\nabla H\,, \quad H = \epsilon + \frac{1}{2}\sum_{i=1}^{n+1}\frac{R}{|\vec{r}-\vec{r}_i|}\,. \tag{14}$$

Clearly, H is a harmonic function on $\mathbf{R}^3$. There appears to be a singularity in (14) when $\vec{r}$ is equal to one of the $\vec{r}_i$, but in fact the manifold is completely smooth, for all $\vec{r}_i$ distinct, provided x^{11} is made periodic with period $2\pi R$. When $k > 1$ of the $\vec{r}_i$ coincide, there is an A_{k-1} singularity. An efficient way to see this is that, with k of the $\vec{r}_i$ coincident, we've made the "wrong" choice of the periodization of x^{11}: the right choice of period, to make the local geometry non-singular, would have been $2\pi kR$. We can get from the right choice to the wrong choice by modding out x^{11} by $\mathbf{Z}_k$, and now what's left is to convince yourself that this is the same $\mathbf{Z}_k$ action that produced A_{k-1} from $\mathbf{C}^2$. If $\epsilon > 0$, the geometry far for large r is metrically the product $S^1 \times \mathbf{R}^3$ (see figure 3). If $\epsilon > 0$, asymptotically the space is a cone over $S^3/\mathbf{Z}_n$: that is, in (14) with $\epsilon = 0$ we have exhibited explicitly the general metric of $SU(2)$ holonomy on a resolved A_n singularity.

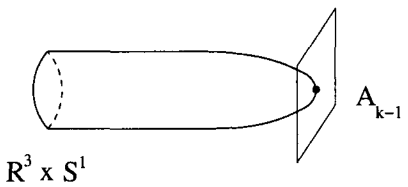


Figure 3. Single-center Taub-NUT ($k = 1$ in (14)) interpolates between $\mathbf{R}^3 \times S^1$ and an $\mathbf{R}^4$ which is well-approximated by the tangent plane to the tip of the cigar. Having k centers coincident amounts to orbifolding by $\mathbf{Z}_k$ in the S^1 direction, and results in an A_{k-1} at the tip of the cigar.

- It's now possible to outline the construction of a compact Calabi-Yau 2-fold, also known as a K3 surface. It's worth remarking that all compact, smooth Calabi-Yau 2-folds with precisely $SU(2)$ holonomy are homeomorphic (not at all an obvious result). Suppose we start with $T^4 = \mathbf{R}^4/\mathbf{Z}^4$, where the lattice $\mathbf{Z}^4$ is just the one generated by the unit vectors $(1,0,0,0)$, $(0,1,0,0)$, $(0,0,1,0)$, and $(0,0,0,1)$. Now let us identify by the action of $\mathbf{Z}_2$ which reflects through the origin: this is precisely the $\mathbf{Z}_2$ action that we used to define the A_1 singularity, so evidently there will be such a singularity at the origin. Actually, on T^4 as a whole, there are 16 fixed points of the $\mathbf{Z}_2$ action, and each is an A_1 singularity: they are at points (r_1, r_2, r_3, r_4), where each r_i can be chosen independently as 0 or 1/2. It's worth verifying that these are all the fixed points. A good way to go about it is to show that the fixed points in $\mathbf{R}^4$ of the combined action of $\mathbf{Z}^4$ and $\mathbf{Z}_2$ are the images of the 16 points we just mentioned under action of the $\mathbf{Z}^4$. A look at figure 4a) may help. At any rate, we now have a compact but singular space, and its holonomy is obviously $\mathbf{Z}_2$, with the usual caveat of avoiding fixed points (the argument is the same as always: translate a vector around the space, and the most it can do is switch its sign). The "Kummer construction" of a smooth K3 space is to cut out a region of radius R around each of the 16 A_1 singularities, and replace it by a copy of the Eguchi-Hansen space, cut off at the same finite radius R, and having an S^2 of radius $\epsilon > 0$ at its tip. This procedure works topologically because the surface $r = R$ of an Eguchi-Hansen space is $S^3/\mathbf{Z}_2$, and that's the same space as we got by cutting out a region around the A_1 singularity: the boundary of $B^4/\mathbf{Z}_2$, where B^4 is a ball with boundary S^3. See figure 4b). The metric does not quite match after we've pasted in copies of EH_2, but it nearly matches: the errors are $O(\epsilon^4/R^4)$. Neglecting these small errors, we have a smooth manifold of $SU(2)$ holonomy: the crucial point here is that each EH_2 has the *same* $SU(2)$ subgroup of $SO(4)$ as its holonomy group, namely the $SU(2)$ which contains the original discrete $\mathbf{Z}_2$ holonomy of the A_1 singularity—and that $\mathbf{Z}_2$ is the same for all 16 fixed points. A *non-trivial* mathematical analysis shows that the $O(\epsilon^4/R^4)$ can be smoothed out without enlarging the holonomy group. It's easy to understand from this analysis that K3 has 22 homologically distinct 2-cycles: T^4 started out with 6 that are undisturbed by the $\mathbf{Z}_2$ orbifolding (think of their cohomological partners, for instance $dr_1 \wedge dr_2$, obviously $\mathbf{Z}_2$ even); and each EH_2 adds one to the total because of the unshrunk S^2 at its tip. As remarked earlier, all K3 surfaces are homeomorphic. Hence all of them have second Betti number $b_2 = 22$.

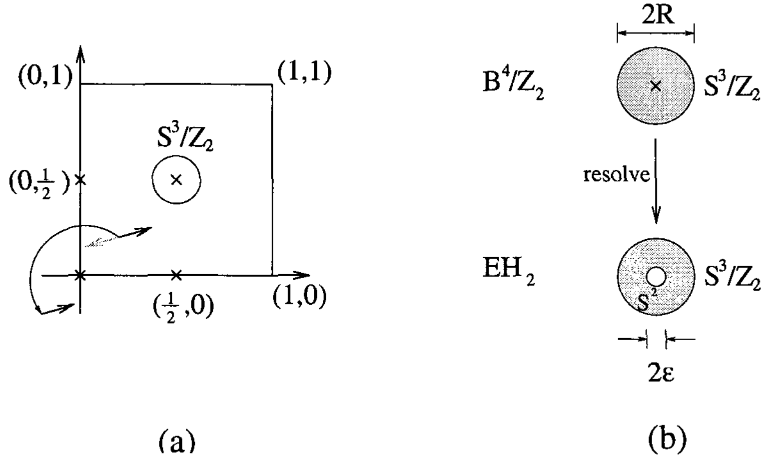


Figure 4. (a): Schematic description of $T^4/\mathbf{Z}_2$. The unit cell of a square torus is quotiented by the action of a $\mathbf{Z}_2$ whose fixed points are indicated by x's. (Actually there would be $2^4 = 16$ such fixed points for T^4, but we could only draw T^2 here). Each fixed point is an A_1 singularity, so the boundary of a region around it is $S^3/\mathbf{Z}_2$ in the quotient space. The quotient is an orbifold of $\mathbf{Z}_2$ holonomy: parallel transport of a vector along a curve, plus its reflected image, are shown. (b) We resolve a $B^4/\mathbf{Z}_2$ region around each A_1 singularity into the central portion of an Eguchi-Hansen space, with an unshrunk S^2 of radius ϵ.

It's worth reflecting for a moment on why we were able to get so far in the study of the A_k spaces just by manipulating complex equations like $z_3^2 = z_1 z_2$. This defining equation for the A_1 space does not determine its metric, but it does determine its complex structure. That is, the notion of holomorphicity is inherited from $\mathbf{C}^3$ to the subspace defined by the algebraic equation. Another way to say it is we automatically have a distinguished way of assembling four real coordinates into two complex coordinates. Note that we haven't said anything yet about the metric! The natural notion of a metric that is "compatible" with a given complex structure is what's called a Kahler metric: it is one which can be expressed locally as

$$ds^2 = 2g_{i\bar{j}}dz^i d\bar{z}^{\bar{j}} \qquad \text{where} \quad g_{i\bar{j}} = \frac{\partial^2 K}{\partial z^i \partial \bar{z}^{\bar{j}}}, \tag{15}$$

for some function $K(z^i, \bar{z}^j)$ which is called the Kahler potential. Evidently, $K(z^i, \bar{z}^j)$ can be modified by the addition of a holomorphic or an anti-holomorphic function. It is quite straightforward to show that the Christoffel connection associated with a Kahler metric preserves the splitting of the tangent plane into holomorphic and anti-holomorphic pieces: for instance, if a vector points in the z_1 direction, then after parallel transport, it may have components in the z_1 and z_2 directions, but none in the $\bar{z}^1$ and $\bar{z}^2$ directions. This is why Kahler metrics on an n-complex-dimensional space necessarily have holonomy $U(n)$.

Yau proved that if a smooth, compact manifold, admitting a complex structure and a Kahler metric, obeys a certain topological condition (vanishing of the first Chern class), then it's possible to find a Ricci-flat Kahler metric. (Some further facts are not so hard to show: the Ricci-flat metric is unique given the cohomology class of the Kahler form; and Ricci-flat Kahler metrics are precisely those with holonomy contained in $SU(n)$). By virtue of Yau's theorem, we can go far in the study of $SU(n)$ holonomy manifolds just by manipulating simple algebraic equations: the equations specify a topology and a complex structure (inherited from the complex structure in the flat space or projective space in which we write the defining equations) and provided we can demonstrate the (rather weak) topological hypotheses of Yau's theorem, we can be sure of the existence of a $SU(n)$ holonomy metric even if we can't write it down. Perhaps the simplest way to look at it is that you get to $U(N)$ holonomy just by knowing the complex structure. The Kahler metric is detailed and difficult information, but a lot of interesting facts can be learned without knowing much about it other than its existence.

We have discussed some of the simplest special holonomy manifolds, and sketched the Kummer construction for a compact K3; but much much more remains unsaid. There are highly developed ways of constructing Calabi-Yau three-folds, of which elliptic fibration, toric geometry, and the intersection of algebraic varieties in complex projective spaces deserve special mention. Far too much is in the literature to even summarize here; but the interested reader will find much already in the references to these lectures.

2.2. *Non-linear sigma models and applications to string theory*

I find it irresistible at this point to detour into some applications of notions from special holonomy to supersymmetry and string theory. In four dimensions, the most general renormalizable lagrangian for a single chiral superfield, $\Phi = \phi + \theta^\alpha \psi_\alpha + \theta^\alpha \theta_\alpha F$, is

$$\mathcal{L} = \int d^4\theta\, \Phi^\dagger \Phi + \left(\int d^2\theta\, W(\Phi) + h.c. \right) , \tag{16}$$

with $W(\Phi)$ some cubic polynomial. Let us work in Euclidean signature. The most general *effective* action for several chiral superfields (that is, a totally general local form up to two derivative) is the following:

$$\mathcal{L} = \int d^4\theta K(\Phi_i, \Phi_i^\dagger) + \left(\int d^2\theta W(\Phi) + h.c.\right)$$
$$= g_{i\bar{j}}\partial_\mu\phi^i\partial^\mu\bar{\phi}^{\bar{j}} + g_{i\bar{j}}\psi^i\partial\!\!\!/\bar{\psi}^{\bar{j}} + g^{i\bar{j}}\frac{\partial W}{\partial\phi^i}\frac{\partial\bar{W}}{\partial\bar{\phi}^{\bar{j}}} + \left(\frac{\partial^2 W}{\partial\phi^i\partial\phi^j}\psi^i\psi^j + h.c.\right) + \dots, \tag{17}$$

where in the last line I have eliminated the auxiliary fields F_i through their algebraic equations of motion. In expanding things out in components I have left out various interaction terms, and I have not been particularly careful with all factors of 2 and signs.

If the superpotential is 0, then the lagrangian is just

$$L = g_{i\bar{j}}(\phi^k, \bar{\phi}^{\bar{k}})\partial_\mu\phi^i\partial^\mu\bar{\phi}^{\bar{j}} + \text{fermions}, \tag{18}$$

which is just a non-linear sigma model with a Kahler target. There are various reasons to be interested in the lagrangians (17) and (18), but let us point out one that is particularly relevant to string theory. If we make a dimensional reduction to two dimensions, setting $\phi^i = Z^i/\sqrt{2\pi\alpha'}$, then we obtain an action

$$S = \frac{1}{2\pi\alpha'}\int d^2z\, g_{i\bar{j}}(Z^k, \bar{Z}^{\bar{k}})\left(\partial_z Z^i\partial_{\bar{z}}\bar{Z}^{\bar{j}} + \partial_{\bar{z}}Z^i\partial_z\bar{Z}^{\bar{j}} + \text{fermions}\right). \tag{19}$$

The bosonic part written out explicitly is precisely the so-called Polyakov action, $S_{\text{Pol}} = \frac{1}{2\pi\alpha'}\int d^2z\, g_{ab}\partial_z X^a\partial_{\bar{z}}X^b$, written in terms of complex variables, $Z^j \propto X^{2j-1} + iX^{2j}$. The action (19) describes strings propagating on a Kahler manifold. We know (see for instance E. D'Hoker's lectures at this school) that conformal invariance forces this manifold to be ten-dimensional and *Ricci flat*, in the leading approximation where α' is small compared to characteristic sizes of the manifold. For instance, the target space could be a Calabi-Yau manifold times flat space: this is part of the standard strategy for getting four-dimensional models out of the heterotic string (more on this later).

A simpler example would be for the target space just to be $\mathbf{R}^6$ times the Eguchi-Hansen space, EH_2. (In fact, we could even use the singular orbifold limit, provided $\int_{S^2} B_2 = \pi$; but it is too much to consider here in detail how string physics can be smooth on a singular geometry). Pursuing our simple $\mathbf{R}^6 \times EH_2$ example a little further: an obvious thing for a string to do is to wrap the S^2 in EH_2. The string is then an instanton with respect to the $\mathbf{R}^6$ directions, and to compute its contribution to the path integral, the first thing we have to know is the minimal classical action for such a string.[b] To this end, it is worth recalling that the Polyakov action coincides with the Nambu-Goto action after the worldsheet metric is eliminated through its algebraic equation of motion. So the minimal action will be attained by a worldsheet wrapped on the minimal area S^2. Finding this S^2 is straightforward work, since we have the explicit metric for EH_2: it's obviously $r = \epsilon$. But for a more general discussion, it's worth introducing a little more technology, in the form of the Kahler form

$$J = i g_{i\bar{j}}(Z^k, Z^{\bar{k}}) dZ^i \wedge d\bar{Z}^{\bar{j}} \,. \tag{20}$$

Since both $g_{i\bar{j}}\partial_z Z^i \partial_{\bar{z}} \bar{Z}^{\bar{j}}$ and $g_{i\bar{j}}\partial_{\bar{z}} Z^i \partial_z \bar{Z}^{\bar{j}}$ are everywhere positive quantities, it's clear that

$$\frac{1}{2\pi\alpha'}\int_{S^2} J = \frac{1}{2\pi\alpha'}\int_{S^2} d^2z\, g_{i\bar{j}}\left(\partial_z Z^i \partial_{\bar{z}}\bar{Z}^{\bar{j}} - \partial_{\bar{z}} Z^i \partial_z \bar{Z}^{\bar{j}}\right) \leq S_{\mathrm{Pol}} \tag{21}$$

with equality precisely if $g_{i\bar{j}}\partial_{\bar{z}} Z^i \partial_z \bar{Z}^{\bar{j}} = 0$, which is equivalent to $\partial_{\bar{z}} Z^i = 0$ for all i. This last equation expresses the condition that the map $z \longrightarrow Z^i(z)$ is a *holomorphic embedding* of the worldsheet into the target spacetime. Obviously, we could consider anti-holomorphic embeddings, and prove in an analogous way that precisely they saturate the inequality $\frac{1}{2\pi\alpha'}\int_{S^2} J \geq -S_{\mathrm{Pol}}$. A string anti-holomorphically embedded in EH_2 would just be one at $r = \epsilon$, wrapping the S^2 with the opposite orientation. Thus we have world-sheet instantons and world-sheet anti-instantons.

The inequality (21) is deceptively simple. Actually it illustrates a very powerful notion: calibration. To see things in a properly general light, first note that we didn't need the two-cycle to be S^2: it could have been any homologically non-trivial two-cycle, call it Σ. Furthermore, we could have derived a pointwise form of the inequality in (21) (obvious since we didn't need any integrations by parts to get the inequality we did derive). That pointwise form would say that the pullback of the Kahler form J to the worldsheet is equal to a multiple of the volume form (defined through the induced metric on the worldsheet), and the multiple is a function that never exceeds 1. A final important ingredient to the setup of a calibration is that J is closed, $dJ = 0$. This arises because $J = i\partial\bar{\partial}K$, where ∂ is the exterior

[b] We would eventually have in mind formulating a string theory in $\mathbf{R}^{5,1}$ via Wick rotation from $\mathbf{R}^6$—or, in the more physically interesting case of a Calabi-Yau three-fold, in $\mathbf{R}^{3,1}$ via Wick rotation from $\mathbf{R}^4$—but we carry on in the hallowed tradition of doing all computations in Euclidean signature until the very end.

derivative with respect to the Z^i's, and $\bar{\partial}$ is the exterior derivative with respect to the $\bar{Z}^{\bar{i}}$'s. So to state the whole setup once and for all and with full generality: a calibration is a closed p-form which restricts (or, more precisely, pulls back) onto any p-submanifold to a scalar multiple of the induced volume form, where the multiple is nowhere greater than 1; and a calibrated cycle is one whose induced volume form precisely coincides with the pullback of the calibration form. An inequality like (21) then ensures that the volume of the calibrated cycle is minimal among all possible cycles in its homology class: this is because the integral of the calibrating form (i.e. the left hand side of (21)) depends only on the homology of what you're integrating it over.

Suppose now we have a compactification of string theory from ten dimensions to four on a (compact) Calabi-Yau three-fold, CY_3. If we pick a basis N^A of homology two-cycles for CY_3, then we could define the Kahler parameters as $v^A = \int_{N^A} J$. From the preceding discussion, v^A is just the minimal area two-cycle in a given equivalence class. A natural complexification of v^A is

$$T^A = \int_{N^A} (J + iB)\,, \tag{22}$$

where B is the NS 2-form, assumed to have $dB = 0$. The T^A are the so-called complexified Kahler moduli of the Calabi-Yau compactification. The claim is that they become massless complex fields in four dimensions. To see this in precise detail, we should perform a rigorous Kaluza-Klein reduction. Without going that far, we can convince ourselves of the claim by expanding

$$J + iB = \sum_A T^A \omega_A\,, \tag{23}$$

where the ω_A are harmonic two-forms with $\int_{N^A} \omega_B = \delta^A_B$; and (23) is basically the beginnings of a Kaluza-Klein reduction, where T^A depends only on the four non-compact dimensions. Since the left hand side of (23) is harmonic (or may at least be made so by a gauge choice) and the ω_A are harmonic, the T^A are indeed massless fields in four dimensions. Compactification on CY_3 preserves $1/4$ of supersymmetry (a theme to be developed more systematically in the next lecture), which means $\mathcal{N} = 1$ supersymmetry in $d = 4$ for a heterotic string compactification, and $\mathcal{N} = 2$ supersymmetry in $d = 4$ for a type II string compactification. Since we have at least $\mathcal{N} = 1$ supersymmetry, the complex scalar fields T^A must be components of chiral superfields, with an action of the form (16), for some

Kahler target manifold that describes all possible values of the complexified Kahler moduli for a given Calabi-Yau compactification.[c] What a mouthful! Now comes the nice part: having learned that the T^A are massless fields based on an argument that applied for *any* Calabi-Yau, we can confidently say that $V = 0$ identically, so also the superpotential $W = 0$. These are classical statements, because the argument that the T^A were massless was based on classical field equations. However, as is often the case, W is protected against contributions from loops by the unbroken $\mathcal{N} = 1$ supersymmetry. More precisely, a Peccei-Quinn symmetry for $\int_{N^A} B$, plus holomorphy, protects W against *all* perturbative string corrections. There are in fact non-perturbative corrections that come from the world-sheet instantons discussed above: the action of such an instanton is

$$S = \frac{1}{2\pi\alpha'} \int_{N^A} (J + iB) = \frac{T^A}{2\pi\alpha'}\,, \tag{24}$$

and because of the explicit T^A-dependence, we obviously must expect some nonperturbative $e^{-T^A/2\pi\alpha'}$ contribution to W to arise from these instantons.

This is about all one can learn about the dynamics of complexified Kahler moduli for CY_3 compactifications of superstrings based on $\mathcal{N} = 1$ supersymmetry. It's actually quite a lot: we have non-linear sigma model dynamics on a Kahler manifold whose complex dimension is b_2 of the CY_3, corrected only non-perturbatively in the small dimensionless parameters $T^A/2\pi\alpha'$. More can be learned, however, if there is $\mathcal{N} = 2$ supersymmetry—that is, for CY_3 compactifications of a type II superstring. Then one can show that the Kahler metric on the moduli space follows from the Kahler potential

$$K = -\log \mathcal{W}(\mathrm{Re}\, T^A) \qquad \mathcal{W} = \int_{CY_3} J \wedge J \wedge J\,, \tag{25}$$

where J is the Kahler form of the CY_3 (but K is the Kahler potential for the many-dimensional moduli space, and as such is a function of T^A and $\bar{T}^{\bar{A}}$). Explaining how (25) arises from $\mathcal{N} = 2$ supersymmetry would take us too far afield; it is enough for us to know that, whereas $\mathcal{N} = 1$ supersymmetry usually protects only the holomorphic object W from corrections, $\mathcal{N} = 2$ supersymmetry tightly constrains the Kahler form as well, protecting it

[c]We have not substantially constrained how these moduli might couple to other sorts of matter. This issue is beyond the scope of the present lectures.

in this case from all perturbative string corrections. There are worldsheet instanton corrections, as before.

A substantial omission in our treatment is that we haven't discussed complex structure moduli. Understanding them, and also the worldsheet origin of both types of moduli, is crucial to the formulation of mirror symmetry in string theory. The reader may wish to consult TASI lectures from previous years (for instance Ref. 7) for an introduction to these fascinating topics.

A truly remarkable property of heterotic string theory dynamics is that the form (25) continues to hold true, modulo similar non-perturbative corrections, in $\mathcal{N}=1$ compactifications of the heterotic string with the "standard embedding" of the spin connection in the gauge group. "Standard embedding" means that one sets gauge potentials $A_\mu{}^I{}_J$ in a particular $SU(3)$ subgroup of $SO(32)$, or of $E_8 \times E_8$, equal to the spin connection $\omega_\mu{}^a_b$ of the CY_3. In contrast to the results presented so far, the fact that (25) persists for these $\mathcal{N}=1$ constructions goes beyond anything one could understand based only on low-energy effective field theory, and is truly stringy in its origin. Before returning to the narrower venue of special holonomy, let us then detour into a demonstration of this claim. Amusingly, almost all the tools we will use have already been introduced.

The basic point is that, for the standard embedding, the CY_3 part of the heterotic worldsheet CFT is identical to the corresponding part of the type II worldsheet CFT. Because the heterotic CFT factorizes into a $\mathbf{R}^{3,1}$ part, a CY_3 part (to be described), and an "extra junk" part, the physical dynamics of the CY_3 is the same for the heterotic and type II constructions. It's as if there were a "secret" $\mathcal{N}=2$ supersymmetry in the heterotic string. To write down type II superstring propagation on a CY_3, we need to make the non-linear sigma model (19) explicitly supersymmetric. With the help of superfields

$$\mathcal{X}^a = X^a + i\theta\psi^a + i\bar{\theta}\tilde{\psi}^a + \theta\bar{\theta}F^a\,, \tag{26}$$

one can write a simple supercovariant worldsheet action:

$$\begin{aligned} S_{CY_3} &= \frac{1}{2\pi\alpha'}\int d^2z d^2\theta\, g_{ab}(\mathcal{X}) D_{\bar{\theta}}\mathcal{X}^a D_\theta \mathcal{X}^b \\ &= \frac{1}{2\pi\alpha'}\int d^2z \Big[g_{ab}(X)\partial X^a \bar{\partial} X^b + g_{ab}(\psi^a D_{\bar{z}}\psi^b + \tilde{\psi}^a D_z \tilde{\psi}^b) \\ &\qquad + \frac{1}{2}R_{\mu\nu\rho\sigma}(X)\psi^\mu\psi^\nu\tilde{\psi}^\rho\tilde{\psi}^\sigma \Big]\,, \end{aligned} \tag{27}$$

where the second equality holds after auxiliary fields have been algebraically eliminated. The covariant derivatives are defined as follows:

$$\begin{aligned} D_\theta &= \partial_\theta + \theta\partial_z \qquad D_{\bar\theta} = \partial_{\bar\theta} + \bar\theta\partial_z \\ D_{\bar z}\psi^a &= \partial_{\bar z}\psi^a + \partial_{\bar z}X^b\Gamma^a_{bc}(X)\psi^c \\ D_z\tilde\psi^a &= \partial_z\tilde\psi^a + \partial_z X^b\Gamma^a_{bc}(X)\tilde\psi^c\,. \end{aligned} \tag{28}$$

The complicated second term in $D_{\bar z}\psi^a$ and $D_z\tilde\psi^a$ are the pull-backs of the Calabi-Yau connection to the string worldsheet. The full action for type II superstrings on $\mathbf{R}^{3,1}\times CY_3$ is

$$S_{II} = \frac{1}{2\pi\alpha'}\int d^2z d^2\theta\,\eta_{\mu\nu}D_{\bar\theta}\mathcal{X}^\mu D_\theta\mathcal{X}^\nu + S_{CY_3}\,. \tag{29}$$

The heterotic string possesses only the anti-holomorphic fermions $\tilde\psi^M$: instead of the corresponding ten holomorphic fermions ψ^M, the heterotic string has 32 holomorphic fermions λ^I. (The choice of GSO projection determines whether we have $SO(32)$ or $E_8\times E_8$ as the gauge group. In the latter case, $SO(16)\times SO(16)$ is manifest in the above description, as rotations of the λ^I's in two sets of 16. For further details about the heterotic string, standard string theory texts should be consulted). The action of the heterotic string is

$$\begin{aligned} S_{Het} = \frac{1}{2\pi\alpha'}\int d^2z\bigg[& g_{MN}\partial X^M\bar\partial X^N + g_{MN}\tilde\psi^M D_z\tilde\psi^N + \delta_{IJ}\lambda^I\mathcal{D}_{\bar z}\lambda^J \\ &+\frac{1}{2}F^{IJ}_{MN}\lambda^I\lambda^J\tilde\psi^M\tilde\psi^N\bigg] \end{aligned} \tag{30}$$

where the only new derivative we need to define is

$$\mathcal{D}_{\bar z}\lambda^I = \partial_z\lambda^I + A^{IJ}_M(X)\partial_{\bar z}X^M\lambda^J\,, \tag{31}$$

the second term being the heterotic gauge field pulled back to the worldsheet. (It's easiest to think of the A^{IJ}_M either as $SO(32)$ gauge fields, or in the $E_8\times E_8$ case as $SO(16)\times SO(16)$ gauge fields, which have to be augmented by some other fields to make up the full $E_8\times E_8$, but these other fields will never be turned on in our construction). Now for the punch-line: we can embed S_{CY_3} into S_{Het} by "borrowing" six of the λ^I to replace the six lost ψ^a. More explicitly,

$$\begin{aligned} &\psi^I \equiv e^I_a(X)\psi^a \to \lambda^I\,,\quad \omega_a^{IJ}\to A_a^{IJ}\,,\quad {R_{ab}}^{IJ}\to {F_{ab}}^{IJ} \\ &\delta_{IJ}e^I_a e^J_b = g_{ab}\quad I=1,\ldots,6 \qquad D_{\bar z}\psi^I = \partial_{\bar z}\psi^I + \partial_{\bar z}X^a\omega_a^{IJ}(X)\psi^J\,. \end{aligned} \tag{32}$$

Thus, quite literally, we are embedding a particular $SU(3) \subset SU(4) = SO(6) \subset SO(16)$, and the $SO(16)$ is either part of $SO(32)$ or E_8. Clearly, $SU(3) \subset SU(4)$ in only one way, and $SO(6) \subset SO(16)$ so that $SO(6)$ rotates only 6 components of the real vector representation of $SO(16)$).

A lesson to remember, even if not all the details registered, is that the spin connection can be thought of as just another connection (acting on the tangent bundle so that $\nabla_a v^I = \partial_a v^I + \omega_a{}^I{}_J v^J$), and it is not only well-defined, but in fact quite convenient, to set some of the gauge fields of the heterotic string equal to the spin connection of $SU(3)$ holonomy that we know exists on any Calabi-Yau. Less minimal choices have been extensively explored—see for example D. Waldram's lectures at this school.

3. Lecture 2: on G_2 holonomy manifolds

Given that all string theories can be thought of as deriving from a single eleven-dimensional theory, M-theory, by a chain of dualities, it is natural to ask what are the sorts of seven-dimensional manifolds we can compactify M-theory on to obtain minimal supersymmetry in four-dimensions.[d] This is the most obvious string theory motivation for studying seven-manifolds of G_2 holonomy, as indeed we shall see that M-theory on such manifolds leads to $\mathcal{N} = 1$ supersymmetry in $d = 4$.

But what is G_2? It can be defined as the subgroup of $SO(7)$ whose action on $\mathbf{R}^7$ preserves the form

$$\begin{aligned}\varphi &= dy^1 \wedge dy^2 \wedge dy^3 + dy^1 \wedge dy^4 \wedge dy^5 + dy^1 \wedge dy^6 \wedge dy^7 + dy^2 \wedge dy^4 \wedge dy^6 \\ &\qquad - dy^2 \wedge dy^5 \wedge dy^7 - dy^3 \wedge dy^4 \wedge dy^7 - dy^3 \wedge dy^5 \wedge dy^6 \\ &\equiv \frac{1}{6}\varphi_{abc} dy^a dy^b dy^c \,.\end{aligned} \tag{33}$$

The φ_{abc} happen to be the structure constants for the imaginary octonions. We will not use this fact, but instead take the above as our *definition* of G_2. Let's now do a little group theory. $SO(7)$ has rank 3 and dimension 21. Three obvious representations are the vector **7**, the spinor **8**, and the adjoint **21**. G_2, on the other hand, has rank 2 and dimension 14. See figure 5.

---

[d]Some people might prefer the phrasing, "All string theories are special limits of a mysterious theory, M-theory, of which another limit is eleven-dimensional supergravity." I will prefer to use M-theory in its more restrictive sense as a theory emphatically tied to eleven dimensions—in other words, the as-yet unknown quantum completion of eleven-

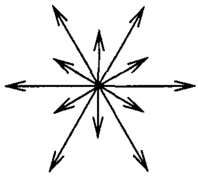


Figure 5. The Dynkin diagram for G_2. The weights comprising the **7** are the six short roots plus one node at the origin.

It has two obvious representations: the fundamental **7** (comprising the short roots plus one weight at the origin) and the adjoint **14**. As a historical note, it's worth mentioning that G_2 enjoyed brief popularity as a possible group to describe flavor physics: the **7** was supposed to be the multiplet of pseudoscalar mesons. That looked OK until it was realized that the η had to be included in this multiplet, which made the **8** of $SU(3)$ clearly superior. Besides, spin 3/2 baryons almost filled out the **10** of $SU(3)$, and then the discovery of the Ω^- completed that multiplet and clinched $SU(3)$'s victory. To return to basic group theory, it's worth noting some branching rules:

$$\begin{array}{cc} SO(7) \supset G_2 & G_2 \supset SU(3) \\ \mathbf{21} = \mathbf{14} \oplus \mathbf{7} & \mathbf{7} \supset \mathbf{3} \oplus \bar{\mathbf{3}} \oplus \mathbf{1} \\ \mathbf{7} = \mathbf{7} & \mathbf{14} = \mathbf{8}_{\rm adj} \oplus \mathbf{3} \oplus \bar{\mathbf{3}} \\ \mathbf{8} = \mathbf{7} \oplus \mathbf{1} & \end{array} \tag{34}$$

The second rule in the right column suggests another construction of G_2, as $SU(3)$ plus generators in the **3** and the $\bar{\mathbf{3}}$—this is similar to the construction of E_8 from $SO(16)$ plus spinor generators.

The construction of G_2 as a subgroup of $SO(7)$ makes it clear that G_2 is a possible holonomy group of seven-manifolds. Before explaining this in detail, let us re-orient the reader on the concept of holonomy. Recall that on a generic seven-manifold, parallel transport of a vector around a closed curve brings it back not to itself, but to the image of itself under an $SO(7)$ transformation which depends on the curve one chooses. See figure 6. The reason that the transformation is in $SO(7)$ is that the *length* of the vector is preserved: parallel transport means $t^\mu \nabla_\mu v^\alpha = 0$ along the curve C, and this implies $t^\mu \nabla_\mu (g_{\alpha\beta} v^\alpha v^\beta) = 0$ (because $\nabla_\mu g_{\alpha\beta} = 0$); so indeed the length

---

dimensional supergravity.

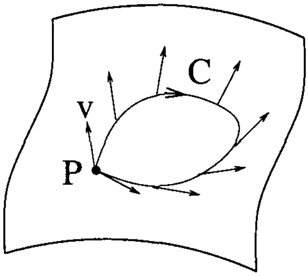


Figure 6. Parallel transport of a vector v around a curve C. Upon returning to the point of origin P, v has undergone some rotation, which for a seven-manifold is an element of $SO(7)$.

of the vector v is the same, all the way around the curve. Suppose we now choose some seven-bein e^a_μ, satisfying $\delta_{ab}e^a_\alpha e^b_\beta = g_{\alpha\beta}$. Parallel transporting all seven of these 1-forms around our closed curve C results in

$$e^a_\alpha \to O^a{}_b e^b_\mu\,, \tag{35}$$

where $O^a{}_b \in SO(7)$. Parallel transport in this context means transport with respect to the covariant derivative $\nabla_\nu e^a_\mu = \partial_\nu e^a_\mu - \Gamma^\rho_{\nu\mu}e^a_\rho$: that is, we treat a merely as a label. One often defines another covariant derivative, D_μ, such that a flat index a results in an extra term involving the spin connection: thus for instance

$$D_\nu e^a_\mu = \partial_\nu e^a_\nu - \Gamma^\rho_{\nu\mu}e^a_\rho + \omega_\nu{}^a{}_b e^b_\mu\,. \tag{36}$$

The spin connection can then be defined by the equation $D_\nu e^a_\mu = 0$.

Thus far our setup has nothing to do with G_2: we have merely explained (or re-explained) some standard aspects of differential geometry. Now suppose our seven-manifold is special, in that for *some* choice of seven-bein e^a_μ, the three-form

$$\begin{aligned}\varphi &= e^1\wedge e^2\wedge e^3 + e^1\wedge e^4\wedge e^5 + e^1\wedge e^6\wedge e^7 + e^2\wedge e^4\wedge e^6\\ &\quad - e^2\wedge e^5\wedge e^7 - e^3\wedge e^4\wedge e^7 - e^3\wedge e^5\wedge e^6\\ &\equiv \frac{1}{6}\varphi_{abc}e^a e^b e^c\end{aligned} \tag{37}$$

satisfies $\nabla_\mu\varphi_{\alpha\beta\gamma} = 0$. That means, in particular, that if we parallel transport φ around C, it comes back to itself. Rephrasing this statement using (35) and the concise form $\varphi = \frac{1}{6}\varphi_{abc}e^a e^b e^c$, we see that $\varphi_{abc}O^a{}_d O^b{}_e O^c{}_f = \varphi_{def}$. So the $SO(7)$ transformation $O^a{}_b$ is actually an element of G_2; and since the curve C was arbitrary, the manifold's holonomy group is G_2.

The presentation of the previous paragraph is in the order that my intuition suggests; however it's actually backwards according to a certain logic. A mathematician might prefer to state it this way: it so happens that preservation of the form (33) under a general linear transformation *implies* preservation of the metric δ_{ab}. So we could start with a manifold M_7 endowed only with differential structure, choose a globally defined three-form φ on it, *determine* the metric $g_{\mu\nu}$ in terms of φ,[e] determine the connection ∇_μ in terms of $g_{\mu\nu}$, and then ask that $\nabla_\mu \varphi_{\alpha\beta\gamma} = 0$ in order to have a G_2 holonomy manifold. This amounts to a set of hugely non-linear differential equations for the three-form coefficients $\varphi_{\alpha\beta\gamma}$.

The decomposition $\mathbf{8} = \mathbf{7} \oplus \mathbf{1}$ of the spinor of $SO(7)$ into representations of G_2 is important, because it means that G_2 holonomy manifolds admit precisely one covariantly constant spinor. To construct it, start at any point P, choose ϵ at P as the singlet spinor according to the above decomposition, and then parallel transport ϵ everywhere over the manifold. There is no path ambiguity because the spinor always stays in the singlet representation of G_2. All other spinors are shuffled around by the holonomy: only the one we have constructed satisfies $\nabla_\mu \epsilon = 0$. The equation for preserved supersymmetry in eleven-dimensional supergravity, with the four-form $G_{(4)}$ set to zero, is

$$\delta \psi_\mu = \nabla_\mu \eta = 0 \,. \tag{38}$$

For an eleven-dimensional geometry $\mathbf{R}^{3,1} \times M_7$, where M_7 has G_2 holonomy, the solutions for η in (38) are precisely ϵ tensored with a spinor in $\mathbf{R}^{3,1}$: that is, compactification on M_7 preserves one eighth of the possible supersymmetry, which amounts to $\mathcal{N} = 1$ in $d = 4$. It can also be shown that if a manifold has precisely one covariantly constant spinor ϵ, then its holonomy group is G_2, or at least a large subgroup thereof. One can in fact construct the covariant three-form φ as a bilinear in ϵ.

It would seem that G_2 holonomy compactifications of 11-dimensional supergravity would be of utmost phenomenological interest; however, one should recall Witten's proof[9] that compactifications of 11-dimensional supergravity on any *smooth* seven-manifold cannot lead to chiral matter in four dimensions. With a modern perspective, we conclude that we should therefore be studying *singularities* in G_2 holonomy manifolds, or branes, or some other defects where chiral fermions might live.

[e] A formula for the metric in terms of φ will be given in section 3.1. The validity of this formula already requires that φ have some non-degeneracy properties. A more careful analysis can be found in Ref. 8.

The usual starting point for investigating singularities in an n-dimensional manifold is to look at non-compact manifolds which are asymptotically conical:

$$ds_n^2 \sim dr^2 + r^2 d\Omega_{n-1}^2 \tag{39}$$

for large r. Note that if $\sim$ were replaced by an exact equality, then the metric ds_n^2 would be singular at $r = 0$ unless $d\Omega_{n-1}^2$ is the metric on a unit $(n-1)$-sphere. In the previous lecture, we encountered a prime example of this sort of singularity: A_k singularities in four-manifolds locally have the form (39) with $d\Omega_3^2$ being the metric of the Lens space $S^3/\mathbf{Z}_{k+1}$. Another frequently discussed example is the conifold singularity in Calabi-Yau three-folds: this is locally a cone over the coset space $T^{11} = SU(2) \times SU(2)/U(1)_{\text{diag}}$. The conifold admits a Calabi-Yau metric that is known explicitly, as are certain resolutions of the singularity which remain Calabi-Yau (much like the resolutions of the A_k singularities discussed in the previous lecture). As remarked previously, one can gain tremendous insight into Calabi-Yau singularities through algebraic equations: for instance, the A_1 space and the conifold can be described, respectively, via the equations $\sum_{i=1}^3 z_i^2 = 0$ and $\sum_{i=1}^4 z_i^2 = 0$. Sadly, there is no such algebraic tool known for describing singular or nearly singular G_2 holonomy manifolds. And in fact, there are essentially only three known asymptotically conical metrics of G_2 holonomy. The bases of the cones are $\mathbf{CP}^3$, $\frac{SU(3)}{U(1)\times U(1)}$, and $S^3 \times S^3$, but the metrics $d\Omega_6^2$ that appear through (39) in the G_2 holonomy metrics are not the obvious metrics on these spaces (just as, in fact, the metric on T^{11} induced by the Calabi-Yau metric on the conifold is not quite the metric suggested by the coset structure). The three metrics admit isometry groups $SO(5)$, $SU(3)$, and $SU(2)^3$, respectively. (Don't get confused between isometry and holonomy: isometry means that after some transformation the metric is the same as before, whereas holonomy tells us how complicated the transformation properties of vectors are under parallel transport). And at the "tip" of the three respective asymptotically conical metrics, an S^4, or a $\mathbf{CP}^2$, or an S^3, remains finite. See figure 7 for a schematic depiction of the S^4 case.

We may describe the explicit G_2 holonomy metrics in terms only slightly more complicated than the explicit metric (11) for EH_2. For the $SO(5)$ symmetric case, one has

$$ds_7^2 = \frac{dr^2}{1-r^{-4}} + \frac{1}{4}r^2(1-r^{-4})(d\mu^i + \epsilon^{ijk}A^j\mu^k)^2 + \frac{1}{2}r^2 ds_4^2\,, \tag{40}$$

where ds_4^2 is the $SO(5)$ symmetric metric on a unit S^4, the μ^i are three

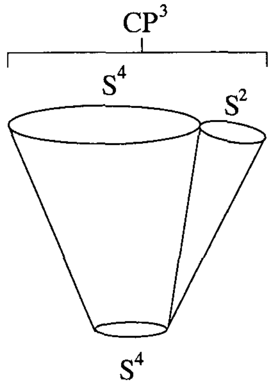


Figure 7. $\mathbf{CP}^3$ a S^2 fibration over S^4, and in the interior, the S^2 shrinks but the S^4 doesn't.

Cartesian coordinates on S^2, subject to $\sum_{i=1}^3 (\mu^i)^2 = 1$, and A^i_μ is an $SU(2)$ gauge field on S^4 carrying unit instanton number. We can be a little more explicit about this gauge field, as follows. S^4 is a space of $SO(4)$ holonomy, but $SO(4) \approx SU(2)_L \times SU(2)_R$, and the spin connection $\omega_{\mu\ ab}$ is decomposable into $SU(2)_L$ and $SU(2)_R$ pieces as $\omega^{L,R}_{\mu\ ab} = \omega_{\mu\ cd}\left(\delta^c_a \delta^d_b \pm \frac{1}{2}\epsilon^{cd}{}_{ab}\right)$. The gauge field A^i_μ can be taken proportional to $\sigma^i_{ab}\omega^L_{\mu\ ab}$, where σ^i_{ab} are the Pauli matrices. The $\mathbf{CP}^2$ case is identical to the above discussion, only one takes ds_4^2 to be the $SU(3)$ symmetric metric on a $\mathbf{CP}^2$ whose size is such that the Ricci curvature is three times the metric (as is true for a unit S^4). Clearly, when $r = 1$, the S^2 part of the metric shrinks to nothing, while the S^4 or $\mathbf{CP}^2$ remains finite. Topologically, the whole space is a bundling of $\mathbf{R}^3$ over S^4 or $\mathbf{CP}^2$, and the base is the corresponding S^2 bundle over S^4 or $\mathbf{CP}^2$.

The $SU(2)^3$ symmetric case is actually more "elementary," in the sense that we do not need to discuss gauge fields. The metric is

$$ds_7^2 = \frac{dr^2}{1 - r^{-3}} + \frac{1}{9}r^2(1 - r^{-3})(\nu_1^2 + \nu_2^2 + \nu_3^2) + \frac{r^2}{12}(\sigma_1^2 + \sigma_2^2 + \sigma_3^2) \quad (41)$$

where $\nu_i = \Sigma_i - \frac{1}{2}\sigma_i$, and Σ_i and σ_i are left-invariant one-forms on two different S^3's. Clearly, only one of these S^3's stays finite as $r \to 1$. Topologically, the whole space is a bundling of $\mathbf{R}^4$ over S^3. Any G_2 holonomy metric can be rigidly rescaled without changing its holonomy group: thus we can claim to have exhibited three one-parameter families of G_2 holonomy metrics, each parametrized by the S^4, or $\mathbf{CP}^2$, or S^3, that remains unshrunk. A perfectly conical metric has an isometry under scaling:

$dr^2 + r^2 d\Omega_6^2 \to d(\Omega r)^2 + (\Omega r)^2 d\Omega_6^2$ for any constant factor Ω. Thus the asymptotics of the rescaled metrics is always the same. And the limit where the unshrunk space at the center goes to zero volume is an exactly conical metric. Considerably more detail on these G_2 holonomy metrics can be found in the original papers.[10,11]

It may seem that our discussion of G_2 holonomy metrics is remarkably unenlightening and difficult to generalize. This is true! Despite more than 15 years since the discovery of the metrics (40) and (41), there are few generalizations of them, and little else known about explicit G_2 holonomy metrics. One interesting generalization of (41) is the discovery of less symmetric versions where, as with Taub-NUT space, there is a $U(1)$ fiber which remains finite at infinity.[12] Nevertheless, there are several generally useful observations to make at this point:

- G_2 holonomy implies Ricci flatness. A mathematically rigorous proof is straightforward, but a nice physical argument is that G_2 holonomy on M_7 implies unbroken supersymmetry for eleven-dimensional supergravity on $\mathbf{R}^{3,1} \times M_7$ with $G_{(4)} = 0$; and supersymmetry implies the equations of motion, which for $G_{(4)} = 0$ are precisely Ricci flatness. Ricci flatness is a common feature of special holonomy manifolds: $SU(n)$ and $Spin(7)$ holonomy manifolds are also necessarily Ricci-flat; but $U(n)$ holonomy manifolds are not.
- The condition $\nabla_\mu \varphi_{\alpha\beta\gamma} = 0$ can be shown to be *equivalent* to the apparently *weaker* condition $d\varphi = 0 = d * \varphi$. These first order equations can be considerably easier to solve than $R_{\mu\nu} = 0$. The three-forms for each of the three "classical" G_2 holonomy metrics are explicitly known, but we would not gain much from exhibiting their explicit forms.
- The three-form φ, as well as its Hodge dual $*\varphi$, are calibrations. Examples of calibrated three- and four-cycles are the unshrunk S^3, S^4, and $\mathbf{CP}^2$ at $r = 1$ in the metrics (40) and (41). An M2-brane on the unshrunk S^3 would be a supersymmetric instanton in M-theory, similar to the worldsheet instantons arising from strings on holomorphic curves. An exploration of such instantons (including their zero modes) can be found in Ref. 13
- M-theory has a 3-form potential, C. Just as we formed $J + iB$ in string theory, so we can form $\varphi + iC$, and then $\int_{S^3}(\varphi + iC)$ is the analog of a complexified Kahler parameter $\int_{N^A}(J + iB)$. As an ex-

ample of the use of this analogy, one may show that $M2-brane$ instantons make a contribution to the superpotential whose dominant behavior is $\exp\left\{-\tau_{M2}\int_{S^3}(\varphi+iC)\right\}$. Perturbative corrections to the classical superpotential are forbidden by the usual holomorphy plus Peccei-Quinn symmetry argument. However, in contrast to the case of type II superstrings, or heterotic strings with the standard embedding, where the Kahler potential could be related to a holomorphic prepotential, it is difficult to say anything systematic about the Kahler potential for G_2 compactifications: no "hidden" supersymmetry is available, and perturbative corrections at all orders seem to be allowed.

Recall that after discussing the A_1 singularity in detail, we were able to go on to construct a smooth, compact $SU(2)$ holonomy manifold by resolving A_1 singularities of an orbifold of T^4 by a discrete subgroup of $SU(2)$. Around each fixed point, we cut out little regions of the orbifold, whose local geometry was $B^4/\mathbf{Z}_2$ (B^4 being a unit ball), and we replaced them by cut-off copies of the Eguchi-Hansen space EH_2. Smoothing out the small discontinuities in the metric at the joining points, without losing $SU(2)$ holonomy, was an interesting subtlety that we left for the mathematical literature. It turns out that a very similar strategy suffices to construct smooth compact G_2 holonomy manifolds. This is called the Joyce construction, and it was the way in which the first explicit examples of compact G_2 holonomy manifolds were found.[4,5] We start with an orbifold T^7/Γ, where T^7 is the square unit torus parametrized by $\vec{x}=(x^1,\ldots,x^7)$, and Γ is a discrete subgroup of G_2, to be specified below. Γ has a set of fixed points S which, in the upstairs picture, is locally a three-dimensional submanifold of T^7. Each fixed point is an A_1 singularity. The key step is to replace $S\times B^4/\mathbf{Z}_2$ by $S\times EH_2$, and then argue that after smoothing out the small discontinuities, the resulting smooth manifold has G_2 holonomy.

A particular example of this strategy begins with the discrete subgroup Γ generated by the following three transformations:

$$\begin{aligned}
&\alpha:\vec{x}\to(-x^1,-x^2,-x^3,-x^4,x^5,x^6,x^7)\\
&\beta:\vec{x}\to(-x^1,\tfrac{1}{2}-x^2,x^3,x^4,-x^5,-x^6,x^7)\\
&\gamma:\vec{x}\to(\tfrac{1}{2}-x^1,x^2,\tfrac{1}{2}-x^3,x^4,-x^5,x^6,-x^7)\,.
\end{aligned}\tag{42}$$

These generators have several nice properties which make the Joyce construction work:

- They commute. The group Γ is $\mathbf{Z}_2^3$.
- They preserve a three-form φ of the form (33) (with an appropriate relabellings of the x^i's as y^j's), so indeed the action of Γ induced by (42) on the tangent space of T^7 is a subgroup of the usual action of $G_2 \subset SO(7)$. (This is what we mean, precisely, by $\Gamma \subset G_2$).
- The generators α, β, and γ each individually has a fixed point set in T^7 consisting of 16 T^3's. β and γ act freely on the fixed point set of α, and similarly for the fixed point sets of β and γ.
- The 48 T^3's coming from the fixed point sets of the generators α, β, and γ are disjoint, but the 16 from α are permuted by β and γ, and similarly for the 16 from β and from γ. Thus on the quotient space, S consists of 12 disjoint T^3's.

Since S has 12 disjoint components in the quotient space, we must ensure when replacing $S \times B^4/\mathbf{Z}_2$ by $S \times EH_2$ that all 12 $SU(2)$'s are in the *same* G_2. To this end, we exploit the fact that EH_2 is hyperkahler, which is to say its metric is Kahler with respect to three different complex structures. In practice, what this means is that there exist covariantly constant ω_1, ω_2, and ω_3 (the three possible Kahler forms), which in local coordinates at any given point can be written as

$$\begin{aligned} \omega_1 &= dy^1 \wedge dy^4 + dy^2 \wedge dy^3 \\ \omega_2 &= dy^1 \wedge dy^3 - dy^2 \wedge dy^4 \\ \omega_3 &= dy^1 \wedge dy^2 + dy^3 \wedge dy^4 \,. \end{aligned} \tag{43}$$

(On the singular space $\mathbf{C}^2/\mathbf{Z}_2$, the y^i could be taken as real coordinates for $\mathbf{C}^2$). Now, the cotangent space of any one of the 12 T^3's is spanned by three one-forms: dx^i, dx^j, and dx^k for some choice of i, j, and k. For a correct ordering of i, j, and k, and correct identification of the y coordinates in (43) with the remaining four x coordinates, the form

$$\varphi = \omega_1 \wedge dx^1 + \omega_2 \wedge dx^2 + \omega_3 \wedge dx^3 + dx^1 \wedge dx^2 \wedge dx^3 \tag{44}$$

is precisely the original three-form (33), written at the location of each T^3 in a way which the replacement $B^4/\mathbf{Z}_2 \to EH_2$ clearly preserves. This is the reasoning that allows us to say that the holonomy group is still contained in G_2 after the resolution. As before, we gloss over the subtlety of smoothing out the discontinuities; this is well treated in Joyce's original papers.[4,5] There it is also shown that the moduli space of G_2 metrics is locally $H^3(M_7, \mathbf{R}) = \mathbf{R}^{43}$ for this example. The moduli space of M-theory on this manifold is locally $H^3(M_7, \mathbf{C})$ because of the complexification $\varphi +$

iC. The Kahler potential on this moduli space is probably *hard* to compute beyond the classical level, for the reasons explained above.

Beautiful and impressive though the Joyce construction is, we still seem as yet rather stuck in mathematics land in our study of G_2 holonomy manifolds. There are two main themes in the relation of G_2 holonomy to string theory. One, which we will not discuss, centers on a relationship with strongly coupled gauge theories, developed in Refs. 14, 15, 16. The other, perhaps more obvious relation, is with configurations of D6-branes in type IIA string theory. To begin, we should recall the basic ansatz relating type IIA string theory to M-theory:

$$ds_{11}^2 = e^{-2\Phi/3} ds_{str}^2 + e^{4\Phi/3}(dx^{11} + C_\nu dx^\nu)^2 \,, \tag{45}$$

where C_μ is the Ramond-Ramond one-form of type IIA, and Φ is the dilaton. The classical geometry for $n+1$ flat, parallel D6-branes can be cast in the form (45):

$$\begin{aligned}
& ds_{11}^2 = ds_{\mathbf{R}^{6,1}}^2 + H d\vec{r}^{\,2} + H^{-1}(dx^{11} + \vec{\omega}\cdot d\vec{r})^2 \\
& \nabla \times \vec{\omega} = -\nabla H \qquad e^{\Phi} = H^{-3/4} \qquad H = 1 + \frac{1}{2}\sum_{i=1}^{n+1} \frac{R}{|\vec{r} - \vec{r}_i|} \\
& ds_{str}^2 = H^{-1/2} ds_{\mathbf{R}^{6,1}}^2 + H^{1/2} d\vec{r}^{\,2} \qquad R = g_{str}\sqrt{\alpha'} \,.
\end{aligned} \tag{46}$$

Here $\vec{r}$ parametrizes the three directions perpendicular to the D6-branes, whose centers are at the various $\vec{r}_i$. Since the eleven-dimensional geometry is the direct product of flat $\mathbf{R}^{6,1}$ and multi-center Taub-NUT, the holonomy group is $SU(2)$, and hence 1/2 of supersymmetry is preserved. It is indeed appropriate, since parallel D6-branes preserve this much supersymmetry. A more general observation is that, since D6-branes act as sources only for the metric, the Ramond-Ramond one-form, and the dilaton, and these fields are organized precisely into the eleven-dimensional metric, *any* configuration of D6-branes that solves the equations of motion must lift to a Ricci-flat manifold in eleven dimensions; and if the configuration of D6-branes is supersymmetric, then the eleven-dimensional geometry must have at least one covariantly constant spinor, and hence special holonomy. In particular, if there is a factor of flat $\mathbf{R}^{3,1}$ in the geometry, and some supersymmetry is unbroken the rest of it must be a seven-manifold whose holonomy is contained in G_2. In the example above, the seven-manifold is $\mathbf{R}^3$ times multi-center Taub-NUT.

Before developing this theme further, it seems worthwhile to explore the dynamics of $n+1$ parallel D6-branes, as described in (46), a little further. Recall that we learned in lecture 1 that there are n homologically non-trivial cycles for the $n+1$-center Taub-NUT geometry: topologically, this is identical to a resolved A_n singularity. Thus there exist n harmonic, normalizable two-forms, call them ω^i. These forms are localized near the centers of the Taub-NUT space, and they are the cohomological forms dual to the n non-trivial homology cycles. Furthermore, there is one additional normalizable 2-form on the Taub-NUT geometry, which can be constructed explicitly for $n = 0$, but which owes its existence to no particular topological property. Let us call this form ω^0. If we expand the Ramond-Ramond three-form of type IIA as

$$C_{(3)} = \sum_{i=0}^{n} \omega^i \wedge A_i + \dots \,, \tag{47}$$

where the A_i depend only on the coordinates of $\mathbf{R}^{6,1}$, then each term represents a seven-dimensional $U(1)$ gauge field localized on a center of the Taub-NUT space. This is very appropriate, because there is indeed a $U(1)$ gauge field on each D6-brane: through (47) we are reproducing this known fact from M-theory. Better yet, recall that there are $n(n+1)/2$ holomorphic embeddings of S^2 in a $n+1$-center Taub-NUT space. An M2-brane wrapped on any of these is some BPS particle, and an anti-holomorphic wrapping is its anti-particle. A closer examination of quantum numbers shows that these wrapped M2-branes carry the right charges and spins to be the non-abelian W-bosons that we know should exist: in the type IIA picture they are the lowest energy modes of strings stretched from one D6-brane to another. It's easy to understand the charge quantum number for the case $n = 1$, that is for two D6-branes. The form ω^0 corresponds to what we would call the center-of-mass $U(1)$ of the D6-branes. The form ω^1 is dual to the holomorphic cycle over which we wrap an M2-brane: this is precisely the holomorphic S^2 at $r = \epsilon$ in the Eguchi-Hansen space (11), as discussed after (21). Thus $\int_{S^2} \omega^1 = 1$, which means that an M2-brane on this S^2 does indeed have charge $+1$ under the $U(1)$ photon which we called A_1 in (47). This is the relative $U(1)$ in the D6-brane description, and the wrapped M2-brane becomes a string stretched between the two D6-branes, which does indeed have charge under the relative $U(1)$.

When two D6-branes come together, one of the holomorphic cycles shrinks to zero size, and there is gauge symmetry enhancement from $U(1) \times U(1)$ to $U(2)$. In the generic situation where the D6-branes are separated, the unbroken gauge group is $U(1)^{n+1}$ on account of the Higgs mechanism. This is a pretty standard aspect of the lore on the relation between M-theory and type IIA, but I find it a particularly satisfying result, because it shows that wrapped M2-branes have to be considered on an equal footing with the degrees of freedom of eleven-dimensional supergravity: in this instance, they conspire to generate $U(n)$ gauge dynamics. See figure 8. Such a conspiracy is one of the reasons why we believe the

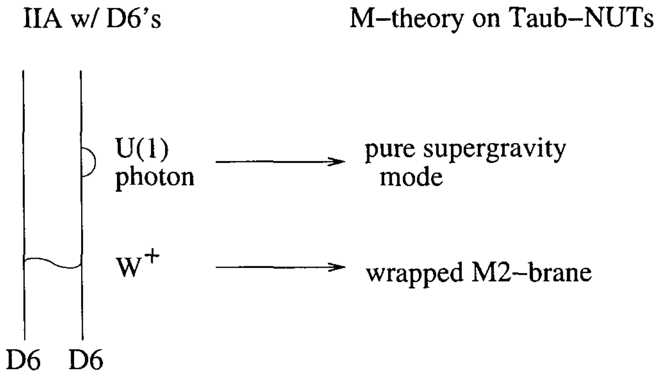


Figure 8. Stretched string degrees of freedom in type IIA lift to very different things in M-theory: a $U(1)$ photon with both its ends on a single D6-brane lifts to a zero-mode of C_{MNP}, whereas a W^+, with one end on one D-brane and the other on another, lifts to a wrapped M2-brane. These states can be tracked reliably from weak to strong string coupling because they are BPS.

type duality between IIA string theory and 11-dimensional M-theory goes beyond the supergravity approximation.

Let's return to our earlier observation that any configuration of D6-branes in type IIA must lift to pure geometry in eleven dimensions. Actually, more is true: any type IIA configuration that involves only the metric, the dilaton, and the Ramond-Ramond one-form should lift to pure geometry in eleven dimensions. This means we can include O6-planes as well as D6-branes. Our focus here, however, will be on D6-branes only. Consider, in particular, a set of D6-branes which all share a common $\mathbf{R}^{3,1}$, which we could regard as our own four dimensions. Assume that the configuration preserves $\mathcal{N} = 1$ supersymmetry in $d = 4$. Then the lift to eleven dimensions should generically be $\mathbf{R}^{3,1}$ times a G_2 holonomy manifold. (We have not entirely ruled out cases where the holonomy group is smaller than G_2, but we expect such configurations to be quite special, if they exist at all).

Suppose, for instance, that the other six dimensions in the type IIA description are non-compact and asymptotically flat (or else compact/curved on a much larger length scale than we are considering for now), and that each D6-brane stretches along some flat $\mathbf{R}^3 \subset \mathbf{R}^6$. When would this configuration preserve $\mathcal{N} = 1$ supersymmetry? The answer to this sort of question was given in one of the early papers on D-branes,[17] and it relies on a fermionic Fock-space trick which is also of use in the study of spinors and differential forms on Calabi-Yau spaces. The best way to state the result of Ref. 17 is to first choose complex coordinates on $\mathbf{R}^6$, call them z^1, z^2, and z^3. Obviously there are many inequivalent ways to form the z^i, but suppose we've made up our mind on one for the moment. Now, the $SU(3)$ acting on the z^i by

$$z^i \to R^i{}_j z^j \qquad \bar{z}^{\bar{i}} \to R^{\dagger}{}^{\bar{i}}{}_{\bar{j}} \bar{z}^{\bar{j}} \,, \tag{48}$$

is obviously a subgroup of all possible $SO(6)$ rotations. Let us construct Dirac gamma matrices obeying $\{\Gamma^{z^i}, \Gamma^{\bar{z}^{\bar{j}}}\} = 2g^{i\bar{j}} \propto \delta^{i\bar{j}}$. Clearly, the Γ^{z^i} and $\Gamma^{\bar{z}^{\bar{j}}}$ are fermionic lowering/raising operators, up to a normalization. We can define a Fock space vacuum ϵ in the spinor representation of the Clifford algebra via $\Gamma^{z^i}\epsilon = 0$ for all i. The full Dirac spinor representation of $SO(6)$ now decomposes under the inclusion $SU(3) \subset SO(6)$ as $\mathbf{8} = \mathbf{1} \oplus \bar{\mathbf{3}} \oplus \mathbf{3} \oplus \mathbf{1}$. The singlets are ϵ itself and $\Gamma^{\bar{z}^1}\Gamma^{\bar{z}^2}\Gamma^{\bar{z}^3}\epsilon$; the $\bar{\mathbf{3}}$ is $\Gamma^{\bar{z}^{\bar{j}}}\epsilon$; and the $\mathbf{3}$ is $\Gamma^{\bar{z}^{\bar{i}}}\Gamma^{\bar{z}^{\bar{j}}}\epsilon$.

With this mechanism in place, we can state and immediately understand the results of Ref. 17: suppose one D6-brane lies along the $\mathbf{R}^3$ spanned by the real parts of z^1, z^2, and z^3. Consider a collection of other D6-branes on $\mathbf{R}^3$'s related by various $SU(3)$ rotations, and, optionally, arbitrary translations. The claim is that this configuration preserves $\mathcal{N} = 1$ supersymmetry. To understand why this is so, we need only recall that the first D6-brane preserves the half of supersymmetry satisfying

$$\tilde{\epsilon}_R = \prod_{i=1}^{3} \left(\Gamma^{z^i} + \Gamma^{\bar{z}^{\bar{i}}} \right) \epsilon_L \,, \tag{49}$$

where $\tilde{\epsilon}_R$ is the right-handed spacetime spinor that comes from the antiholomorphic sector on the worldsheet, and ϵ_L is the left-handed spacetime spinor that comes from the holomorphic sector. The rotated D6-branes also preserve half of supersymmetry, but a different half, namely

$$\tilde{\epsilon}_R = \prod_{i=1}^{3} \left(R^i{}_j \Gamma^{z^j} + R^{\dagger}{}^{\bar{i}}{}_{\bar{j}} \Gamma^{\bar{z}^{\bar{j}}} \right) \epsilon_L \,. \tag{50}$$

Some supersymmetry is preserved by the total collection of D6-branes iff we can find simultaneous solutions to (49) and (50) for the various $SU(3)$ rotations $R^i{}_j$ that appear. In fact, if ϵ_L is the Fock space vacuum ϵ tensored with an arbitrary chiral spinor in four-dimensions, and $\tilde{\epsilon}_R$ is $\Gamma^{\bar{z}^1}\Gamma^{\bar{z}^2}\Gamma^{\bar{z}^3}\epsilon$ times the same four-dimensional chiral spinor, then (49) is obviously satisfied; but also (50) is satisfied, because

$$\tilde{\epsilon}_R = \prod_{i=1}^{3} R^{\dagger\,\bar{i}}{}_{\bar{j}}\Gamma^{\bar{z}^{\bar{j}}}\epsilon_L = (\det R^{\dagger\,\bar{i}}{}_{\bar{j}})\Gamma^{\bar{z}^1}\Gamma^{\bar{z}^2}\Gamma^{\bar{z}^3}\epsilon_L \tag{51}$$

when ϵ_L is as stated above; and $\det R^{\dagger\,\bar{i}}{}_{\bar{j}} = 1$ for an $SU(3)$ matrix.

More in fact was shown in Ref. 17: it turns out that $\mathcal{N} = 1$ $d = 4$ chiral matter lives at the intersection of D6-branes oriented at unitary angles in the manner discussed in the previous paragraph. We will not here enter into the discussion in detail, but merely state that the GSO projection that acts on string running from one D6-brane to another one at a unitary angle from the first winds up projecting out the massless fermions of one chirality and leaving the other. Clearly, such strings carry bifundamental charges under the gauge group on the D6-branes they end on; so if one intersects, say, a stack of two coincident D6-branes with another stack of three, the four-dimensional dynamics of the intersection is $U(3) \times U(2)$ gauge theory with chiral "quarks" in the $(\mathbf{3}, \bar{\mathbf{2}})$. This is remarkably similar to the Standard Model! See figure 9. After these lectures were given,

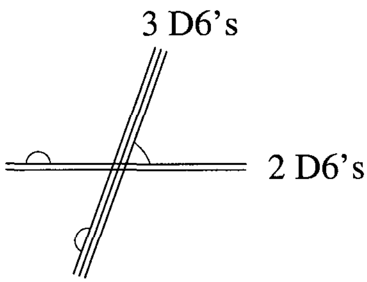


Figure 9. Three coincident D6-branes intersecting two coincident D6-branes at a unitary angle. Strings are shown that give rise to $U(3) \times U(2)$ gauge fields (on the respective D6-brane worldvolumes) and chiral fermions in the $(\mathbf{3}, \bar{\mathbf{2}})$ (at the intersection).

work has appeared[18] where explicit compact constructions are given, involving D6-branes and O6-planes, whose low-energy dynamics includes the supersymmetric Standard Model (as well as some other, possibly innocuous, extra degrees of freedom). See also the related work.[19] As is clear from the previous discussion, such constructions lift in M-theory to configurations which involve only the metric, not $G_{(4)}$. Only there are singularities in the eleven-dimensional metric where D6-branes cross. As shown in Ref. 16, the singularity in the G_2 holonomy lift of two D6-branes intersecting at unitary angles is precisely the cone over $\mathbf{CP}^3$ exhibited in (40) (or, properly, the limit of this geometry where the S^4 shrinks); whereas the singularity in the G_2 holonomy lift of of three D6-branes (all at different, unitarily related angles) is the cone over $\frac{SU(3)}{U(1)\times U(1)}$ (of a form very similar to (40), as discussed above). More detail can be found in Ref. 16, and also in the earlier work,[20] on the D6-brane interpretation of resolving the conical singularity. Oddly, it seems that the generalizations of these G_2 holonomy cones to any number of D6-branes, intersecting all at different angles, is not known. Also, the full geometry interpolating between the near-intersection region, where the metric is nearly conical, and the asymptotic region, where the metric is nearly Taub-NUT close to any single D6-brane, is not known. Finding either sort of generalization of the existing result (40) would be very interesting, and possibly useful for studying the dynamics of M-theory compactifications.

A word of explanation is perhaps in order on why we have focused so exclusively on M-theory geometries with $G_{(4)} = 0$. Really there are two. First, on a compact seven-manifold, there are rather tight constraints on how $G_{(4)}$ may be turned on—see for example Ref. 21. The main context of interest where non-zero $G_{(4)}$ seems necessary is compactifications of Horava-Witten theory: there is seems impossible to satisfy the anomaly condition $\operatorname{tr} R \wedge R - \frac{1}{2} \operatorname{tr} F \wedge F = 0$ on each E_8 plane individually, so some $G_{(4)}$ is needed to "soak out" the anomaly. The second reason to consider M-theory geometries first with $G_{(4)} = 0$ is that, in an expansion in the gravitational coupling, the zeroth order equations of motion are indeed Ricci-flatness. For instance, in Horava-Witten compactifications, the necessary $G_{(4)}$ is only a finite number of Dirac units through given four-cycles. As long as only finitely many quanta of $G_{(4)}$ are turned on, and provided the compactification scale is well below the eleven-dimensional Planck scale, one would expect to learn much by starting out ignoring $G_{(4)}$ altogether. This is not to say that nonzero $G_{(4)}$ won't have some interesting and novel effects: see for example.[22,23] It is fair to say that our understanding of M-theory compactifications is in a very primitive state, as compared, for instance, to compactifications of type II or heterotic strings. It is to be hoped that this topic will flourish in years to come.

3.1. *Addendum: further remarks on intersecting D6-branes*

My original TASI lectures ended here, but in view of the continuing interest in G_2 compactifications of M-theory, it seems worthwhile to present a little more detail on intersecting D6-branes and their M-theory lift. This in fact is a subject where rather little is known, so I will in part be speculating about what might be accomplished in further work.

First it's worthwhile to reconsider the work of Ref. 17 in light of a particular calibration on $\mathbf{R}^6$. Consider the complex three-form $\Omega = dz^1 \wedge dz^2 \wedge dz^3$, where the z^i are, as before, complex coordinates on $\mathbf{R}^6$ such that the metric takes the standard Kahler form. Ω is called the holomorphic three-form, or the holomorphic volume form, and if space had permitted, some elegant results could have been presented about how the analogous form on a curved complex manifold relates to its complex structure, as well as to covariantly constant spinors, if they exist. Our purpose here is to note that $\mathrm{Re}\,\Omega$ is a calibration, in the sense explained in section 2.2. Clearly $\mathrm{Re}\,\Omega$ calibrates the plane in the $\mathrm{Re}\,z^1$, $\mathrm{Re}\,z^2$, $\mathrm{Re}\,z^3$ directions. Any $SU(3)$ change of the coordinates z^i preserves Ω; in fact such a map is the most general linear map that does so. So it is not hard to convince oneself that all planes related to the one we first mentioned by a $SU(3)$ rotation are also calibrated by $\mathrm{Re}\,\Omega$. One can now concisely restate the result that D6-branes stretched on $\mathbf{R}^{3,1}$ must be at unitary angles in the remaining $\mathbf{R}^6$ to preserve supersymmetry: supersymmetric intersecting D6-branes must all be calibrated by $\mathrm{Re}\,\Omega$, for some suitable choice of the z^i. Choice of the z^i here includes the ability to rotate $\mathrm{Re}\,\Omega$ by a phase. A more general truth is that supersymmetric D6-branes on a three-cycle of a Calabi-Yau manifold are those calibrated by $\mathrm{Re}\,\Omega$. Such three-cycles are called special lagrangian manifolds.[f]

It can be shown (see for example Ref. 16) that the G_2 cones over $\mathbf{CP}^3$ and $\frac{SU(3)}{U(1)\times U(1)}$ are limits of the M-theory lift of two and three D6-branes intersecting at a common point and at unitary angles. It is even known how the G_2 resolutions of these cones corresponds to deformed world-volumes of the D6-branes: for instance, for the cone over $\mathbf{CP}^3$, the D6-brane world-volume has an hour-glass shape which is topologically $S^2 \times \mathbf{R}^2$.

We can describe in a straightforward fashion, though without mathematical rigor, how supersymmetric D6-brane configurations spanning $\mathbf{R}^{3,1}$

[f]We have not been quite precise in the main text: more accurately, D6-branes should wrap special lagrangian manifolds, and such a manifold has the properties that both the Kahler form and $\mathrm{Im}\,\Omega$ pull back to zero, as well as being calibrated by $\mathrm{Re}\,\Omega$.

and a special lagrangian manifold in $\mathbf{R}^6$ can be lifted to manifolds of G_2 holonomy in eleven dimensions. As explained in section 3, the eleven-dimensional lift of a single flat D6-brane is Taub-NUT space times $\mathbf{R}^{6,1}$, which we will conveniently write as $\mathbf{R}^{3,1} \times \mathbf{R}^3$. Intuitively speaking, we should be able to lift any D6-brane configuration with no coincident or intersecting D6-branes, just by making an affine approximation to the curving D6-brane world volume at each point, and lifting to Taub-NUT times the D6-brane world-volume times $\mathbf{R}^{3,1}$. From now own let's ignore the $\mathbf{R}^{3,1}$ part. Then in the seven remaining dimensions, the geometry far from any brane is $\mathbf{R}^6 \times S^1$. Near the D6-brane world volume, we cut out a region in $\mathbf{R}^6$ that surrounds the brane, and since locally this region is $\mathbf{R}^3 \times B^3$, we can replace $\mathbf{R}^3 \times B^3 \times S^1$ in the seven dimensional geometry by $\mathbf{R}^3$ times a cut-off Taub-NUT. Gluing in the Taub-NUT space should cause only very small discontinuities, which hopefully could be erased through some real analysis.

There is a meaningful point to check, though: in our putative almost-G_2 manifold, formed by gluing into $\mathbf{R}^6 \times S^1$ a cut-off Taub-NUT snaking along what was the D6-brane world volume, we'd like to see that the holonomy on different parts of the "snake" is always (nearly) contained in the same G_2. To this end, we need to write down a covariantly constant three-form φ for $\mathbf{R}^3$ times Taub-NUT. This can be done in different ways, because there are different embeddings of $SU(2)$ into G_2. Let x^1, x^2, and x^3 be coordinates on $\mathbf{R}^3$, and let y^1, y^2, y^3, and x^{11} be coordinates on Taub-NUT. One choice suggested by the discussion in section 3 is to use the fact that Taub-NUT is hyperkahler, and construct

$$\varphi = \omega_1 \wedge dx^1 + \omega_2 \wedge dx^2 + \omega_3 \wedge dx^3 + dx^1 \wedge dx^2 \wedge dx^3 \,, \tag{52}$$

where the ω_i are the Kahler structures. This is *not* the choice of φ that we will be particularly interested in. Instead, we want a φ which will have some transparent connection with the complex structure of $\mathbf{R}^6$. If we choose complex coordinates $z^j = x^j + iy^j$ for $j = 1,2,3$, then a D6-brane in the x^1-x^2-x^3 plane (or any $SU(3)$ image of it) is calibrated by $\mathrm{Re}\,\Omega$. The standard Kahler form on $\mathbf{R}^6 = \mathbf{C}^3$ is

$$J = \frac{i}{2}\left(dz_1 \wedge d\bar{z}_1 + dz_2 \wedge d\bar{z}_2 + dz_3 \wedge d\bar{z}_3\right) \,, \tag{53}$$

and one can readily verify that

$$\varphi_0 = \mathrm{Re}\,\Omega - J \wedge dx^{11} \tag{54}$$

is a G_2-structure on $\mathbf{R}^6 \times S^1$ (whose holonomy is certainly a subgroup in G_2). Actually, much more is true: the φ_0 in (54) can be constructed in the same way for any $CY_3 \times S^1$, and it represents an inclusion of $SU(3)$ in G_2. In fact the Joyce construction we explained in section 3 is believed to generalize to $\mathbf{Z}_2$ orbifolds of $CY_3 \times S^1$, acting with two fixed points on the S^1 and as an anti-holomorphic involution on the CY_3.

The obvious vielbein for $X = \mathbf{R}^3 \times$ (Taub-NUT) is

$$\begin{aligned} e^1 &= dx^1 \quad e^2 = dx^2 \quad e^3 = dx^3 \\ e^4 &= \sqrt{1+H}dy^1 \quad e^5 = \sqrt{1+H}dy^2 \quad e^6 = \sqrt{1+H}dy^3 \\ e^7 &= \frac{1}{\sqrt{1+H}}\left(dx^{11}+V\right) , \end{aligned} \tag{55}$$

where

$$dV = *_y dH \qquad \text{and} \quad H = \frac{R}{2|\vec{y}|} \, . \tag{56}$$

Here $*_y$ represents the Hodge dual in the y^j directions only, and $x^{11} \sim x^{11} + 2\pi R$. One can now modify φ_0 slightly to give a G_2-structure on X:

$$\begin{aligned} \varphi &= dx^1 \wedge dx^2 \wedge dx^3 - (1+H)\big(dx^1 \wedge dy^2 \wedge dy^3 + dy^1 \wedge dx^2 \wedge dy^3 \\ &\qquad + dy^1 \wedge dy^2 \wedge dx^3\big) - J \wedge (dx^{11} + V) \\ &= \varphi_0 - H\left(dx^1 \wedge dy^2 \wedge dy^3 + dy^1 \wedge dx^2 \wedge dy^3 + dy^1 \wedge dy^2 \wedge dx^3\right) - J \wedge V \, . \end{aligned} \tag{57}$$

Now, $\varphi \to \varphi_0$ as $|\vec{y}| \to \infty$. The key point is that φ_0 is invariant under $SU(3)$ changes of the coordinates z^i: this is so because both Ω and J are $SU(3)$ singlets. So the φ we would construct locally at any point along the lift of the D6-brane world-volume asymptotes to the *same* φ_0. This is the desired verification that the holonomy of the entire approximation to the seven-dimensional manifold is (nearly) in the same G_2. I have included the "(nearly)" because of the errors in the affine approximation to the D6-brane world-volume. This error can be uniformly controlled if there are no coincident or intersecting D6-branes. A way to think about it is that we make R much smaller than the closest approach of one part of the world-volume to another.

A remarkable fact is that the deformation of the G_2-structure, $\varphi - \varphi_0$, in (57), is *linear* in H. This is true despite the fact that the vielbein and the metric are complicated non-linear functions of H. It is tempting to conjecture that an appropriate linear modification of φ_0, along the lines of (57), will be an *exact* G_2 structure on the whole seven-manifold, even in cases where D6-branes intersect (of course, in such a case one must exclude the singularity right at the intersection). However, we have been unable to verify this.[g] Knowing the covariantly constant three-form suffices to determine the metric, via the explicit formula

$$\begin{aligned} g_{\mu\nu} &= (\det s_{\mu\nu})^{-1/9} s_{\mu\nu} \qquad \text{where} \\ s_{\mu\nu} &= \frac{1}{144} \varphi_{\mu\lambda_1\lambda_2} \varphi_{\nu\lambda_3\lambda_4} \varphi_{\lambda_5\lambda_6\lambda_7} \epsilon^{\lambda_1\lambda_2\lambda_3\lambda_4\lambda_5\lambda_6\lambda_7}, \end{aligned} \tag{58}$$

where $\epsilon^{1234567} = \pm 1$, and the sign is chosen to make $s_{\mu\nu}$ positive definite. Convincing oneself of the truth of this formula (which appeared quite early, see for instance Ref. 10) is pretty straightforward: $s_{\mu\nu}$ is symmetric, but scales the wrong way under rigid rescalings of the manifold to be a metric. The determinant factor in the definition of $g_{\mu\nu}$ corrects this problem. The strategy of finding φ first and then deducing $g_{\mu\nu}$ has been of use in a recent investigation in the string theory literature.[24]

Acknowledgments

I thank A. Brandhuber, J. Gomis, S. Gukov, and E. Witten for extensive discussions about G_2 holonomy, and J. Schwarz for a careful reading of the manuscript. I would like to express my special appreciation to K. T. Mahanthappa for his many years of dedication to TASI, and to Kathy Oliver for her help in making the school function. This work was supported in part by the DOE under grant DE-FG03-92ER40701.

References

1. T. Eguchi, P. B. Gilkey, and A. J. Hanson, "Gravitation, gauge theories and differential geometry," *Phys. Rept.* **66** (1980) 213.
2. M. B. Green, J. H. Schwarz, and E. Witten, *Superstring theory. vol. 2: loop amplitudes, anomalies and phenomenology.* Cambridge University Press, 1987.
3. J. Polchinski, *String theory. Vol. 2: Superstring theory and beyond.* Cambridge University Press, 1998.
4. D. Joyce, "Compact Riemannian 7-manifolds with holonomy G_2. Part I," *J. Diff. Geom.* **43** (1996) 291.
5. D. Joyce, "Compact Riemannian 7-manifolds with holonomy G_2. Part II," *J. Diff. Geom.* **43** (1996) 329.
6. D. Joyce, "Lectures on special Lagrangian geometry," `math.dg/0111111`.
7. B. R. Greene, "String theory on Calabi-Yau manifolds," `hep-th/9702155`.
8. N. Hitchin, "The geometry of three-forms in six and seven dimensions," `math.dg/0010054`.

[g] We thank I. Mitra, O. Evnin, and A. Brandhuber for discussions on this point.

9. E. Witten, "Fermion quantum numbers in Kaluza-Klein theory," in *Quantum field theory and the fundamental problems of physics*, pp. 227–277, Shelter Island 1983, proceedings. Also in Appelquist, T. (ed.) et al.: *Modern Kaluza-Klein theories*, 438-511.
10. R. Bryant and S. Salamon, "On the construction of some complete metrics with exceptional holonomy," *Duke Math. J.* **58** (1989) 829.
11. G. W. Gibbons, D. N. Page, and C. N. Pope, "Einstein metrics on S**3 R**3 and R**4 bundles," *Commun. Math. Phys.* **127** (1990) 529.
12. A. Brandhuber, J. Gomis, S. S. Gubser, and S. Gukov, "Gauge theory at large N and new G(2) holonomy metrics," *Nucl. Phys.* **B611** (2001) 179–204, `hep-th/0106034`.
13. J. A. Harvey and G. W. Moore, "Superpotentials and membrane instantons," `hep-th/9907026`.
14. C. Vafa, "Superstrings and topological strings at large N," *J. Math. Phys.* **42** (2001) 2798–2817, `hep-th/0008142`.
15. B. S. Acharya, "On realising N = 1 super Yang-Mills in M theory," `hep-th/0011089`.
16. M. Atiyah and E. Witten, "M-theory dynamics on a manifold of G(2) holonomy," `hep-th/0107177`.
17. M. Berkooz, M. R. Douglas, and R. G. Leigh, "Branes intersecting at angles," *Nucl. Phys.* **B480** (1996) 265–278, `hep-th/9606139`.
18. M. Cvetic, G. Shiu, and A. M. Uranga, "Three-family supersymmetric standard like models from intersecting brane worlds," *Phys. Rev. Lett.* **87** (2001) 201801, `hep-th/0107143`.
19. B. Acharya and E. Witten, "Chiral fermions from manifolds of G(2) holonomy," `hep-th/0109152`.
20. S. Kachru and J. McGreevy, "M-theory on manifolds of G(2) holonomy and type IIA orientifolds," *JHEP* **06** (2001) 027, `hep-th/0103223`.
21. K. Becker and M. Becker, "Compactifying M-theory to four dimensions," *JHEP* **11** (2000) 029, `hep-th/0010282`.
22. K. Becker and M. Becker, "M-Theory on Eight-Manifolds," *Nucl. Phys.* **B477** (1996) 155–167, `hep-th/9605053`.
23. M. Cvetic, G. W. Gibbons, H. Lu, and C. N. Pope, "Supersymmetric non-singular fractional D2-branes and NS-NS 2-branes," *Nucl. Phys.* **B606** (2001) 18–44, `hep-th/0101096`.
24. A. Brandhuber, "G(2) holonomy spaces from invariant three-forms," `hep-th/0112113`.

This page intentionally left blank

JOANNE L. HEWETT

This page intentionally left blank

PHENOMENOLOGY OF EXTRA DIMENSIONS

JOANNE L. HEWETT
Stanford Linear Accelerator Center
Stanford, CA 94309
E-mail: hewett@slac.stanford.edu

String theorists and their predecessors have studied the effects of higher dimensional spacetimes at very short distances for decades. New theories, which offer a novel approach to the hierarchy problem, have proposed that extra dimensions might be visible at larger distances, comparable to the TeV scale. This allows for experiment to explore their presence. Here, we review the principal features of these scenarios and their resulting phenomenology.

1. Introduction

In particle physics, 4-dimensional Minkowski spacetime is the underlying fundamental framework under which the laws of nature are formulated and interpreted. Relativistic quantum fields exist in spacetime, interactions occur at spacetime points and the laws governing these fields and their interactions are constructed using weighted averages over their spacetime histories. According to the general theory of relativity, fluctuations of the spacetime curvature provide gravitational dynamics. Indeed, experiments show evidence for the predictions of general relativity and hence that spacetime is dynamical at very long length scales. However, gravitational dynamics have yet to be probed at short distances, and it is possible that they are quite different from that implied by a simple extrapolation of the long range theory.

Early attempts to extend general relativity in order to unify gravity and electromagnetism within a common geometrical framework trace back to Gunnar Nordström (1914),[1] Theodor Kaluza (1921) and Oscar Klein (1926).[2] They proposed that unification of the two forces occurred when spacetime was extended to a five dimensional manifold and imposed the condition that the fields should not depend on the extra dimension. A difficulty with the acceptance of these ideas at the time was a lack of both experimental implications and a quantum description of gravitational dynamics. Today, one of the most striking requirements of modern string theory, which incorporates both gauge theories and gravitation, is that there must be six or seven extra spatial dimensions. Otherwise the theory is anomalous. Recently, concepts developed within string theory have led to new phenomenological ideas which relate the physics of extra dimensions to observables in a variety of physics experiments.

These new theories have been developed to address the hierarchy problem, *i.e.*, the large disparity between the scale where electroweak symmetry breaking occurs and the traditional scale of gravity defined by the Planck scale. The source of physics which generates and stabilizes this sixteen order of magnitude difference between the two scales is unknown and represents one of the most puzzling aspects of nature. The novel approach to this long-standing problem proposed in these recent theories is that the geometry of extra spatial dimensions may be responsible for the hierarchy: the gravitational field lines spread throughout the full higher dimensional space and modify the behavior of gravity. Indeed, the fact that gravity has yet to be measured at energy scales much above 10^{-3} eV in laboratory experiments admits for the possibility that at higher energies gravity behaves quite differently than expected. The first scenario of this type to be proposed[3] suggested that the apparent hierarchy between these two important scales of nature is generated by a large volume of the extra dimensions, while in a later theoretical framework[4,5] the observed hierarchy results from a strong curvature of the extra dimensional space. If new dimensions are indeed relevant to the source of the hierarchy, then they should provide detectable signatures at the electroweak scale. These physics scenarios with additional dimensions hence afford concrete and distinctive phenomenological predictions for high energy colliders, as well as producing observable consequences for astrophysics and short-range gravity experiments.

Theoretical frameworks with extra dimensions have some general features, we briefly introduce a couple of the principal properties here. In most scenarios, our observed 3-dimensional space is a 3-brane (sometimes called a wall), where the terminology is derived from a generalization of a 2-dimensional membrane. This 3-brane is embedded in a higher D-dimensional spacetime, $D = 3 + \delta + 1$, with δ extra spatial dimensions which are orthogonal to our 3-brane. The higher D-dimensional space is known as the "bulk". String theory contains branes upon which particles can be naturally confined or localized.[6] In a general picture, branes carry the Standard Model (SM) gauge charges and the ends of open strings are stuck to the branes and represent the Standard Model fields. Fields, such as gravitons, which do not carry Standard Model gauge charges correspond to closed strings and may pop off the brane and propagate throughout the bulk.

The picture is thus one where matter and gauge forces are confined to our 3-dimensional subspace, while gravity propagates in a higher dimensional volume. In this case, the Standard Model fields maintain their usual behavior, however, the gravitational field spreads throughout the full $3 + \delta$ spatial volume. Conventional wisdom dictates that if the additional dimensions are too large, this would result in observable deviations from Newtonian gravity. The extra dimensional space must then be *compactified*, *i.e*, made finite. However, in some alternative theories ,[5,7] the extra dimensions are infinite and the gravitational deviations are suppressed by other means.

If the additional dimensions are small enough, the Standard Model fields are phenomenologically allowed to propagate in the bulk. This possibility allows for new model-building techniques to address gauge coupling unification,[8] supersymmetry breaking,[9,10] the neutrino mass spectrum,[11] and the fermion mass hierarchy.[12] Indeed, the field content which is allowed to propagate in the bulk, as well as the size and geometry of the bulk itself, varies between different models.

As a result of compactification, fields propagating in the bulk expand into a series of states known as a Kaluza-Klein (KK) tower, with the individual KK excitations being labeled by mode numbers. Similar to a particle in a box, the momentum of the bulk field is then quantized in the compactified dimensions. For an observer trapped on the brane, each quanta of momentum in the compactified volume appears as a KK excited state with mass $m^2 = \vec{p}_\delta^{\,2}$. This builds a KK tower of states, where each state carries identical spin and gauge quantum numbers. If the additional dimensions are infinite instead of being compactified, the δ-dimensional momentum and resulting KK spectrum is continuous.

More technically, in the case where gravity propagates in a compactified bulk, one starts from a D-dimensional Einstein-Hilbert action and performs a KK expansion about the metric field of the higher dimensional spacetime. The graviton KK towers arise as a solution to the linearized equation of motion of the metric field in this background.[13] The resulting 4-dimensional fields are the Kaluza-Klein modes. Counting the degrees of freedom within the original higher dimensional metric, the reduction of a spin-2 bulk field results in three distinct classes of towers of KK modes: symmetric tensor, vector fields and scalar fields. The KK zero-mode fields are massless, while the excitation states acquire mass by 'eating' lower spin degrees of freedom. This results in a single 5-component tensor KK tower of massive graviton states, $\delta - 1$ gauge KK towers of massive vector states, and $\delta(\delta-1)/2$ scalar towers. The zero-mode scalar states are radius moduli fields associated with the size of the additional dimensions.

A generalized calculation of the action for linearized gravity in D dimensions can be used to compute the effective 4-dimensional theory. The spin-2 tower of KK states couples to Standard Model fields on the brane via the conserved symmetric stress-energy tensor. The spin-1 KK tower does not induce interactions on the 3-brane. The scalar KK states couple to the Standard Model fields on the brane via the trace of the stress-energy tensor.

The possible experimental signals for the existence of extra dimensions are: (i) the direct or indirect observation of a KK tower of states, or (ii) the observation of deviations in the inverse-square law of gravity in short-range experiments. The detailed properties of the KK states are determined by the geometry of the compactified space and their measurement would reveal the underlying geometry of the bulk.

We now discuss each of the principal scenarios and how they may be probed in experiment.

2. Large Extra Dimensions

The large extra dimensions scenario postulated by Arkani-Hamed, Dimopoulos and Dvali (ADD)[3] makes use of the string inspired braneworld hypothesis. In this model, the Standard Model gauge and matter fields are confined to a 3-dimensional brane that exists within a higher dimensional bulk. Gravity alone propagates in the δ extra spatial dimensions which are compactified. Gauss' Law relates the Planck scale of the effective 4-d low-energy theory, $M_{\rm Pl}$, to the scale where gravity becomes strong in the $4+\delta$-dimensional spacetime, M_D, through the volume of the compactified dimensions V_δ via

$$M_{\rm Pl}^2 = V_\delta M_D^{2+\delta} \, . \tag{1}$$

Taking $M_D \sim$ TeV, as assumed by ADD, eliminates the hierarchy between $M_{\rm Pl}$ and the electroweak scale. $M_{\rm Pl}$ is simply generated by the large volume of the higher dimensional space and is thus no longer a fundamental scale. The hierarchy problem is now translated to the possibly more tractable question of why the compactification scale of the extra dimensions is large.

If the compactified dimensions are flat, of equal size, and of toroidal form, then $V_\delta = (2\pi R_c)^\delta$. For $M_D \sim$ TeV, the radius R_c of the extra dimensions ranges from a fraction of a millimeter to $\sim$ 10 fermi for δ varying between 2 and 6. The compactification scale ($1/R_c$) associated with these parameters then ranges from $\sim 10^{-4}$ eV to tens of MeV. The case of one extra dimension is excluded as the corresponding dimension (of size $R_c \approx 10^{11}$ m) would directly alter Newton's law at solar-system distances. Our knowledge of the electroweak and strong forces extends with great precision down to distances of order 10^{-15} mm, which corresponds to $\sim (100 \text{ GeV})^{-1}$. Thus the Standard Model fields do not feel the effects of the large extra dimensions present in this scenario and must be confined to the 3-brane. Therefore in this model only gravity probes the existence of the extra dimensions [a].

The existence of large extra dimensions would affect a broad range of physical processes. Their presence may be detected in tests of short range gravity, astrophysical considerations, and collider experiments. We now review each of these in turn.

2.1. *Short Range Test of Gravity*

Until very recently, the inverse square force law of Newtonian gravity had been precisely tested only down to distances of order a centimeter.[14–18] Such tests are performed by short range gravity experiments that probe new interactions by searching for deviations from Newtonian gravity at small distances. There are several parameterizations which describe these potential deviations;[19] the one most widely used by experiments is that where the classical gravitational potential is expanded to include a Yukawa interaction:

$$V(r) = -\frac{1}{M_{\rm Pl}^2}\frac{m_1 m_2}{r}(1 + \alpha e^{-r/\lambda}) \ . \qquad (2)$$

Here, r is the distance between two masses m_1 and m_2 and is fixed by the experimental apparatus, α is a dimensionless parameter relating the strength of the additional Yukawa interaction to that of gravity, and λ is the range of the new interaction.

The two-body potential given by Gauss' Law in the presence of addi-

[a]However, any Standard Model singlet field, *e.g.*, a right handed neutrino, could also be in the bulk in this scenario.

tional dimensions (for distances $r < R_c$) is expressed as:

$$V(r) = -\frac{1}{8\pi M_D^{2+\delta}} \frac{m_1 m_2}{r^{\delta+1}} \tag{3}$$

in the conventions of Ref. 20, which we emply throughout these lectures. When the two masses are separated by a distance $r > R_c$ and the dimensions are assumed to be compactified on a torus of radius R_c the potential becomes:

$$V(r) = -\frac{1}{8\pi M_D^{2+\delta}} \frac{m_1 m_2}{R_c^\delta} \frac{1}{r}, \tag{4}$$

i.e., the usual $1/r$ Newtonian potential is recovered using Gauss' Law. The parameters in the general form of the two-body potential in Eq. (2), (α and λ) depend on the number of extra dimensions and the type of compactification;[21] for the simple case of compactifying on a torus, the range λ of the new interaction is the compactification radius R_c, and $\alpha = 2\delta$. It should be noted that the dependence of Eq. (4) on M_D is related to the compactification scheme through the precise form of the volume factor.

The most recent Cavendish-type experiment[22] excludes scenarios with $\alpha \geq 3$ for $\lambda \geq 140$ μm at 95% confidence level. These results are displayed as the curve labeled Eöt-Wash in Figure 1. Interpreted within the framework of two large additional spatial dimensions, the results imply that $R_c < 190$ μm. The relation of this bound to the fundamental scale M_D depends on the compactification scheme. For $\delta > 2$, R_c is too small for the effects of extra dimensions to be probed in mechanical experiments. Results from other searches are also shown in the $\alpha - \lambda$ plane in this figure. The predictions and allowed regions in the $\alpha - \lambda$ parameter space from other theoretical considerations are also presented in the figure; they include scenarios with axions, dilatons and scalar moduli fields from string theory, and attempted solutions to the cosmological constant problem.

2.2. *Astrophysical and Cosmological Constraints*

Astrophysical and cosmological considerations impose strict constraints on theories of extra dimensions; in particular, early universe cosmology can be drastically altered from the standard picture. The typical energy scale associated with such considerations is of order 100 MeV, and models with KK states that can be produced in this energy regime are highly restricted.

For the case of large extra dimensions of flat and toroidal form, the astrophysical bounds far surpass those from collider or short range gravity

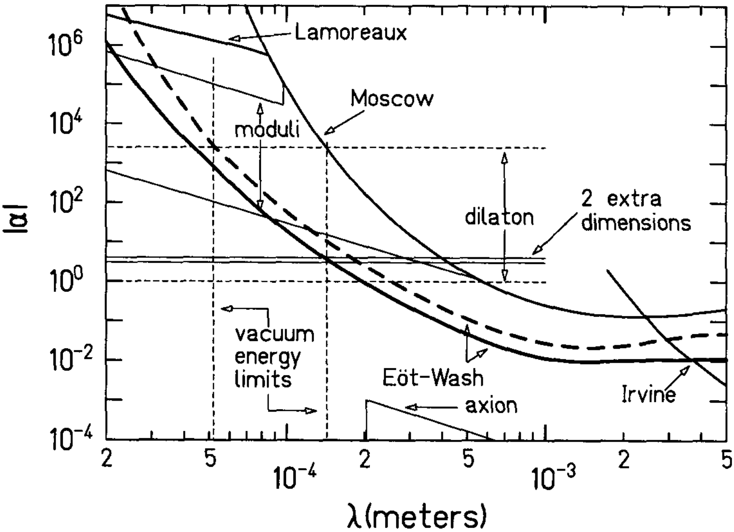


Figure 1. 95% confidence level upper limits on the strength α (relative to Newtonian gravity) as a function of the range λ of additional Yukawa interactions. The region excluded by previous experiments[16,17,23] lies above the curves labeled Irvine, Moscow and Lamoreaux, respectively. The most recent results[22] correspond to the curves labeled Eöt-Wash. The heavy dashed line is the result from Ref. 22 the heavy solid line shows the most recent analysis that uses the total data sample.

experiments for $\delta = 2$. If these large additional dimensions are compactified on a hyperbolic manifold instead, then the astrophysical constraints are avoided[24] as the modified spectrum of KK graviton states admits for a first excitation mass of order several GeV. Alternatively, these bounds are also weakened if a Ricci term is present on the brane since that serves to suppress graviton emission rates.[25]

We now describe the various astrophysical and cosmological considerations that restrict this scenario. These processes include graviton emission during the core collapse of supernovae, the heating of neutron stars from graviton decays, considerations of the cosmic diffuse γ-ray background, overclosure of the universe, matter dominated cooling of the universe, and reheating of the universe. The restrictions obtained from processes that include effects from the decays of KK states rely on the assumption that the KK modes can only decay into Standard Model particles on one brane, *i.e.*, there are no additional branes in the theory, and that decays into other

KK modes with smaller bulk momenta do not occur.

During the core collapse of type II supernova (SN), most of the gravitational binding energy is radiated by neutrinos. This hypothesis has been confirmed by measurements of neutrino fluxes from SN1987A by the Kamiokande and IMB collaborations.[26] Any light, neutral, weakly interacting particle which couples to nucleons, such as bulk gravitons, will compete with neutrinos in carrying energy away from the stellar interior. The rate at which the supernova core can lose energy through emission of KK states can then be used to constrain the fundamental scale M_D.[27,28] The graviton emission process is nucleon 'gravisstrahlung', $N + N \to N + N + X$, where N can be a proton or neutron, and X represents the contributions from massive KK graviton states, ordinary gravitons, and the KK dilaton (scalar) modes which are a remnant of the bulk graviton decomposition. If present, this gravisstrahlung process would provide an additional heat sink and accelerate the supernova cooling in violation with the observations of SN1987A. This process is highly dependent on the temperature of the core at collapse, which is estimated to be $T \approx 30 - 70$ MeV, and on the core density, $\rho \approx (3 - 10) \times 10^{14}\mathrm{gcm}^{-3}$. The most conservative constraint on KK emission[28] yields $R_c \leq 7.1 \times 10^{-4}$ mm for $\delta = 2$ and $R_c \leq 8.5 \times 10^{-7}$ mm for $\delta = 3$, taking $T_{SN1987A} = 30$ MeV.

A complementary bound arises from the radiative decay of the Kaluza Klein gravitons produced by the core collapse of all supernovae that have exploded during the history of the universe (SNe). The two photon decay mode is kinematically favored for the lower mass KK modes[29] with this lifetime being $\tau_{\gamma\gamma} \approx 3 \times 10^9\mathrm{yr}\left(\frac{100\ \mathrm{MeV}}{m_{KK}}\right)^3$. Over the age of the universe, a significant fraction of the KK states emitted from supernovae cores will have decayed into photons, contributing to the cosmic diffuse γ-ray background. This is estimated using the present day supernova rate and the gravisstrahlung rate discussed above. A bound on the size of the additional dimensions is then imposed from the measured cosmic γ-ray background. For a choice of cosmological parameters the predicted γ-flux exceeds the observations by EGRET or COMPTEL[30] unless the fraction of the SN energy released via gravisstrahlung is less than about 0.5-1% of the total. For two extra dimensions, the limit on the compactification radius is $R_c \leq 0.9\ 10^{-4}$ mm and for three extra dimensions the bound is $R_c \leq 0.19\ 10^{-6}$ mm .[31] Additional contributions to the cosmic diffuse γ-ray background arise when the KK gravitons are produced from other sources such as neutrino annihilation, $\nu\bar{\nu} \to G_n \to \gamma\gamma$. These were considered in Ref. 32, and by placing a bound on the normalcy temperature required by Big Bang Nucleosynthesis the limit $R_c \leq 5.1 \times 10^{-5}$ mm for $\delta = 2$ is obtained. In Ref. 32, it is assumed that the universe enters the radiation dominated epoch instantaneously at the reheating temperature. However, it is plausible that the universe enters the radiation epoch after being reheated by the decay of a massive scalar field or by some other means of entropy production. If a large number of KK states are produced during reheating, they are non-relativistic and hence are not diluted by entropy production. Their subsequent decays contribute to the diffuse γ-ray background. Using data from COMPTEL and EGRET,[30] the constraints on M_D are tightened and are 167, 21.7, 4.75, and 1.55 TeV for $\delta = 2, 3, 4$, and 5, respectively, assuming that a 1 GeV maximum temperature is reached during reheating.[33]

The escape velocity of a neutron star is similar to the average speed of thermally produced KK states in a SN core collapse, and hence a large fraction of the KK states can become trapped within the core halo. The decays of these states will continue to be a source of γ-rays long after the SN explosion. Comparisons of the expected contributions to the γ-ray flux rate from this source with EGRET data[30] from nearby neutron stars and pulsars constrains[34] the fraction of the SN energy released via gravisstrahlung to be less than about 10^{-5} of the total. For two extra dimensions this yields the bound $M_D \gtrsim 450$ TeV and for $\delta = 3$ the constraint is $M_D \gtrsim 30$ TeV. The expected sensitivity from GLAST[35] will increase these limits by a factor of 2 to 3.

The Hubble space telescope has observed that the surface temperature of several older neutron stars is higher than that expected in standard cooling models. A possible source for this excess heat is the decays of the KK graviton states trapped in the halo surrounding the star. The γ's, electrons, and neutrinos from the KK decays then hit the star and heat it. For the estimated heating rate from this mechanism not to exceed the observed luminosity, the fraction of the SN energy released via gravisstrahlung must be $\lesssim 5 \times 10^{-8}$ of the total,[34] with the exact number being uncertain by a factor of a few due to theoretical and experimental uncertainties. This is by far the most stringent constraint yielding $M_D \gtrsim 1700, 60$ TeV for $\delta = 2, 3$, respectively. Although the calculations for SN emissions have not been performed for $\delta > 4$, simple scaling suggests that this mechanism results in $M_D \gtrsim 4, 0.8$ TeV for $\delta = 4, 5$, respectively.

Once produced, the massive KK gravitons are sufficiently long-lived as to potentially overclose the universe. Comparisons of KK graviton production rates from photon, as well as neutrino, annihilation to the critical density of the universe results[32] in $R_c < 1.5h \times 10^{-5}$ m for 2 additional dimensions, where h is the current Hubble parameter in units of 100 km/sMpc. While this constraint is milder than those obtained above, it is less dependent on assumptions regarding the existence of additional branes.

Overproduction of Kaluza Klein modes in the early universe could initiate an early epoch of matter radiation equality which would lead to a too low value for the age of the universe. For temperatures below $\sim$ 100 MeV, the cooling of the universe can be accelerated by KK mode production and evaporation into the bulk, as opposed to the normal cosmological expansion. Using the present temperature of the cosmic microwave background of 2.73 K ($= 2.35 \times 10^{-10}$ MeV) and taking the minimum age of the universe to be 12.8 Gyrs ($= 6.2 \times 10^{39}$ MeV^{-1}), as determined by the mean observed age of globular clusters, a maximum temperature can be imposed at radiation-matter equality which cannot be exceeded by the overproduction of KK modes at early times. The resulting lower bounds are M_D are 86, 7.4, and 1.5 TeV for $\delta = 2, 3$, and 4 respectively.[36] Further considerations of the effects from overproduction of KK states on the characteristic scale of the turn-over of the matter power spectra at the epoch of matter radiation equality show that the period of inflation must be extended down to very low temperatures in order to be consistent with the latest data from galaxy surveys.[37]

We collect the constraints from these considerations in Table 1, where we state the restrictions in terms of bounds on the fundamental scale M_D. We note that the relation of the above constraints to M_D is tricky as numerical conventions, as well as assumptions regarding the compactification scheme, explicitly enter some of the computations; in particular, that of gravisstrahlung production during supernova collapse. In addition, all of these bounds assume that all of the additional dimensions are of the same size. The constraints in the table are thus merely indicative and should not be taken as exact.

To conclude this section, we discuss the possible contribution of graviton KK states to the production of high-energy cosmic rays beyond the GZK cut-off of 10^{20} eV. About 20 super-GZK events have been observed and their origin is presently unknown. KK graviton exchange can contribute to high-energy ν-nucleon scattering and produce hadronic sized cross sections above the GZK cut-off for M_D in the range of 1 to 10 TeV.[38]

Table 1. Summary of constraints on the fundamental scale M_D in TeV from astrophysical and cosmological considerations as discussed in the text.

	δ			
	2	3	4	5
Supernova Cooling[28]	30	2.5		
Cosmic Diffuse γ-Rays:				
Cosmic SNe[31]	80	7		
$\nu\bar{\nu}$ Annihilation[32]	110	5		
Re-heating[33]	170	20	5	1.5
Neutron Star Halo[34]	450	30		
Overclosure of Universe[32]	$6.5/\sqrt{h}$			
Matter Dominated Early Universe[36]	85	7	1.5	
Neutron Star Heat Excess[34]	1700	60	4	1

2.3. *Collider Probes*

If such dimensions are present and quantum gravity becomes strong at the TeV scale, then observable signatures at colliders operating at TeV energies must be induced.

We begin by discussing the derivation of the 4-dimensional effective theory, which is computed within linearized quantum gravity. The flat metric is expanded via

$$G_{AB} = \eta_{AB} + \frac{h_{AB}}{M_D^{\delta/2+1}} \tag{5}$$

where the upper case indices extend over the full D-dimensional spacetime and h_{AB} represents the bulk graviton fluctuation. The interactions of the graviton are then described by the action

$$S_{int} = -\frac{1}{M_D^{\delta/2+1}} \int d^4x\, d^\delta y_i\, h_{AB}(x_\mu, y_i) T_{AB}(x_\mu, y_i)\,, \tag{6}$$

with T_{AB} being the symmetric conserved stress-energy tensor. Upon compactification, the bulk graviton decomposes into the various spin states as described in the introduction and Fourier expands into Kaluza-Klein towers of spin-0, 1, and 2 states which have equally spaced masses of $m_{\vec{n}} = \sqrt{\vec{n}^2/R_c^2}$, where $\vec{n} = (n_1, n_2, ...n_\delta)$ labels the KK excitation level. The spin-1 states do not interact with fields on the 3-brane, and the spin-0 states couple to the trace of the stress-energy tensor and will not be considered here. Their phenomenology is described in Ref. 39. Performing the KK expansion for the spin-2 tower, setting $T_{AB} = \eta_A^\mu \eta_B^\nu T_{\mu\nu}\delta(y_i)$ for the Standard Model fields confined to the brane, and integrating the action over the extra dimensional coordinates y_i gives the interactions of the graviton KK states with the Standard Model fields. All the states in the KK tower, including the $\vec{n} = 0$ massless state, couple in an identical manner with universal strength of $M_{\rm Pl}^{-1}$. The corresponding Feynman rules are catalogued in Refs. 20, 29.

One may wonder how interactions of this type can be observable at colliders since the coupling strength is so weak. In the ADD scenario, there are $(ER_c)^\delta$ massive Kaluza-Klein modes that are kinematically accessible in a collider process with energy E. For $\delta = 2$ and $E = 1$ TeV, that totals 10^{30} graviton KK states which may individually contribute to a process! It is the sum over the contribution from each KK state which removes the Planck scale suppression in a process and replaces it by powers of the fundamental scale $M_D \sim$ TeV. The interactions of the massive Kaluza-Klein graviton modes can then be observed in collider experiments either through missing energy signatures or through their virtual exchange in Standard Model processes. At future colliders with very high energies, it is possible that quantum gravity phenomena are accessible resulting in explicit signals for string or brane effects; these will be discussed briefly at the end of this section. We now discuss in detail the two classes of collider signatures for large extra dimensions.

The first set of collider reactions involves the real emission of Kaluza-Klein graviton states in the scattering processes $e^+e^- \to \gamma(Z) + G_n$, and $\overset{(-)}{p}p \to g + G_n$, or in $Z \to f\bar{f} + G_n$. The produced graviton behaves as if it were a massive, non-interacting, stable particle and thus appears as missing energy in the detector. The cross section is computed for the production of a single massive KK excitation and then summed over the full tower of KK states. Since the mass splittings between the KK states is so small, the sum over the states may be replaced by an integral weighted by the density of KK states. The specific process kinematics cut off this integral, rendering a finite and model independent result. The expected suppression from the $M_{\rm Pl}^{-1}$ strength of the graviton KK couplings is exactly compensated by a $M_{\rm Pl}^2$ enhancement in the phase space integration. The cross section for on-shell production of massive Kaluza Klein graviton modes then scales as simple powers of $\sqrt{s}/M_D$,

$$\sigma_{KK} \cdot \sim \frac{1}{M_{\rm Pl}^2}(\sqrt{s}R_c)^\delta \sim \frac{1}{M_D^2}\left(\frac{\sqrt{s}}{M_D}\right)^\delta . \tag{7}$$

The exact expression may be found in Refs. 20, 40. It is important to note that due to integrating over the effective density of states, the radiated graviton appears to have a continuous mass distribution; this corresponds to the probability of emitting gravitons with different extra dimensional momenta. The observables for graviton production, such as the γ/Z angular and energy distributions in e^+e^- collisions, are then distinct from those of other physics processes involving fixed masses for the undetectable particles. In particular, the SM background is given by the 3-body production $e^+e^- \rightarrow \nu\bar{\nu}\gamma$.

The cross section for $e^+e^- \rightarrow \gamma G_n$ as a function of the fundamental Planck scale is presented in Fig. 2 for $\sqrt{s} = 1$ TeV. The level of SM background is also shown, with and without electron beam polarization set at 90%. We note that the signal (background) increases (decreases) with increasing $\sqrt{s}$. Details of the various distributions associated with this process can be found in Ref. 40.

Searches for direct KK graviton production in the reaction $e^+e^- \rightarrow G_n + \gamma(Z)$ at LEP II, using the characteristic final states of missing energy plus a single photon or Z boson, have excluded[41,42] fundamental scales up to $M_D \sim 1.45$ TeV for two extra compactified dimensions and $M_D \sim 0.6$ TeV for six extra dimensions. These analyses use both total cross section measurements and fits to angular distributions to set a limit on the graviton production rates as a function of the number of extra dimensions. The expected discovery reach from this process has been computed in Ref. 43 at a high energy linear e^+e^- collider with $\sqrt{s} = 800$ GeV, 1000 fb^{-1} of integrated luminosity, and various configurations for the beam polarization. These results are displayed in Table 2 and include kinematic acceptance cuts, initial state radiation, and beamsstrahlung.

The emission process at hadron colliders, for example, $q\bar{q} \rightarrow g + G_n$, results in a monojet plus missing transverse energy signature. For larger numbers of extra dimenisons the density of the KK states increases rapidly and the KK mass distribution is shifted to higher values. This is not reflected in the missing energy distribution: although the heavier KK gravitons are more likely to carry larger energy, they are also more likely to be produced at threshold due to the rapidly decreasing parton distribution functions. These two effects compensate each other, leaving nearly identical missing energy distributions. In addition, the effective low-energy theory breaks down for some regions of the parameter space as the parton-level center of mass energy can exceed the value of M_D. Experiments are then sensitive to the new physics appearing above M_D that is associated with the low-scale quantum gravity.

Searches from the Tevatron Run I yield similar results as those from LEP II,[44] however it is anticipated that Run II at the Tevatron will have a higher

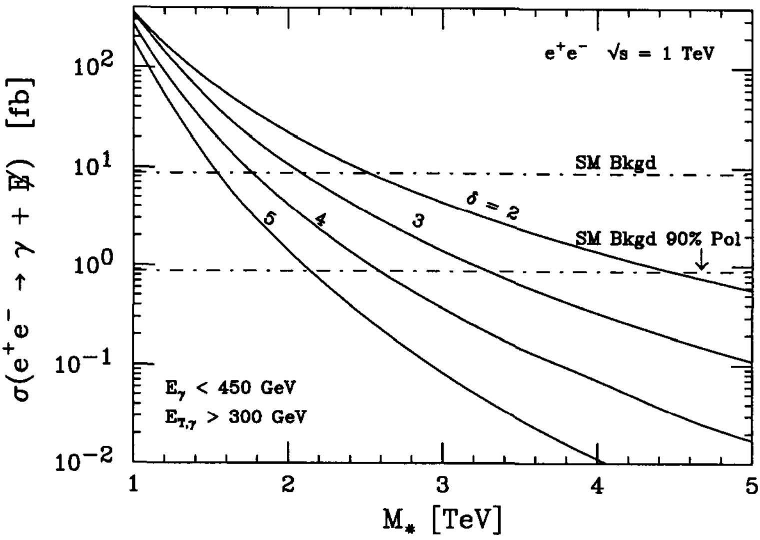


Figure 2. The cross section for $e^+e^- \to \gamma G_n$ for $\sqrt{s} = 1$ TeV as a function of the fundamental Planck scale, labled here by M_*, for various values of δ as indicated. The cross sections for the SN background, with and without 90% beam polarization, correspond to the horizontal lines as indicated. The signal and background are computed with the requirement $E_\gamma < 450$ GeV in order to eliminate the $\gamma Z \to \nu\bar{\nu}\gamma$ contribution to the background. From Ref. 20.

sensitivity.[40] An ATLAS simulation[45] of the missing transverse energy in signal and background events at the LHC with 100 fb^{-1} is presented in Fig. 3 for various values of M_D and δ. This study results in the discovery range displayed in Table 2. The lower end of the range corresponds to where the ultraviolet physics sets in and the effective theory fails, while the upper end represents the boundary where the signal is observable above background.

If an emission signal is observed, one would like to determine the values of the fundamental parameters, M_D and δ. The measurement of the cross section at a linear collider at two different values of $\sqrt{s}$ can be used to determine these parameters[43] and test the consistency of the data with the large extra dimensions hypothesis. This is illustrated in Fig. 4.

Lastly, we note that the cross section for the emission process can be reduced somewhat if the 3-brane is flexible, or soft, instead of being rigid.[46] In this case, the brane is allowed to recoil when the KK graviton is radiated; this can be parameterized as an exponential suppression of the cross section,

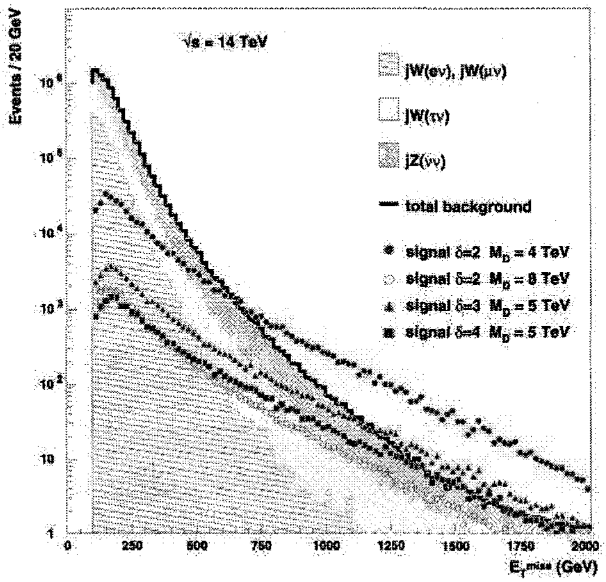


Figure 3. Distribution of the missing transverse energy in background events and signal events for 100 fb^{-1}. The contribution of the three principal Standard Model background processes is shown as well as the distribution of the signal for several values of δ and M_D. From Ref. 45.

with the exponential being a function of the brane tension Δ_τ

$$\frac{d^\sigma}{dx_\gamma d\cos\theta}(\text{soft}) = \frac{d^\sigma}{dx_\gamma d\cos\theta}(\text{stiff})e^{s(1-x_\gamma)/\Delta_\tau^2}\,, \tag{8}$$

where x_γ is the scaled energy of the photon. For reasonable values of the brane tension, this suppression is not numerically large. The brane tension can also be determined by mapping out the cross section as a function of $\sqrt{s}$ as shown in Fig. 4.

The second class of collider signals for large extra dimensions is that of virtual graviton exchange[20,47] in $2 \to 2$ scattering. This leads to deviations in cross sections and asymmetries in Standard Model processes, such as $e^+e^- \to f\bar{f}$. It may also give rise to new production processes which are not present at tree-level in the Standard Model, such as $gg \to \ell^+\ell^-$. The signature is similar to that expected in composite theories and provides a good experimental tool for searching for large extra dimensions for the case $\sqrt{s} < M_D$.

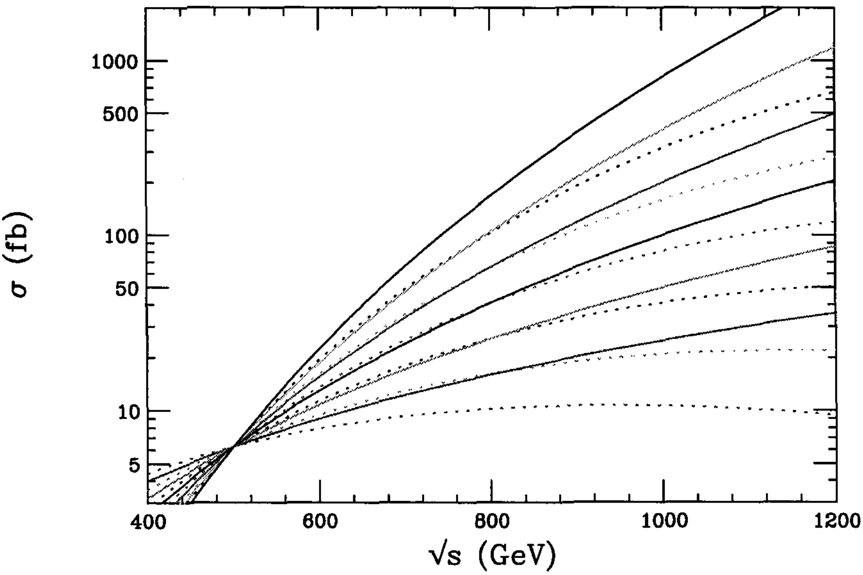


Figure 4. Dependence of the cross section for the process $e^+e^- \to \gamma G_n$ on the center-of-mass energy, normalized to the case where $M_D = 5$, $\delta = 2$ at $\sqrt{s} = 500$ GeV. From top to bottom, the curves correspond to $\delta = 7, 6, 5, 4, 3, 2$. The dashed curves correspond to the case where the brane tension is set to $\Delta_T = 800$ GeV.

Table 2. 95% CL sensitivity to the fundamental scale M_D in TeV for different values of δ, from the emission process for various polarization configurations and different colliders as discussed in the text. $\sqrt{s} = 800$ GeV and 1 ab^{-1} has been assumed for the LC and 100 fb^{-1} for the LHC. Note that the LHC only probes M_D within the stated range.

$e^+e^- \to \gamma + G_n$		2	4	6
LC	$P_{-,+} = 0$	5.9	3.5	2.5
LC	$P_- = 0.8$	8.3	4.4	2.9
LC	$P_- = 0.8$, $P_+ = 0.6$	10.4	5.1	3.3
$pp \to g + G_n$		2	3	4
LHC		$4 - 8.9$	$4.5 - 6.8$	$5.0 - 5.8$

Graviton exchange is governed by the effective Lagrangian

$$\mathcal{L} = i\frac{4\lambda}{M_H^4} T_{\mu\nu}T^{\mu\nu} + h.c. \tag{9}$$

The amplitude is proportional to the sum over the propagators for the graviton KK tower which may be converted to an integral over the density of KK states. However, in this case, there is no specific cut-off associated with the process kinematics and the integral is divergent for $\delta > 1$. This introduces a sensitivity to the unknown ultraviolet physics which appears at the fundamental scale. This integral needs to be regulated and several approaches have been proposed: (i) a naive cut-off scheme[20,47] (ii) brane fluctuations,[48] or (iii) the inclusion of full weakly coupled TeV-scale string theory in the scattering process.[49] The most model independent approach which does not make any assumptions as to the nature of the new physics appearing at the fundamental scale is that of the naive cut-off. Here, the cut-off is set to $M_H \neq M_D$; the exact relationship between M_H and M_D is not calculable without knowledge of the full theory. The parameter $\lambda = \pm 1$ is also usually incorporated in direct analogy with the standard parameterization for contact interactions[50] and accounts for uncertainties associated with the ultraviolet physics. The substitution

$$\mathcal{M} \sim \frac{i^2\pi}{M_{\rm Pl}^2} \sum_{\vec{n}=1}^{\infty} \frac{1}{s - m_{\vec{n}}^2} \to \frac{\lambda}{M_H^4} \tag{10}$$

is then performed in the matrix element for s-channel KK graviton exchange with corresponding replacements for t- and u-channel scattering. As above, the Planck scale suppression is removed and superseded by powers of $M_H \sim$TeV.

The resulting angular distributions for fermion pair production are quartic in $\cos\theta$ and thus provide a unique signal for spin-2 exchange. An illustration of this is given in Fig. 5 which displays the angular dependence of the polarized Left-Right asymmetry in $e^+e^- \to b\bar{b}$.

The experimental analyses also make use of the cut-off approach. Using virtual Kaluza-Klein graviton exchange in reactions with diphoton, diboson and dilepton final states, $e^+e^- \to G_n \to \gamma\gamma, VV, \ell\ell$, LEP experiments[51] exclude $M_H \lesssim 0.5-1.0$ TeV independent of the number of extra dimensions. At the Tevatron,[52] the combined Drell-Yan and diphoton channels exclude exchange scales up to ~ 1.1 TeV. In addition, H1 and ZEUS at HERA[53] have both placed the bound $M_H \gtrsim 800$ GeV.

The potential search reach for virtual KK graviton exchange in processes at future accelerators are listed in Table 3. These sensitivities are estimated for the LHC,[54] a high energy e^+e^- linear collider,[47] as well as for a $\gamma\gamma$ collider,[55] where the initial photon beams originate from Compton laser back-scattering. Note that the $\gamma\gamma \to WW$ process has the highest sensitivity to graviton exchange.

In summary, present facilities have searched for large extra dimensions and excluded their existence for fundamental scales up to $\sim$ TeV. The reach of future facilities will extend this reach to a sensitivity of ~ 10 TeV. If this

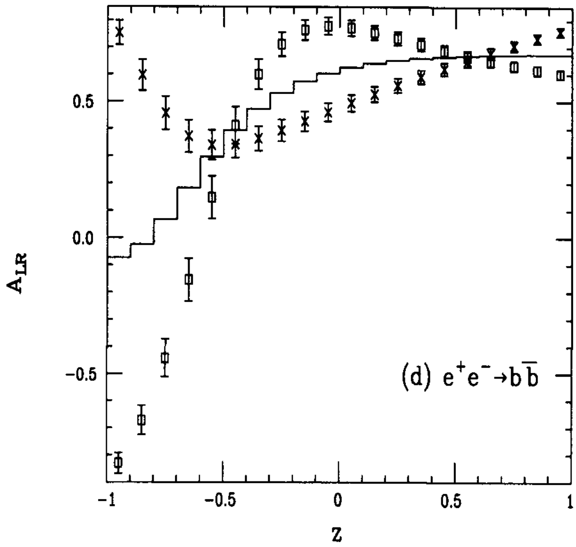


Figure 5. Distribution of the angular dependence ($z = \cos\theta$) of the polarized Left-Right asymmetry in $e^+e^- \to b\bar{b}$ with $\sqrt{s} = 500$ GeV, taking $M_D = 1.5$ TeV and $\lambda = \pm 1$. The solid histogram is the Standard Model expectation. The two sets of data points correspond to the two choices of sign for λ, and the error bars represent the statistics in each bin for an integrated luminosity of 75 fb^{-1}. From Ref. 47.

Table 3. The estimated 95% CL search reach for M_H from various processes at future accelerators.

		$\sqrt{s}$ (TeV)	M_H (TeV)
LC	$e^+e^- \to f\bar{f}$	0.5	4.1
	$e^+e^- \to f\bar{f}$	1.0	7.2
	$\gamma\gamma \to \gamma\gamma$	1.0	3.5
	$\gamma\gamma \to WW$	1.0	13.0
	$e\gamma \to e\gamma$	1.0	8.0
LHC	$pp \to \ell^+\ell^-$	14.0	7.5
	$pp \to \gamma\gamma$	14.0	7.1

scenario is indeed relevant to the hierarchy, then it should be discovered in the next round of experiments. In addition, future experiments will have the capability to determine the geometry of the higher dimensional space, such as the size and number of extra dimensions, as well as the degree of the brane tension.

If the fundamental scale of gravity is at roughly a TeV, then future colliders will directly probe new exotic degrees of freedom in addition to the Kaluza Klein modes of extra dimensions, including the effects of quantum gravity itself. We do not yet have unambiguous predictions for this new and unknown physics, but it could take the form of new strongly interacting gauge sectors or string or brane excitations. For example, the exchange of string Regge excitations of Standard Model particles in $2 \to 2$ scattering[49] would appear as a contact-like interaction, similar to that of graviton KK exchange, but with a large strength.

It is possible that inelastic scattering at energies $\gg$ TeV could be dominated by the production of strongly coupled objects such as microscopic black holes.[56] Assuming that these decay via Hawking radiation, they would then be observable in future very high-energy colliders.[57]

3. TeV^{-1}-Sized Extra Dimensions

The possibility of TeV^{-1}-sized extra dimensions naturally arises in braneworld theories.[9,58] By themselves, they do not allow for a reformulation of the hierarchy problem, but they may be incorporated into a larger structure in which this problem is solved. In these scenarios, the Standard Model fields are phenomenologically allowed to propagate in the bulk. This presents a wide variety of choices for model building: (i) all, or only some, of the Standard Model gauge fields exist in the bulk; (ii) the Higgs field may lie on the brane or in the bulk; (iii) the Standard Model fermions may be confined to the brane or to specific locales in the extra dimension. The phenomenological consequences of this scenario strongly depend on the location of the fermion fields. Unless otherwise noted, our discussion assumes that all of the Standard Model gauge fields propagate in the bulk.

The masses of the excitation states in the gauge boson KK towers depend on where the Higgs boson is located. If the Higgs field propagates in the bulk, the zero-mode state of the Higgs KK tower receives a vacuum expectation value (vev) which is responsible for the spontaneous breaking of the electroweak gauge symmetry. In this case, the resulting mass matrix for the states in the gauge boson KK towers is diagonal and the excitation masses are shifted by the mass of the gauge zero-mode, which corresponds to the Standard Model gauge field, giving

$$m_{\vec{n}} = (m_0^2 + \vec{n} \cdot \vec{n}/R_c^2)^{1/2} \,. \tag{11}$$

However, if the Higgs is confined to the brane, its vev induces mixing, amongst the gauge KK states of order $(m_0 R_c)^2$. The KK mass matrix must then be diagonalized in order to determine the excitation masses. For the case of 1 extra TeV^{-1}-sized dimension, the coupling strength of the gauge KK states to the Standard Model fermions on the brane is $\sqrt{2}g$, where g is the corresponding Standard Model gauge coupling.

We first discuss the case where the Standard Model fermions are rigidly fixed to the brane and do not feel the effects of the additional dimensions. For models in this class, precision electroweak data place strong constraints[59] on the mass of the first gauge KK excitation. Contributions to electroweak observables arise from the virtual exchange of gauge KK states and a summation over the contributions from the entire KK tower must be performed. For $D > 5$, this sum is divergent. In the full higher dimensional theory, some new, as of yet unknown, physics would regularize this sum and render it finite. An example of this is given by the possibility that the brane is flexible or non-rigid,[48] which has the effect of exponentially damping the sum over KK states. Due to our present lack of knowledge of the full underlying theory, the KK sum is usually terminated by an explicit cut-off, which provides a naive estimate of the magnitude of the effects.

Since the $D = 5$ theory is finite, it is the scenario that is most often discussed and is sometimes referred to as the 5-dimensional Standard Model (5DSM). In this case, a global fit to the precision electroweak data including the contributions from KK gauge interactions yields[59] $m_1 \sim R_c^{-1} \gtrsim 4$ TeV. In addition, the KK contributions to the precision observables allow for the mass of the Higgs boson to be somewhat heavier than the value obtained in the Standard Model global fit. Given the constraint on R_c from the precision data set, the gauge KK contributions to the anomalous magnetic moment of the muon are small.[60]

Such a large mass for the first gauge KK state is beyond the direct reach at present accelerators, as well as a future e^+e^- linear collider. However, they can be produced as resonances at the LHC in the Drell-Yan channel provided $m_1 \lesssim 6$ TeV. Lepton colliders can indirectly observe the existence of heavy gauge KK states in the contact interaction limit via their s-channel exchanges. In this case the contribution of the entire KK tower must be summed, and suffers the same problems with divergences discussed above. The resulting sensitivities to the gauge KK tower in the 5DSM from direct and indirect searches at various facilities is displayed in Table 4.

We now discuss the scenario where the Standard Model fermions are localized at specific points in the extra TeV^{-1}-sized dimensions. In this

Table 4. 95% CL search reach for the mass m_1 of the first KK gauge boson excitation.[59]

	m_1 Reach (TeV)
Tevatron Run II 2 fb^{-1}	1.1
LHC 100 fb^{-1}	6.3
LEP II	3.1
LC $\sqrt{s} = 0.5$ TeV 500 fb^{-1}	13.0
LC $\sqrt{s} = 1.0$ TeV 500 fb^{-1}	23.0
LC $\sqrt{s} = 1.5$ TeV 500 fb^{-1}	31.0

case, the fermions have narrow gaussian-like wave functions in the extra dimensions with the width of their wave function being much smaller than R_c^{-1}. The placement of the different fermions at distinct locations in the additional dimensions, along with the narrowness of their wavefunctions, can then naturally suppress[12] operators mediating dangerous processes such as proton decay. The exchange of gauge KK states in $2 \to 2$ scattering processes involving initial and final state fermions is sensitive to the placement of the fermions and can be used to perform a cartography of the localized fermions,[61] *i.e.*, measure the wavefunctions and locations of the fermions. At very large energies, it is possible that the cross section for such scattering will tend rapidly to zero since the fermions' wavefunctions will not overlap and hence they may completely miss each other in the extra dimensions.[62]

Lastly, we discuss the case of universal extra dimensions,[63] where all Standard Model fields propagate in the bulk, and branes need not be present. Translational invariance in the higher dimensional space is thus preserved. This results in the tree-level conservation of the δ-dimensional momentum of the bulk fields, which implies that KK parity, $(-1)^n$, is conserved to all orders. The phenomenology of this scenario is quite different from the cases discussed above. Since KK parity is conserved, KK excitations can no longer be produced as s-channel resonances; they can now only be produced in pairs. This results in a drastic reduction of the collider sensitivity to such states, with searches at the Tevatron yielding the bounds[63,64] $m_1 \gtrsim 400$ GeV for two universal extra dimensions. The constraints from electroweak precision data are also lowered and yield similar bounds. Since the KK states are allowed to be relatively light, they can produce observable effects[64,65] in loop-mediated processes, such as $b \to s\gamma$, $g-2$ of the muon, and rare Higgs decays.

4. Warped Extra Dimensions with Localized Gravity

In this scenario, the hierarchy between the Planck and electroweak scales is generated by a large curvature of the extra dimensions.[4,5] The simplest such framework is comprised of just one additional spatial dimension of finite size, in which gravity propagates. The geometry is that of a 5-dimensional anti-de-Sitter space (AdS_5), which is a space of constant negative curvature. The extent of the 5^{th} dimension is $y = \pi R_c$. Every slice of the 5^{th} dimension corresponds to a 4-d Minkowski metric. Two 3-branes, with equal and opposite tension, sit at the boundaries of this slice of AdS_5 space. The Standard Model fields are constrained to the 3-brane located at the boundary $y = \pi R_c$, known as the TeV-brane, while gravity is localized about the opposite brane at the other boundary $y = 0$. This is referred to as the Planck brane.

The metric for this scenario preserves 4-d Poincare invariance and is given by

$$ds^2 = e^{-2ky}\eta_{\mu\nu}dx^\mu dx^\nu - dy^2\,, \tag{12}$$

where the exponential function of the 5^{th} dimensional coordinate multiplying the usual 4-d Minkowski term indicates a non-factorizable geometry. This exponential is known as a warp factor. Here, the parameter k governs the degree of curvature of the AdS_5 space; it is assumed to be of order the Planck scale. Consistency of the low-energy theory sets $k/\overline{M}_{\rm Pl} \lesssim 0.1$, with $\overline{M}_{\rm Pl} = M_{\rm Pl}/\sqrt{8\pi} = 2.4 \times 10^{18}$ being the reduced 4-d Planck scale. The relation

$$\overline{M}^2_{\rm Pl} = \frac{\overline{M}^3_5}{k} \tag{13}$$

is derived from the 5-dimensional action and indicates that the (reduced) 5-dimensional fundamental scale $\overline{M}_5$ is of order $\overline{M}_{\rm Pl}$. Since $k \sim \overline{M}_5 \sim \overline{M}_{\rm Pl}$, there are no additional hierarchies present in this model.

The scale of physical phenomena as realized by a 4-dimensional flat metric transverse to the 5^{th} dimension is specified by the exponential warp factor. The scale $\Lambda_\pi \equiv \overline{M}_{\rm Pl}e^{-kR_c\pi}$ then describes the scale of all physical processes on the TeV-brane. With the gravitational wavefunction being localized on the Planck brane, Λ_π takes on the value ~ 1 TeV providing $kR_c \simeq 11 - 12$. It has been demonstrated[66] that this value of kR_c can be stabilized within this configuration without the fine tuning of parameters. The hierarchy is thus naturally established by the warp factor. Note that since $kR_c \simeq 10$ and it is assumed that $k \sim 10^{18}$ GeV, this is not a model with a large extra dimension.

Two parameters govern the 4-d effective theory of this scenario:[67] Λ_π and the ratio $k/\overline{M}_{Pl}$. Note that the approximate values of these parameters are known due to the relation of this model to the hierarchy problem. As in the case of large extra dimensions, the Feynman rules are obtained by a linear expansion of the flat metric,

$$G_{\alpha\beta} = e^{-2ky}(\eta_{\alpha\beta} + 2h_{\alpha\beta}/M_5^{3/2}) \,, \tag{14}$$

which for this scenario includes the warp factor multiplying the linear expansion. After compactification, the resulting KK tower states are the coefficients of a Bessel expansion with the Bessel functions replacing the Fourier series of a flat geometry due to the strongly curved space and the presence of the warp factor. Here, the masses of the KK states are $m_n = x_n k e^{-kR_c\pi} = x_n \Lambda_\pi k/\overline{M}_{\rm Pl}$ with the x_n being the roots of the first-order Bessel function, *i.e.*, $J_1(x_n) = 0$. The first excitation is then naturally of order a TeV and the KK states are not evenly spaced. The interactions of the graviton KK tower with the Standard Model fields on the TeV-brane are given by

$$\mathcal{L} = \frac{-1}{\overline{M}_{Pl}} T^{\mu\nu}(x) h^{(0)}_{\mu\nu}(x) - \frac{1}{\Lambda_\pi} T^{\mu\nu}(x) \sum_{n=1}^{\infty} h^{(n)}_{\mu\nu}(x) \,. \tag{15}$$

Note that the zero-mode decouples and that the couplings of the excitation states are inverse TeV strength. This results in a strikingly different phenomenology than in the case of large extra dimensions.

In this scenario, the principal collider signature is the direct resonant production of the spin-2 states in the graviton KK tower. To exhibit how this may appear at a collider, Figure 6 displays the cross section for $e^+e^- \to \mu^+\mu^-$ as a function of $\sqrt{s}$, assuming $m_1 = 500$ GeV and varying $k/\overline{M}_{\rm Pl}$ in the range $0.01 - 0.05$. The height of the third resonance is greatly reduced as the higher KK excitations prefer to decay to the lighter graviton states, once it is kinematically allowed.[68] In this case, high energy colliders may become graviton factories! If the first graviton KK state is observed, then the parameters of this model can be uniquely determined by measurement of the location and width of the resonance. In addition, the spin-2 nature of the graviton resonance can be determined from the shape of the angular distribution of the decay products. This is demonstrated in Figure 7, which displays the angular distribution of the final state leptons in Drell-Yan production, $pp \to \ell^+\ell^-$, at the LHC.[69]

Searches for the first graviton KK resonance in Drell-Yan and dijet data from Run I at the Tevatron restrict[67] the parameter space of this model,

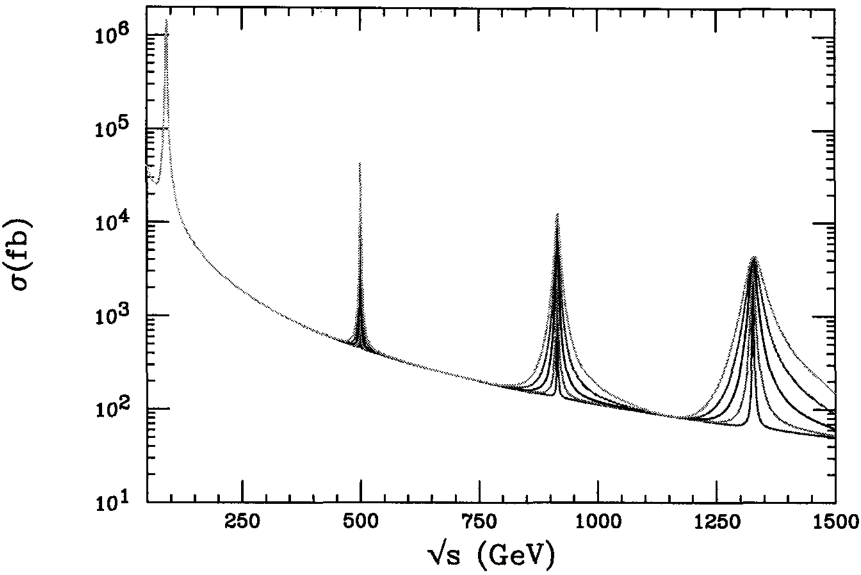


Figure 6. The cross section for $e^+e^- \to \mu^+\mu^-$ including the exchange of a KK tower of gravitons in the Randall-Sundrum model with $m_1 = 500$ GeV. The curves correspond to $k/\overline{M}_{\rm Pl}$ = in the range 0.01 – 0.05.

as shown in Figure 8. These data exclude larger values of $k/\overline{M}_{\rm Pl}$ for values of m_1 which are in kinematic reach of the accelerator.

Gravitons may also contribute to precision electroweak observables. A precise description of such contributions requires a complete understanding of the full underlying theory due to the non-renormalizability of gravity. However, naive estimates of the size of such effects can be obtained in an effective field theory by employing a cut-off to regulate the theory.[70] The resulting cut-off dependent constraints indicate[71] that smaller values of $k/\overline{M}_{\rm Pl}$ are inconsistent with precision electroweak data, as shown in Figure 9.

These two constraints from present data, taken together with the theoretical assumptions that (i) $\Lambda_\pi \lesssim 10$ TeV, *i.e.*, the scale of physics on the TeV-brane is not far above the electroweak scale so that an additional hierarchy is not generated, and (ii) $k/\overline{M}_{\rm Pl} \lesssim 0.1$ from bounds on the curvature of the AdS_5, result in a closed allowed region in the two parameter space. This is displayed in Figure 9, which also shows the expected search reach for resonant graviton KK production in the Drell-Yan channel at the LHC. We see that the full allowed parameter space can be completely explored at

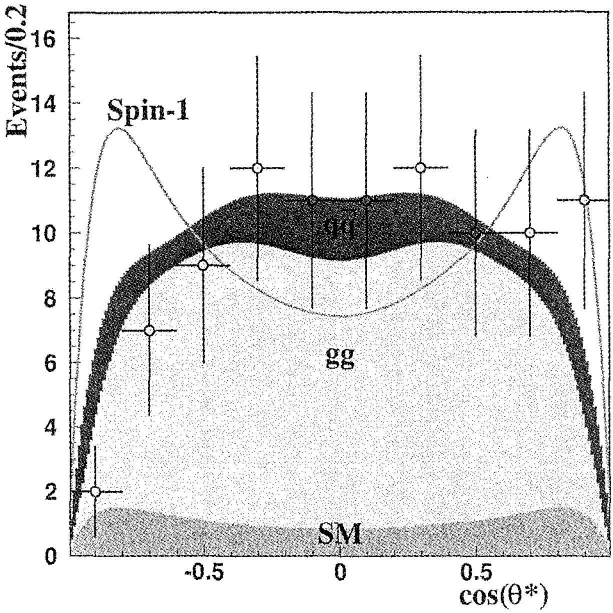


Figure 7. The angular distribution of "data" at the LHC from Drell-Yan production of the first graviton KK excitation with $m_1 = 1.5$ TeV and 100 fb^{-1} of integrated luminosity. The stacked histograms represent the Standard Model contributions, and gg and $q\bar{q}$ initiated graviton production as labeled. The curve shows the expected distribution from a spin-1 resonance. From Ref. 69.

the LHC, given our theoretical prejudices, and hence the LHC will either discover or exclude this model.

If the above theoretical assumptions are evaded, the KK gravitons may be too massive to be produced directly. However, their contributions to fermion pair production may still be felt via virtual exchange. In this case, the uncertainties associated with the introduction of a cut-off are avoided, since there is only one additional dimension and the KK states may be neatly summed. The resulting sensitivities[67] to Λ_π at current and future colliders are listed in Table 5.

The Goldberger-Wise mechanism[66] for stabilizing the separation of the two 3-branes in this configuration with $kR_c \sim 10$ leads to the existence of a new, relatively light scalar field. This field is the radion and it is related to the radial fluctuations of the extra dimension, and to the scalar remnant of the bulk graviton KK decomposition. The radion couples to the Standard Model fields via the trace of the stress energy tensor with strength $\sim T^\mu_\mu/\Lambda_\pi$. These interactions are similar to those of the Higgs

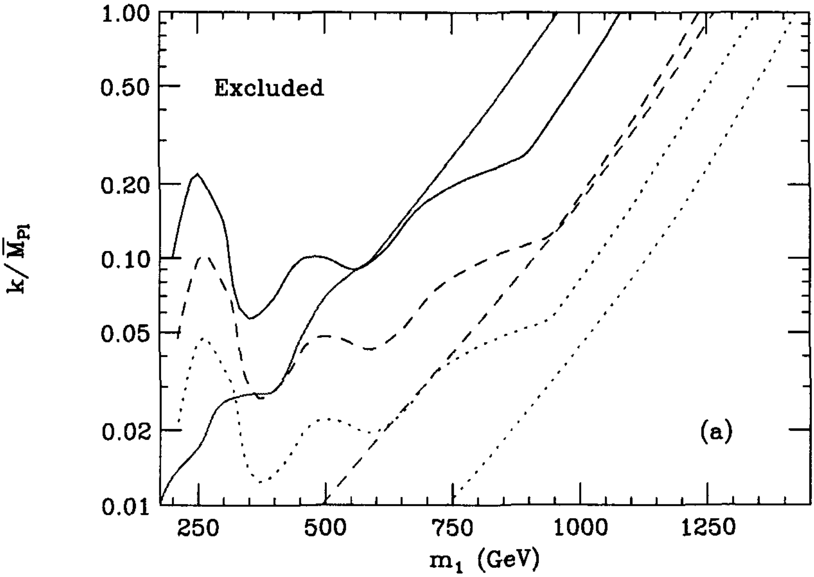


Figure 8. Exclusion regions for resonance production for the first KK graviton excitation in the Drell-Yan and dijet channels at the Tevatron. The solid curves represent the results from Run I, while the dashed, dotted curves correspond to Run II with 2, 300 fb^{-1} of integrated luminosity respectively. The excluded region lies to the left of the curves. From Ref. 67.

Table 5. 95% CL search reach for Λ_π in TeV in the contact interaction regime taking 500, 2.5, 2, and 100 fb^{-1} of integrated luminosity at the LC, LEP II, Tevatron, and LHC, respectively. From Ref. 67.

	$k/\overline{M}_{Pl}$		
	0.01	0.1	1.0
LEP II	4.0	1.5	0.4
LC $\sqrt{s} = 0.5$ TeV	20.0	5.0	1.5
LC $\sqrt{s} = 1.0$ TeV	40.0	10.0	3.0
Tevatron Run II	5.0	1.5	0.5
LHC	20.0	7.0	3.0

boson, and it is allowed to mix with the Higgs, which alters the couplings of both fields. The phenomenology of this field is detailed in Refs. 39, 72.

Astrophysical bounds are not present in this scenario since the first graviton KK state occurs at a ~TeV. However, the TeV scale graviton KK states can induce high energy cosmic rays. In this case, neutrino annihi-

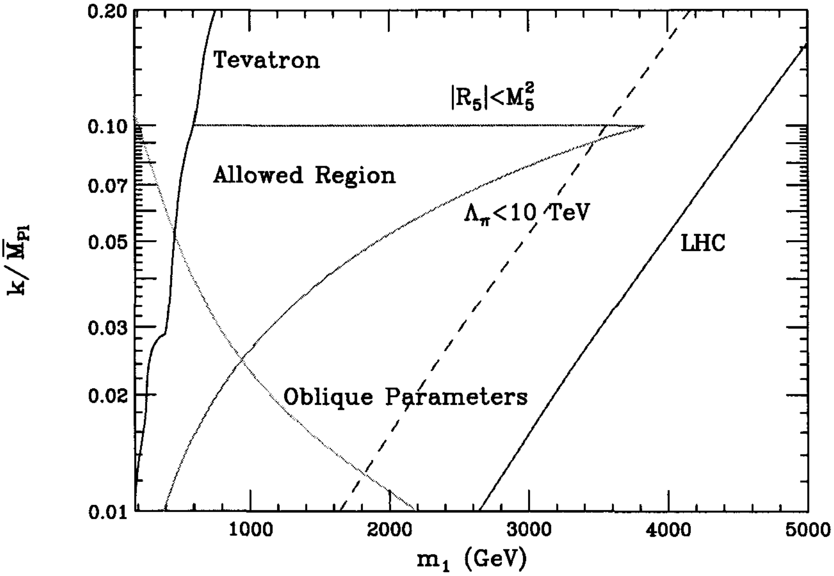


Figure 9. Summary of experimental and theoretical constraints on the Randall-Sundrum model in the two-parameter plane $k/\overline{M}_{\rm Pl}-m_1$, for the case where the Standard Model fields are constrained to the TeV-brane. The allowed region lies in the center as indicated. The LHC sensitivity to graviton resonances in the Drell-Yan channel is represented by the diagonal dashed and solid curves, corresponding to 10 and 100 fb^{-1} of integrated luminosity, respectively. From Ref. 71.

lation within a GZK distance of the earth can produce a single graviton KK state on resonance which subsequently decays hadronically.[73] For neutrinos of mass $m_\nu \sim 10^{-2}$ to 10^{-1} eV, and graviton resonances of order a TeV, super-GZK events can be produced. Under the assumption that the incident neutrino spectrum extends in neutrino energy with a reasonably slow fall-off, the existence of a series of s-channel KK graviton resonances will lead to a series of ultra-GZK events. The rates for these bursts are generally at or near the present level of observability for a wide range of model parameters. The fact that such events are not as yet observed can be used to constrain the parameter space of this model once a specific form of the neutrino energy spectrum is assumed. This is demonstrated in Fig. 10.

In a variant of this model, the Standard Model fields may propagate in the bulk. This is desirable for numerous model building reasons as mentioned in the introduction. As a first step, one can study the effect of placing the Standard Model gauge fields in the bulk and keeping the fermions on

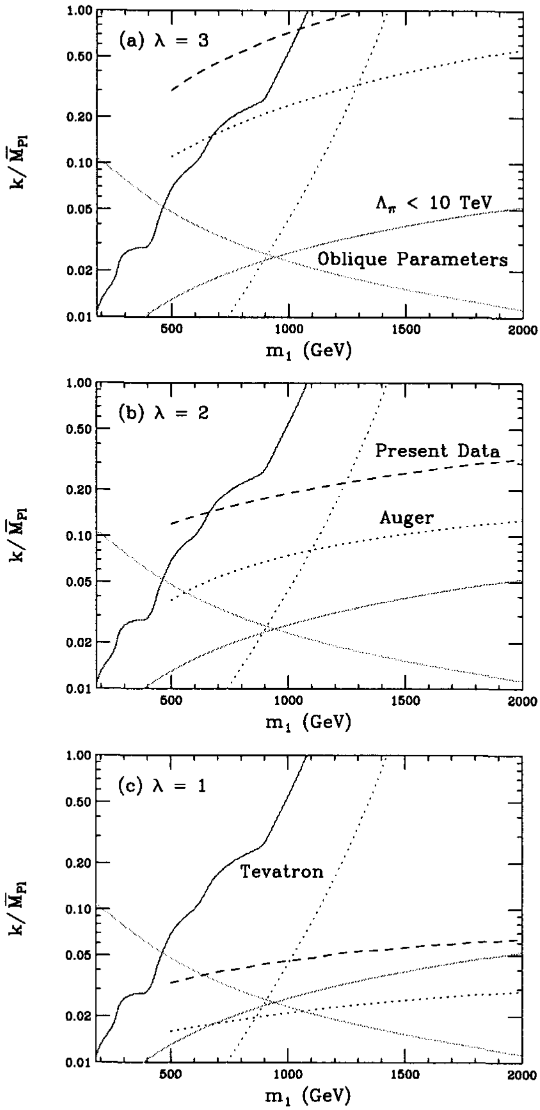


Figure 10. Allowed region in the $(k/\overline{M}_{Pl})$-m_1 plane for various values of λ which describes the fall with energy of the neutrino flux as $E^{-\lambda/2}$. The solid (dotted) diagonal red curve excludes the region above and to the left from direct searches for graviton resonances at the Run I (II, 30 fb^{-1}) Tevatron. The light blue (green) curve is an indirect bound from the oblique parameter analysis (based on the hierarchy requirement that $\Lambda_\pi < 10$ TeV) and excludes the region below it. The black dashed (dotted) curves excluding the regions above them at 95% CL based on present (anticipated future Auger) cosmic ray data.

the TeV-brane. In this case, one finds[74] that the fermions on the brane couple to the KK gauge fields ~ 9 times more strongly than they couple to the Standard Model gauge fields. This results in strong bounds on gauge KK states from their contributions to electroweak precision data. A global fit to the electroweak data set yields the constraint $m_1 \gtrsim 25$ TeV on the first gauge KK mass, implying $\Lambda_\pi \gtrsim 100$ TeV.

This bound can be relaxed if the fermions also reside in the bulk.[71,75] In this case, a third parameter is introduced, corresponding to the bulk fermion mass which is given by $m_5 = \nu k$ with ν being of order one. The parameter ν controls the shape of the fermion zero mode wavefunction. The resulting phenomenology is markedly different, and is highly dependent on the parameter ν. In particular, large mixing is induced between the zero-mode top-quark and the states in its KK tower. This results[76] in substantial shifts to the ρ-parameter and forces the third generation of fermions to be confined to the TeV-brane with only the first two generations of fermions being allowed to reside in the bulk.

An alternate scenario is possible[5] when the second brane is taken off to infinity, *i.e.*, $R_c \to \infty$, and the Standard Model fields are confined to the brane at $y = 0$ where gravity is localized. In this case, the graviton KK modes become continuous, *i.e.*, the gap between KK states disappears, and their couplings to the Standard Model fields are much weaker than $M_{\rm Pl}$. This configuration no longer allows for a reformulation of the hierarchy problem, but can potentially be observable[77] in sub-mm gravitational force experiments.

Another consistent scenario of this type involves two branes, both with positive tension, separated in a five-dimensional Anti-de Sitter geometry of infinite extent.[78] The graviton is localized on one of the branes, while a gapless continuum of additional gravity modes probe the infinite fifth dimension. The phenomenological effects of this framework are similar to the process of real graviton emission in the ADD scheme with six large toroidal dimensions. In this scenario, the resulting cosmological constraints are found to be very mild.[79]

5. Summary

If the structure of spacetime is different than that readily observed, gravitational physics, particle physics and cosmology are all immediately affected. The physics of extra dimensions offers new insights and solutions to fundamental questions arising in these fields. Here, we have summarized the pioneering frameworks with extra spatial dimensions which have observable consequences at the TeV scale. We have outlined some of the experimental observations in particle and gravitational physics as well as astrophysical and cosmological considerations that can constrain or confirm these scenarios. These developing ideas and the wide interdisciplinary experimental program that is charted out to investigate them mark a renewed effort to describe the dynamics behind spacetime. We look forward to the discovery of a higher dimensional spacetime.

Acknowledgements

JH is supported by the Department of Energy, Contract DE-AC03-76SF00515.

References

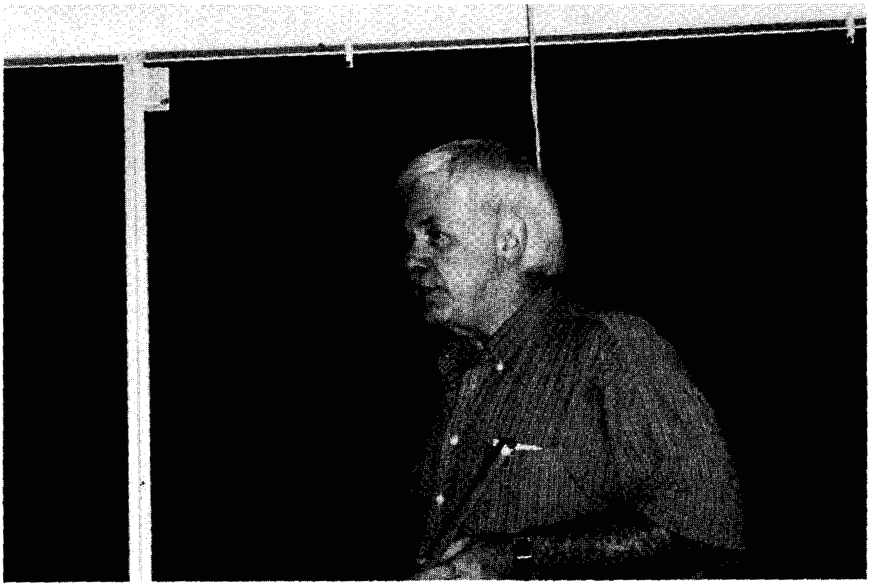
1. G. Nordstrom, Z. Phys. **15**, 504 (1914).
2. T. Kaluza, Preuss.Akad.Wiss, Berlin, Math. Phys. K **1**, 966 (1921); O. Klein Z. Phys. **37** 895; Nature **118**, 516 (1926).
3. N. Arkani-Hamed, S. Dimopoulos and G. R. Dvali, Phys. Lett. B **429**, 263 (1998) [arXiv:hep-ph/9803315]; I. Antoniadis, N. Arkani-Hamed, S. Dimopoulos and G. R. Dvali, Phys. Lett. B **436**, 257 (1998) [arXiv:hep-ph/9804398]; N. Arkani-Hamed, S. Dimopoulos and G. R. Dvali, Phys. Rev. D **59**, 086004 (1999) [arXiv:hep-ph/9807344].
4. L. Randall and R. Sundrum, Phys. Rev. Lett. **83**, 3370 (1999) [arXiv:hep-ph/9905221].
5. L. Randall and R. Sundrum, Phys. Rev. Lett. **83**, 4690 (1999) [arXiv:hep-th/9906064].
6. J. Polchinski, String Theory, Cambridge University Press (1998).
7. N. Arkani-Hamed, S. Dimopoulos, G. R. Dvali and N. Kaloper, Phys. Rev. Lett. **84**, 586 (2000) [arXiv:hep-th/9907209].
8. K. R. Dienes, E. Dudas and T. Gherghetta, Phys. Lett. B **436**, 55 (1998) [arXiv:hep-ph/9803466]; Nucl. Phys. B **537**, 47 (1999) [arXiv:hep-ph/9806292]; Nucl. Phys. B **567**, 111 (2000) [arXiv:hep-ph/9908530].
9. I. Antoniadis, Phys. Lett. B **246**, 377 (1990).
10. D. E. Kaplan, G. D. Kribs and M. Schmaltz, Phys. Rev. D **62**, 035010 (2000) [arXiv:hep-ph/9911293], Z. Chacko, M. A. Luty, A. E. Nelson and E. Ponton, JHEP **0001**, 003 (2000) [arXiv:hep-ph/9911323]; N. Arkani-Hamed, L. J. Hall, D. R. Smith and N. Weiner, Phys. Rev. D **63**, 056003 (2001) [arXiv:hep-ph/9911421].
11. K. R. Dienes, E. Dudas and T. Gherghetta, Nucl. Phys. B **557**, 25 (1999) [arXiv:hep-ph/9811428]; N. Arkani-Hamed, S. Dimopoulos, G. R. Dvali and J. March-Russell, Phys. Rev. D **65**, 024032 (2002) [arXiv:hep-ph/9811448]; Y. Grossman and M. Neubert, Phys. Lett. B **474**, 361 (2000) [arXiv:hep-ph/9912408].

12. N. Arkani-Hamed and M. Schmaltz, Phys. Rev. D **61**, 033005 (2000) [arXiv:hep-ph/9903417].
13. "Flavor Physics For The Millenium", Proceedings of the Theoretical Advanced Study Institute in Elementary Particle Physics edited by J.L. Rosner (University of Chicago) 2000.
14. Particle Data Group (C. Caso *et al.*), Eur. Phys. J. C **3**, 781 (1998).
15. D.E. Krause and E. Fischbach, arXiv:hep-ph 9912276; G.L. Smith, C.D. Hoyle, J.H. Gundlach, E.G. Adelberger, B.R. Heckel and H.E. Swanson, Phys. Rev. D **61**, 022001 (2000).
16. J. K. Hoskins, R. D. Newman, R. Spero, and J. Schultz, Phys. Rev D **32**, 3084 (1985).
17. V. P. Mitrofanov and O. I. Ponomareva, Sov. Phys. JETP **67**, 1963 (1988).
18. J.C. Long, H.W. Chan, J.C. Price, Nucl. Phys. B **539**, 23 (1999).
19. A. H. Cook, Contemp. Phys. **28**, 159 (1987).
20. G. F. Giudice, R. Rattazzi and J. D. Wells, Nucl. Phys. B **544**, 3 (1999) [arXiv:hep-ph/9811291].
21. A. Kehagias and K. Sfetsos, Phys. Lett. B **472**, 39 (2000) [arXiv:hep-ph/9905417].
22. C. D. Hoyle, U. Schmidt, B. R. Heckel, E. G. Adelberger, J. H. Gundlach, D. J. Kapner and H. E. Swanson, Phys. Rev. Lett. **86**, 1418 (2001) [arXiv:hep-ph/0011014]; E. G. Adelberger [EOT-WASH Group Collaboration], arXiv:hep-ex/0202008.
23. S.K. Lamoreaux, Phys. Rev. Lett. **78**, 5 (1997); arXiv:quant-ph/9907076.
24. N. Kaloper, J. March-Russell, G. D. Starkman and M. Trodden, Phys. Rev. Lett. **85**, 928 (2000) [arXiv:hep-ph/0002001].
25. G. R. Dvali, G. Gabadadze, M. Kolanovic and F. Nitti, Phys. Rev. D **64**, 084004 (2001) [arXiv:hep-ph/0102216].
26. K. Hirata *et.al*, Phys. Rev. Let. **58**, 1490 (1987); R.M. Bionta *et.al*, Phys. Rev. Let. **58**, 1494 (1987).
27. S. Cullen and M. Perelstein, Phys. Rev. Lett. **83**, 268 (1999) [arXiv:hep-ph/9903422]; V. D. Barger, T. Han, C. Kao and R. J. Zhang, Phys. Lett. B **461**, 34 (1999) [arXiv:hep-ph/9905474]; C. Hanhart, J. A. Pons, D. R. Phillips and S. Reddy, Phys. Lett. B **509**, 1 (2001) [arXiv:astro-ph/0102063].
28. C. Hanhart, D. R. Phillips, S. Reddy and M. J. Savage, Nucl. Phys. B **595**, 335 (2001) [arXiv:nucl-th/0007016].
29. T. Han, J. D. Lykken and R. J. Zhang, Phys. Rev. D **59**, 105006 (1999) [arXiv:hep-ph/9811350].
30. S. C. Kappadath *et al.*, Astron. Astrophys. Suppl. **120**, 619 (1996); D. A. Knitten *et al.*, Astron. Astrophys. Suppl. **120**, 615 (1996); D. J. Thompson *et al.*, Astrophys. J. Suppl. **101**, 259 (1995); *ibid.*, **107**, 227 (1996); R. C. Hartman *et al.*, Astrophys. J. Suppl. **123**, 79 (1999).
31. S. Hannestad and G. Raffelt, Phys. Rev. Lett. **87**, 051301 (2001) [arXiv:hep-ph/0103201].
32. L. J. Hall and D. R. Smith, Phys. Rev. D **60**, 085008 (1999) [arXiv:hep-ph/9904267].

33. S. Hannestad, Phys. Rev. D **64**, 023515 (2001) [arXiv:hep-ph/0102290].
34. S. Hannestad and G. G. Raffelt, Phys. Rev. Lett. **88**, 071301 (2002) [arXiv:hep-ph/0110067].
35. See, for example, www-glast.stanford.edu.
36. M. Fairbairn, Phys. Lett. B **508**, 335 (2001) [arXiv:hep-ph/0101131].
37. M. Fairbairn and L. M. Griffiths, JHEP **0202**, 024 (2002) [arXiv:hep-ph/0111435].
38. P. Jain, D. W. McKay, S. Panda and J. P. Ralston, Phys. Lett. B **484**, 267 (2000) [arXiv:hep-ph/0001031]; A. Goyal, A. Gupta and N. Mahajan, Phys. Rev. D **63**, 043003 (2001) [arXiv:hep-ph/0005030].
39. G. F. Giudice, R. Rattazzi and J. D. Wells, Nucl. Phys. B **595**, 250 (2001) [arXiv:hep-ph/0002178].
40. E. A. Mirabelli, M. Perelstein and M. E. Peskin, Phys. Rev. Lett. **82**, 2236 (1999) [arXiv:hep-ph/9811337]; K. M. Cheung and W. Y. Keung, Phys. Rev. D **60**, 112003 (1999) [arXiv:hep-ph/9903294].
41. For a summary of LEP results on graviton emission, see, G. Landsberg, arXiv:hep-ex/0105039. See also, P.Abreu *et al.*, The DELPHI Collaboration, Eur. Phys. J. C **17**, 53 (2000); G. Abbiendi *et al.*, The OPAL Collaboration, Eur. Phys. J. C **18**, 253 (2000).
42. M.Acciari *et al.*, The L3 collaboration, Phys. Lett. B **464**, 135 (1999).
43. J. A. Aguilar-Saavedra *et al.* [ECFA/DESY LC Physics Working Group Collaboration], TESLA Technical Design Report, Part III, Editors R.D. Heuer, D.J. Miller, F. Richard, P.M. Zerwas, 2001, arXiv:hep-ph/0106315.
44. G. Bernardi, presented at the *31st International Conference on High Energy Physics*, Amsterdam, July 2002; D. Acosta *et al.* [CDF Collaboration], arXiv:hep-ex/0205057.
45. L. Vacavant and I. Hinchliffe, J. Phys. G **27**, 1839 (2001).
46. H. Murayama and J. D. Wells, Phys. Rev. D **65**, 056011 (2002) [arXiv:hep-ph/0109004].
47. J. L. Hewett, Phys. Rev. Lett. **82**, 4765 (1999) [arXiv:hep-ph/9811356].
48. M. Bando, T. Kugo, T. Noguchi and K. Yoshioka, Phys. Rev. Lett. **83**, 3601 (1999) [arXiv:hep-ph/9906549].
49. E. Dudas and J. Mourad, Nucl. Phys. B **575**, 3 (2000) [arXiv:hep-th/9911019]; E. Accomando, I. Antoniadis and K. Benakli, Nucl. Phys. B **579**, 3 (2000) [arXiv:hep-ph/9912287]; S. Cullen, M. Perelstein and M. E. Peskin, Phys. Rev. D **62**, 055012 (2000) [arXiv:hep-ph/0001166].
50. E. Eichten, K. D. Lane and M. E. Peskin, Phys. Rev. Lett. **50**, 811 (1983).
51. ALEPH Collaboration, ALEPH Note CONF-2000-005; DELPHI Collaboration, DELPHI-CONF-464 (2001); L3 Collaboration, L3 Note 2650 (2001); OPAL Collaboration, OPAL Note PN469 (2001); D. Bourilkov, arXiv:hep-ex/0103039.
52. B. Abbott *et al.*, D0 Collaboration, Phys. Rev. Lett. **86**, 1156 (2001).
53. C. Adloff *et al.*, Phys. Lett. B **479**, 358 (2000); A. Zarnecki, presented at *Workshop on Higgs and Supersymmetry*, Orsay, France, March 2001.
54. F. Gianotti *et al.*, arXiv:hep-ph/0204087; A. Miagkov, talk presented at *Workshop on Higgs and Supersymmetry*, Orsay, France, March 2001.

55. T. G. Rizzo, Phys. Rev. D **60**, 115010 (1999) [arXiv:hep-ph/9904380]; H. Davoudiasl, Phys. Rev. D **60**, 084022 (1999) [arXiv:hep-ph/9904425]; Phys. Rev. D **61**, 044018 (2000) [arXiv:hep-ph/9907347].
56. S. B. Giddings and S. Thomas, Phys. Rev. D **65**, 056010 (2002) [arXiv:hep-ph/0106219]; S. Dimopoulos and G. Landsberg, Phys. Rev. Lett. **87**, 161602 (2001) [arXiv:hep-ph/0106295]; M. B. Voloshin, Phys. Lett. B **518**, 137 (2001) [arXiv:hep-ph/0107119].
57. T. G. Rizzo, JHEP **0202**, 011 (2002) [arXiv:hep-ph/0201228].
58. J. D. Lykken, Phys. Rev. D **54**, 3693 (1996) [arXiv:hep-th/9603133]; I. Antoniadis and M. Quiros, Phys. Lett. B **392**, 61 (1997) [arXiv:hep-th/9609209].
59. T. G. Rizzo and J. D. Wells, Phys. Rev. D **61**, 016007 (2000) [arXiv:hep-ph/9906234]; M. Masip and A. Pomarol, Phys. Rev. D **60**, 096005 (1999) [arXiv:hep-ph/9902467]; W. J. Marciano, Phys. Rev. D **60**, 093006 (1999) [arXiv:hep-ph/9903451].
60. P. Nath and M. Yamaguchi, Phys. Rev. D **60**, 116006 (1999) [arXiv:hep-ph/9903298].
61. T. G. Rizzo, Phys. Rev. D **64**, 015003 (2001) [arXiv:hep-ph/0101278].
62. N. Arkani-Hamed, Y. Grossman and M. Schmaltz, Phys. Rev. D **61**, 115004 (2000) [arXiv:hep-ph/9909411].
63. T. Appelquist, H. C. Cheng and B. A. Dobrescu, Phys. Rev. D **64**, 035002 (2001) [arXiv:hep-ph/0012100].
64. T. G. Rizzo, Phys. Rev. D **64**, 095010 (2001) [arXiv:hep-ph/0106336].
65. K. Agashe, N. G. Deshpande and G. H. Wu, Phys. Lett. B **514**, 309 (2001) [arXiv:hep-ph/0105084]; T. Appelquist and B. A. Dobrescu, Phys. Lett. B **516**, 85 (2001) [arXiv:hep-ph/0106140]; F. J. Petriello, arXiv:hep-ph/0204067.
66. W. D. Goldberger and M. B. Wise, Phys. Rev. Lett. **83**, 4922 (1999) [arXiv:hep-ph/9907447].
67. H. Davoudiasl, J. L. Hewett and T. G. Rizzo, Phys. Rev. Lett. **84**, 2080 (2000) [arXiv:hep-ph/9909255].
68. H. Davoudiasl and T. G. Rizzo, Phys. Lett. B **512**, 100 (2001) [arXiv:hep-ph/0104199].
69. B. C. Allanach, K. Odagiri, M. A. Parker and B. R. Webber, JHEP **0009**, 019 (2000) [arXiv:hep-ph/0006114].
70. For a discussion of this point, see, R. Contino, L. Pilo, R. Rattazzi and A. Strumia, JHEP **0106**, 005 (2001) [arXiv:hep-ph/0103104].
71. H. Davoudiasl, J. L. Hewett and T. G. Rizzo, Phys. Rev. D **63**, 075004 (2001) [arXiv:hep-ph/0006041].
72. W. D. Goldberger and M. B. Wise, Phys. Lett. B **475**, 275 (2000) [arXiv:hep-ph/9911457]; K. M. Cheung, Phys. Rev. D **63**, 056007 (2001) [arXiv:hep-ph/0009232]; J. L. Hewett and T. G. Rizzo, arXiv:hep-ph/0202155.
73. H. Davoudiasl, J. L. Hewett and T. G. Rizzo, arXiv:hep-ph/0010066.
74. H. Davoudiasl, J. L. Hewett and T. G. Rizzo, Phys. Lett. B **473**, 43 (2000) [arXiv:hep-ph/9911262]; A. Pomarol, Phys. Lett. B **486**, 153 (2000) [arXiv:hep-ph/9911294].

75. S. Chang, J. Hisano, H. Nakano, N. Okada and M. Yamaguchi, Phys. Rev. D **62**, 084025 (2000) [arXiv:hep-ph/9912498]; T. Gherghetta and A. Pomarol, Nucl. Phys. B **586**, 141 (2000) [arXiv:hep-ph/0003129].
76. J. L. Hewett, F. J. Petriello and T. G. Rizzo, arXiv:hep-ph/0203091.
77. D. J. Chung, L. L. Everett and H. Davoudiasl, Phys. Rev. D **64**, 065002 (2001) [arXiv:hep-ph/0010103].
78. J. Lykken and L. Randall, JHEP **0006**, 014 (2000) [arXiv:hep-th/9908076].
79. A. Hebecker and J. March-Russell, Nucl. Phys. B **608**, 375 (2001) [arXiv:hep-ph/0103214].



GORDON L. KANE

This page intentionally left blank

WEAK SCALE SUPERSYMMETRY — A TOP-MOTIVATED-BOTTOM-UP APPROACH

GORDON L. KANE
Michigan Center for Theoretical Physics, Randall Physics Lab,
University of Michigan
E-mail: gkane@umich.edu

CONTENTS:

- Introduction, Perspective – since particle physics beyond the SM is presently in an incoherent state, with lots of static, a long introduction is needed, including some history of the supersymmetry revolutions, physics not described by the SM, indirect evidence for low energy supersymmetry and how flavor physics should be approached.
- Derive the supersymmetric Lagrangian – the superpotential W.
- Soft supersymmetry breaking – underlying physics – L_{soft} – the MSSM.
- The μ opportunity – R-parity conservation.
- Count of parameters – constraints – measuring the parameters.
- Connecting the weak and unification scale.
- Derivation of the Higgs mechanism – in what sense does supersymmetry *explain* the Higgs physics.
- The Higgs spectrum – $\tan\beta$, Yukawa couplings, constraints.
- LEP Higgs physics – Tevatron Higgs physics can confirm the Higgs mechanism and coupling proportional to mass – Higgs sector measurements.
- $\tilde{g}$, $\tilde{N}$, $\tilde{C}$ – cannot in general measure $\tan\beta$ at hadron colliders.
- Effects of soft phases – all observables, not only CPV ones, $g_\mu - 2$, EDMs, $\tilde{g}$ phase, LSP CDM, possible connections to stringy physics.
- Phase structure of simple D-brane models.
- Tevatron superpartner searches, signatures.
- Extensions of the MSSM.
- The importance of low scale supersymmetry is not only that we learn of another profound aspect of our world, but also to provide a window to Planck scale physics, in order to connect string theory and our world.

1. INTRODUCTION

For about 400 years we have improved our understanding of the physical world until we discovered and tested the Standard Model (SM) of particle physics, which provides a complete description of our world, of all that we see. We know that the basic constituents are quarks and leptons, and we have a complete theory of the strong, weak, and electromagnetic forces.

We also know that much is unexplained, such as what is the cold dark matter (CDM) of the universe, and why is the universe made of matter rather than being an equal mixture of matter and antimatter, neither of which can be explained by the SM. Below I will make a longer list of questions the SM cannot answer. And there are conceptual reasons also why we expect to find new physics beyond the SM.

While a number of approaches to physics beyond the SM have been worth considering, only one so far has actually explained and predicted phenomena beyond the SM, the supersymmetric extension of the SM, which will be the focus of these lectures. As we will see, the supersymmetric SM has a number of successes, and as yet no failures. It is not yet a complete theory in the sense that we do not yet understand fully the physics of all of its parameters, but it is a complete effective theory because we can write the full effective Lagrangian of the theory. One of its important successes is that it can be a valid theory to very high energy scales or very short distances, near to the Planck scale. Another is that it is not sensitive to new physics at some high scale.

Superstring theories have become very attractive in recent years as well. They are formulated near the Planck scale with ten dimensions and presumably unbroken supersymmetry. String theories only predict or explain that there is a gravitational force, and that we live in at most 10 dimensions. What is exciting about them from our point of view is that they seem to be able to accommodate the SM forces and quarks and leptons, and possibly explain how these forces and particles originate. So in these lectures we will assume that the basic theory is a string theory at the Planck scale (loosely speaking). We do not distinguish string theory from M-theory for our top-motivated purposes, since the effective low scale theory from both will be parameterized by the same Lagrangian.

Historically, physics has progressed by one basic method, with experiment and theory intermingling as each level of the world was understood. That method will continue to work as supersymmetry is established experimentally, as its parameters (masses, flavor rotation angles, phases, vacuum expectation values) are measured, and as supersymmetry breaking is studied. But the historical approach can only take us to a broken supersymmetric theory near the unification and string scales. It cannot be used to learn the form of the 10D supersymmetric string theory. There is a barrier that can only be crossed by human imagination. To cross it we must know the Lagrangian of the broken supersymmetric theory near the unification and string scales, and we must understand the 10 D supersymmetric string theory very well. Then it will be possible to guess how to jump the barrier. In my view it will not be possible to do that until the broken supersymmetric theory near the Planck scale is known. No amount of thinking will tell us how to compactify the string theory, or to break supersymmetry, or to recognize the vacuum of the theory, because there is no practical way to recognize if one has it right.

Let's pursue this in a little more detail. Sometimes people argue that calculating fermion masses will be a convincing way to learn when a compactification is correct. But the hierarchy of fermion masses implies that will be very hard to do. The small masses are unlikely to arise at the tree level, but rather depend on non-renormalizable operators and possibly on supersymmetry breaking effects. So perhaps the large masses can be calculated, but not the smaller ones, and if the large ones have Yukawa couplings of order unity that will be common to many theories. It is of course known that huge numbers of manifolds give three families of chiral fermions. A little thought suggests very strongly that most of the usual "string phenomenology" is of a similar nature, and is very unlikely to point toward the correct vacuum. Indeed, suppose some string theorist already knew how to compactify and to break supersymmetry and to find the correct vacuum. How would they convince themselves, or anyone else?

However, the supersymmetry soft-breaking Lagrangian, L_{soft} may offer more hope for testing theory. The parameters of L_{soft} are measurable, though little has been known until recently about how to measure most of them, and much of these lectures will be about how to measure them. If a theorist has an approach to compactification and to breaking supersymmetry, then L_{soft} is likely to be calculable more easily than the full Yukawa matrix in that approach, and thus knowledge of L_{soft} may test ideas better than knowledge of the fermion masses. The parameters of L_{soft} may be less sensitive to uncertain higher order corrections (unless the leading term vanishes in which case the one-loop radiative correction is usually not hard to work out). The flavor structure of L_{soft} depends on the flavor structure of the Yukawas and may help untangle that. Progress will come from measuring L_{soft} at the weak scale, and extrapolating it to the unification scale. The patterns of the soft parameters may be typical of one approach or another to compactification or supersymmetry breaking or the vacuum structure, so the measured L_{soft} may focus attention toward particular solutions to these problems. Superpartners should be directly detected in the next few years, and once the initial excitement is past we will turn to the challenging and delightful opportunity to untangle the data and measure the Lagrangian.

Supersymmetry is an idea as old as the SM, and it has not been the most fashionable way to describe the real world in recent years. Consequently many students have not become familiar with supersymmetry as a practical theory, nor have they seen the arguments for its validity. Once superpartners are directly observed it will not be necessary to include these arguments, but at the present time there is some static in the messages most students get, so it is worthwhile to include some tables summarizing why classic supersymmetry is the best approach. In these largely pedagogical lectures I will also not focus on extensive referencing, with apologies to many authors. Some references are given to help the reader find the relevant additional papers. Many topics are integrated into these lectures, and most have been worked on by many authors, so I either have to provide extensive referencing or little referencing, and the latter seems reasonable here in a pedagogical treatment. For thorough referencing to the literature before the past three years the chapters in Ref. 1 are useful. I will in places follow the approach of Martin in Ref. 1, and he has very good referencing; the larger version of his chapter on the web is more valuable than the printed chapter.[2]

It is good to recall some of the history of supersymmetry. We can basically split it into five "revolutions":

HISTORY OF SUPERSYMMETRY REVOLUTIONS

1^{st}	1970-72	The idea
2^{nd}	1974	Supersymmetric relativistic quantum theory
3^{rd}	1975	Local supersymmetry, supergravity
4^{th}	1979-83	Supersymmetry solves many problems
5^{th}	2000-03	Higgs boson and superpartners observed

Next let us consider a list of important questions that the SM does not deal with. Consequently, these can point the way beyond the SM.

2. PHYSICS NOT DESCRIBED BY THE STANDARD MODEL

- Gravity
- Cosmological Constant
- Dark Energy

- What is (are) the inflation(s)?
- Strong CP problem
- Hierarchy problem
- How is the electroweak symmetry broken (EWSB)?
- Gauge coupling unification
- Matter asymmetry of the universe
- Cold dark matter
- 3 families
- Neutrino masses
- Values of quark and charged lepton masses
- Approximate Yukawa unification of bottom, tau, and perhaps top
- The value of the Higgs boson mass 115 GeV if the LEP signal is confirmed

The SM *cannot* account for or explain any of these. It can accommodate some of them. Any approach that claims to be making any progress (such as large extra dimension ideas) should be able to deal with some or most of these simultaneously. Where do they lead us? Supersymmetry of the form we are focusing on in these lectures is relevant to most or all of these. (There are additional reported deviations from the SM that could be relevant and arise from superpartner loops (a) the condition for charged current universality, or unitarity of the CKM matrix,[3] and (b) the number of neutrinos is slightly less than 3.[4])

3. THE HIERARCHY PROBLEM

The hierarchy problem is the SM problem that quantum corrections raise the Higgs boson mass up to the highest mass scale there is. It is a serious problem — as someone said, the quantum corrections are not only infinite, they are large. The high mass scales do not have to couple directly to the Higgs boson; the coupling can be through several loops, as Martin explains in some detail. All SM masses (W and Z and quarks and charged leptons) are proportional to m_h so if m_h is large they are too.

Supersymmetry was not invented or designed to solve this problem (contrary to what is often stated), but it did. If supersymmetry is unbroken then loops with particles cancel loops with their superpartners in general. For broken supersymmetry the effect is proportional to a power of some couplings times the square of the difference of the masses of superpartner pairs, and a log of mass ratios. Any solution of the hierarchy problem must be insensitive to high scales, and to higher order corrections. If an approach is claimed to deal with the hierarchy problem it must explain how the weak and gravitational scales are determined. Later when we discuss EWSB we examine in what sense supersymmetry provides this explanation. Sometimes people make connections between the cosmological constant problem and the Higgs hierarchy problem, but they are not the same because the calculation of the cosmological constant sums over all states, while the calculation of the Higgs mass only sums over states with SM gauge quantum numbers. Another way to think of the supersymmetry solution is that the Higgs doublet becomes a chiral supermultiplet so h and its superpartner have the same mass, and the fermion masses are not quadratically divergent so its superpartner mass is not quadratically divergent.

4. GAUGE COUPLING UNIFICATION

One of the most important things we have learned from LEP is that the gauge couplings unify at an energy above about 10^{16} GeV in a world described by a supersymmetric theory, though not in the SM. Further, where they meet points toward a unification with gravity near the Planck scale. Together these make one of the strongest indications of the validity of the view of physics at the foundation of these lectures. Any other view has to claim this unification is a coincidence! The gauge coupling unification implies two important results:

(1) The underlying theory is perturbative up to the unification scale. Sometimes it is said there should be a desert (apart from the superpartners) but that is not so — only that whatever is in that range (such as right handed neutrinos) does not destroy the perturbativity of the theory.

(2) Physics is simpler at or near the unification scale. That need not have happened — nothing in the SM implies such a result.

There is another important clue. The supersymmetric gauge coupling unification misses by about 10%. More precisely, the experimental value of the strong coupling α_3 is about 10-15% lower than the value computed by running down theoretically from the point were the SU(2) and U(1) couplings meet. The details are interesting here — the one-loop result is somewhat small because of a cancellation, and the two-loop contribution therefore not negligible. If one only took into account the one-loop effect the theoretical value would be close to the experimental one but the two-loop effect increases the separation. Nature is kind here, on the one hand giving us information about the need for supersymmetric unification, and on the other giving us a further clue about the physics near the unification scale, or about particles that occur in the "desert" and change the running somewhat.

5. (INDIRECT) EVIDENCE FOR WEAK SCALE SUPERSYMMETRY

We have described two of the strongest pieces of evidence for weak scale supersymmetry. The third and to some the strongest is that this approach can explain the central problem of how the electroweak symmetry is broken — we will consider that in great detail after we derive the supersymmetric Lagrangian. First we list here additional evidence for weak scale supersymmetry. Sometimes people wrongly imagine that supersymmetry was invented to explain some of what it explains so the approximate date when it was realized that each of these pieces of evidence existed is listed. Of course the theory existed even before it was realized that it solved these problems — it was not invented for any of them. For completeness we include the evidence we have already examined.

1980 — Can stabilize hierarchy of mass scales.
1982 — Provides an explanation for the Higgs mechanism.
1982 — Gauge coupling unification.
1982 — Provides cold dark matter candidate.
1982 — Heavy top quark predicted.
1992 — Can explain the baryon asymmetry of the universe.
1993 — Higgs boson must be light in general supersymmetric theory.
1990 — Realization that either superpartners are light enough to find at LEP, or their effects on precision data must be very small and unlikely to be observed at LEP/SLC. Supersymmetry effects at lower energies arise only from loops, which explains why the SM works so well even though it is incomplete.
1982/1995 — Starting from a high scale with a value for $\sin^2\theta_W$ of 3/8, which arises in any theory with a unified gauge group that contains SU(5), and also in a variety of string-based theories, the value for $\sin^2\theta_W$ at the weak scale is 0.2315 and agrees very accurately with the measured value.

Some of these successes are explanations, and some are correct predictions. It is also very important that all of them are simultaneously achieved — often efforts to deal with the real world can apparently work for one effect, but cannot describe the range of phenomena we know.

There are theoretical motivations for low energy supersymmetry too.

It is the last four-dimensional space-time symmetry not yet known to be realized in some way in nature, it adds a fermionic or quantum structure to space-time, it allows theories to be extrapolated to near the Planck scale where they can be related to gravity, local supersymmetry is supergravity which suggests a connection of the supersymmetric SM to gravity, it allows many problems in string theory and string field theory to be solved, including stabilizing the string vacuum. It is expected, though not yet demonstrated, that low energy supersymmetry is implied by string theory. Not all of these necessarily require low energy supersymmetry. In any case, improving the theory is nice but is not strong motivation for something to exist in nature, so we have emphasized the evidence that actually depends on data.

6. CURRENT LIMITS ON SUPERPARTNER MASSES

The general limits from direct experiments that could produce superpartners are not very strong. They are also all model dependent, sometimes a little and sometimes very much. Limits from LEP on charged superpartners are near the kinematic limits except for certain models, unless there is close degeneracy of the charged sparticle and the LSP, in which case the decay products are very soft and hard to observe, giving weaker limits. So in most cases charginos and charged sleptons have limits of about 95 GeV. Gluinos and squarks have typical limits of about 250 GeV, except that if one or two squarks are lighter the limits on them are much weaker. For stops and sbottoms the limits are about 85 GeV separately.

There are no general limits on neutralinos, though sometimes such limits are quoted. It is clear no general limits exist — suppose the LSP was pure photino. Then it could not be produced at LEP through a Z which does not couple to photinos, and suppose selectrons were very heavy so it's production via selectron exchange is very small in pair or associated production. Then no cross section at LEP is large enough to set limits. There are no general relations between neutralino masses and chargino or gluino masses, so limits on the latter do not imply limits on neutralinos. In typical models the limits are $M_{LSP} \gtrsim 40$ GeV, $M_{\tilde{N}_2} \gtrsim 85$ GeV. Superpartners get mass from both the Higgs mechanism and from supersymmetry breaking, so one would expect them to typically be heavier than SM particles. All SM particles would be massless without the Higgs mechanism, but superpartners would not. Many of the quark and lepton masses are small presumably because they do not get mass from Yukawa couplings of order unity in the superpotential, so one would expect naively that the normal mass scale for the Higgs mechanism was of order the Z or top masses. In models chargino and neutralino masses are often of order Z and top masses, with the colored gluino mass a few times the Z mass.

There are no firm indirect limits on superpartner masses. If the $g_\mu - 2$ deviation from the SM persists as the data and theory improve the first such upper limits will be deduced. If in fact supersymmetry explains all that we argue above it is explaining, particularly the EWSB, then there are rather light upper limits on superpartner masses, but they are not easily made precise. Basically, what is happening is that EWSB produces the Z mass in terms of soft-breaking masses, so if the soft-breaking masses are too large such an explanation does not make sense. The soft parameters that are most sensitive to this issue are M_3 (basically the gluino mass) and μ which strongly affects the chargino and neutralino masses. Qualitatively one therefore expects rather light gluino, chargino, and neutralino masses. If one takes this argument *seriously* one is led to expect $M_{\tilde{g}} \lesssim 500$ GeV; $M_{\tilde{N}_2}$, $M_{\tilde{C}} \lesssim 250$ GeV; and $M_{\tilde{N}_1} \lesssim 100$ GeV. These are upper limits, seldom saturated in models. There are no associated limits on sfermions. They suggest that these gaugino states should be produced in significant quantities at the Tevatron in the next few years.

There are some other clues that some superpartners may be light. If the baryon number is generated at the EW phase transition then the lighter stop and charginos should be lighter than the top. If the LSP is indeed the cold dark matter, then at least one scalar fermion is probably light enough to allow enough annihilation of relic LSPs, but there are loopholes to this argument.

7. WHAT CAN SUPERSYMMETRY EXPLAIN?

Supersymmetry can explain much that the SM cannot, as described above, particularly the Higgs physics as we will discuss in detail below. Sometimes people who do not understand supersymmetry say it can "explain or fit anything". In fact it is the opposite. Supersymmetry is a full theory, and all that is unknown is the masses (which are matrices in flavor space) and the vacuum expectation values, exactly as for the SM. There are many conceivable phenomena that supersymmetry could not explain, including sharp peaks in spectra at colliders, a world with no Higgs boson below about 200 GeV, a top quark lighter than the W, deviations from SM predictions greater than about 1% for any process with a tree-level SM contribution (including Z decay to $c\bar{c}$), leptoquarks, wide WW or ZZ resonances, excess high-P_t leptons at HERA, large violations of μ/e universality, and much more. None of these has occurred, consistent with supersymmetry, but a number of them have been reported and then gone away, and supersymmetry did not "explain" them while they were around. Supersymmetry alone also cannot explain some real questions such as why there are three families or the $\mu - \tau$ mass ratio.

8. HOW DOES FLAVOR PHYSICS ENTER THE THEORY?

The "flavor problem" is one of the most basic questions in physics. By this usually three questions are intended. First, why are there three families of quarks and leptons, and not more or less? Second, why are the symmetry eigenstates different from the mass eigenstates? Third, why do the quarks and leptons have the particular mass values they do? Supersymmetry does not provide the answers to those questions directly, though it will affect the answers. The second and third questions are of course related, but different. We could know the answer to the second question but not the third. For example, the actual values of masses of the lighter quarks and leptons could depend on operators beyond the tree level in the superpotential. The u,d,e masses are so small that they could get large corrections from a number of sources.

Where to look for those answers is not something that is agreed on — many people have tried to understand flavor physics at the TeV scale. Supersymmetry does suggest where to find the answers. Supersymmetry is like the SM in that it accommodates the three families and the flavor rotations but does not explain them. It clearly suggests that the flavor physics has basically entered once the superpotential is determined, i.e. when the Yukawa couplings in the superpotential are fixed. That occurs as soon as a 4D theory is written and depends in a basic way on the string physics and on the compactification and on the determination of the string vacuum. Since the superpotential of the observable sector does not know about supersymmetry breaking, the basic flavor physics probably does not depend on supersymmetry breaking either, though how the flavor physics manifests itself in L_{soft} may. That in turn suggests that learning the Yukawa couplings and the off-diagonal structure of the trilinears and squark and slepton mass matrices can guide us to the formulation of how to compactify and how to find the string vacuum, and can test ideas about such physics. The role of supersymmetry breaking is unclear. For example, the structure of the trilinear soft-breaking terms can be calculated in terms of the Yukawa couplings and their derivatives, but may depend on how supersymmetry is broken as well.

An important point is that we are likely to learn more from data on the superpartner masses than we did from the quark masses (as we will discuss later). That is because the parameters of L_{soft} are rather directly related to an underlying theory, while the quark and lepton masses probably are not. Probably what we learned from the fermion masses is that some Yukawa couplings are of order unity while others are small at tree level, arising from non-renormalizable operators and/or breaking of discrete symmetries and supersymmetry. The masses of the first and second family quarks and leptons are probably determined by or very sensitive to small effects that are hard to calculate (the first family masses are in the MeV range, while the theory makes sense for the 100 GeV range), while the squark and slepton masses, and probably the phases, and the approximate size of off-diagonal flavor dependent squark and slepton masses and trilinears all generally emerge from the theory at leading order and are thus much more easily interpretable than the fermion masses.

The next question is how to measure the flavor-dependent elements of L_{soft}, which has 112 flavor-dependent parameters not counting neutrino physics. Although certain combinations of them affect collider physics, and the masses of the mass eigenstates can be measured at colliders, most of them affect rare decays, mixing, and CP violation experiments. Collider studies of superpartners may tell us little about flavor physics directly. If they are to have an observable effect, of course, the supersymmetric contributions to the decays and mixing and CP violation must be significant, which is most likely for processes that are forbidden at tree level such as $b \to s + \gamma$, mixing, penguin diagrams, $\mu \to e + \gamma$, etc.

The absence of flavor-changing decays for many systems puts strong constraints on some soft parameters. If the off-diagonal elements of the squark or slepton mass matrices and trilinears were of order the typical squark or slepton masses then in general there would be large flavor mixing effects, since the rotations that diagonalize the quarks and charged leptons need not diagonalize the squarks and sleptons. However, many of the constraints from flavor-changing processes in the literature have been evaluated with assumptions that may not apply, so people should reevaluate them for any approach they find attractive for other reasons. Much effort has gone into constructing models of L_{soft} that guarantee without tuning the absence of FCNC, and several approaches exist. If one of them is confirmed when data exists it will be a major clue to the structure of the high energy theory. Our view that the flavor physics is determined at the high scale implies that the resulting structure of the squark and slepton mass matrices, and the trilinear coefficients, is also determined at the high scale and not by TeV-scale dynamics. Thus the absence of FCNC is not and should not be explained by an effective supersymmetric theory. Rather, the pattern of soft-breaking terms that is measured and gives small FCNC will help us learn about the underlying (presumably string) theory. Similar remarks could be made about proton decay.

Once the soft flavor parameters are measured it is necessary to deduce their values at the unification or string scales in order to compare with the predictions of string-based models, or to stimulate the development of string-based models. There are two main issues that arise. One is how to relate measured values of the CKM matrix and soft parameters to the values of Yukawa matrices and soft parameters at the unification scale, assuming no other physics enters between the scales. This is subtle because the number of independent parameters is considerably less than the number of apparent parameters in L_{soft} and the superpotential Yukawas, as discussed in Section 17, and the RGE running will for a generic procedure involve non-physical parameters. This problem has recently been solved,[5] giving a practical technique to convert measurements into the form of the high scale theory.

The second issue is that presumably there is not a desert between the high and low scales. Both gauge coupling unification and radiative electroweak symmetry breaking imply that no part of the theory becomes strongly interacting below the unification scale. But we expect heavy RH neutrinos, axion physics, and perhaps "exotic" states such as those often generated in stringy models, e.g. vector multiplets, fractionally charged uncolored fermions, etc. This issue has not been studied much.[6] Perhaps by examining appropriate combinations of quantities for the RGE running, and by imposing appropriate conditions, it will be possible to use consistency checks to control the effects of intermediate scale physics.

9. DERIVATION OF THE SUPERSYMMETRY LAGRANGIAN

In order to understand the predictions and explanations of supersymmetry, particularly for the Higgs sector, we must learn the derivation of the supersymmetry Lagrangian. I will present the arguments fully though not all the algebra. I will largely follow the approach of Martin.

Consider a massless and therefore two-component fermion, ψ whose superpartner is a complex scalar ϕ. Both have two real degrees of freedom. But in the off-shell field theory the fermion is a four-component field with four degrees of freedom, and we want supersymmetry to hold for the full field theory. So we introduce an additional complex scalar F so that there are four scalar degrees of freedom also. F is called an auxiliary field. The combined fields (ψ, ϕ, F) are called a chiral superfield or chiral supermultiplet. I will not be systematic or careful about the two-component vs. four-component notation since the context usual is clear. The Lagrangian can be taken to be

$$-L_{free} = \sum_i (\partial^\mu \phi_i^* \partial_\mu \phi_i + \bar{\psi}_i \gamma^\mu \partial_\mu \psi_i + F_i^* F_i). \qquad (1)$$

The sum is over all chiral supermulitplets in the theory. Note that the dimensions of F are $[F] = m^2$. The Euler-Lagrange equations of motion for F are $F = F^* = 0$, so on-shell we revert to only two independent degrees of freedom. One can define supersymmetry transformations that take bosonic degrees of freedom into fermionic ones; we will look briefly at them later. The supersymmetry transformations can be defined so that L_{free} is invariant. Next we write the most general set of renormalizable interactions,

$$L_{chiral} = L_{free} + L_{int} \qquad (2)$$

$$L_{int} = -\frac{1}{2} W^{ij} \psi_i \psi_j + W^i F_i + c.c. \qquad (3)$$

Here W^{ij} and W^i are any functions of only the scalar fields, remarkably, and W^{ij} is symmetric. If W^{ij} or W^i depended on the fermion or auxiliary fields the associated terms would have dimension greater than four, and would therefore not be renormalizable. There can be no terms in L_{int} that depend on ϕ_i^* or ϕ_i since such terms would not transform into themselves under the supersymmetry transformations.

Now imagine supersymmetry transformations that mix fermions and bosons, $\phi \to \phi + \varepsilon\psi, \psi \to \psi + \varepsilon\phi$. We should go through these transformations in detail with indices, but one can see the basic argument simply.

Here ε must be a spinor so each term behaves the same way in spin space, and we can take ε to be a constant spinor in space-time, and infinitesimal. Then the variation of the Lagrangian (which must vanish or change only by a total derivative if the theory is invariant under the supersymmetry transformation) contains two terms with four spinors:

$$\delta L_{int} = -\frac{1}{2}\frac{\delta W^{ij}}{\delta \phi_k}(\varepsilon\psi_k)\psi_i\psi_j - \frac{1}{2}\frac{\delta W^{ij}}{\delta \phi_k^*}(\varepsilon^\dagger\psi_k^\dagger)\psi_i\psi_j + c.c. \tag{4}$$

Neither term can cancel against some other term. For the first term there is a Fierz identity $(\varepsilon\psi_i)(\psi_j\psi_k) + (\varepsilon\psi_j)(\psi_k\psi_i) + (\varepsilon\psi_k)(\psi_i\psi_j) = 0$, so if and only if $\delta W^{ij}/\delta\phi_k$ is totally symmetric under interchange of i,j,k the first term vanishes identically For the second term the presence of the hermitean conjugation allows no similar identity, so it must vanish explicitly, which implies $\delta W^{ij}/\delta\phi_k^* = 0$,and thus W^{ij} cannot depend on ϕ^*! W^{ij} must be an analytic function of the complex field ϕ.

Therefore we can write

$$W^{ij} = M^{ij} + y^{ijk}\phi_k, \tag{5}$$

where M^{ij} is a symmetric matrix that will be the fermion mass matrix, and y^{ijk} can be called Yukawa couplings since it gives the strength of the coupling of boson k with fermions i, j; y^{ijk} must be totally symmetric. Then it is very convenient to define

$$W = \frac{1}{2}M^{ij}\phi_i\phi_j + \frac{1}{6}y^{ijk}\phi_i\phi_j\phi_k \tag{6}$$

and $W^{ij} = \delta^2 W/\delta\phi_i\delta\phi_j$. W is the "superpotential", an analytic function of ϕ, and a central function of the formulation of the theory. W is fully supersymmetric and gauge invariant and Lorentz invariant, and an analytic function of ϕ (i.e. it cannot depend explicitly on ϕ^*), so it is highly constrained. It determines the most general non-gauge interactions of the chiral superfields.

A similar argument for the parts of δL_{int} which contain a spacetime derivative imply that W^i is determined in terms of W as well,

$$W^i = \frac{\delta W}{\delta\phi_i} = M^{ij}\phi_j + \frac{1}{2}y^{ijk}\phi_j\phi_k. \tag{7}$$

Because interactions are now present, the equations for F are non-trivial,

$$F_i = -W_i^*. \tag{8}$$

The scalar potential is related to the Lagrangian by $L = T - V$, so

$$V = \sum_i |F_i|^2 \tag{9}$$

This contribution is called an "F-term", and is automatically bounded from below, an important improvement.

The above analysis was appropriate for chiral superfields, which will contain the fermions and their superpartners. Now we repeat the logic for the gauge supermultiplets that contain the gauge bosons and their superpartners. Initially they are massless gauge bosons, like photons, A_μ^a, with gauge index a, and two degrees of freedom. Their superpartners are two-component spinors λ^a. But as above, off shell the fermion has four degrees of freedom, while the massive boson has three, the two transverse polarizations and a longitudinal polarization. So again it is necessary to add an auxiliary field, a real one since only one degree of freedom is needed, called D^a. Then the Lagrangian has additional pieces

$$L_{gauge} = -\frac{1}{4}F_{\mu\nu}^a F_a^{\mu\nu} - i\lambda^{\dagger a}\gamma^\mu D_\mu \lambda^a + \frac{1}{2}D^a D^a, \tag{10}$$

where

$$F_{\mu\nu}^a = \partial_\mu A_\nu^a - \partial_\nu A_\mu^a - g f^{abc} A_\mu^b A_\nu^c \tag{11}$$

and the covariant derivative is

$$D_\mu \lambda^a = \partial_\mu \lambda^a - g f^{abc} A_\mu^b \lambda^c. \tag{12}$$

Note that the notation is unfortunate, with both the covariant derivative and the new field being denoted by the standard "D". Also, I have not been careful about two component vs. four component spinors. It is crucial for gauge invariance that the same coupling g appears in the definition of the tensor F and in the covariant derivative. Lagrangians always have to contain all of the terms allowed by gauge invariance, etc., and here we can see another term to add,

$$(\phi_i^* T^a \phi_i) D^a. \tag{13}$$

There is one more term that can be added that mixes the fields, $\lambda^{\dagger a}(\psi^\dagger T^a \phi)$, and its conjugate, with an arbitrary dimensionless coefficient. Requiring the entire Lagrangian to be invariant under supersymmetry transformations determines the arbitrary coefficient and gives a resulting Lagrangian

$$L = L_{gauge} + L_{chiral} + g_a(\phi^* T^a \phi)D^a - \sqrt{2}g_a[(\phi^* T^a \psi)\lambda^a + \lambda^{\dagger a}(\psi^\dagger T^a \phi)] \tag{14}$$

where all derivatives in earlier forms are replaced by covariant ones. Remarkably, the requirement of supersymmetry fixed the couplings of the last terms to be gauge couplings even though they are not normal gauge interactions! The equations of motion for D^a give $D^a = -g(\phi^* T^a \phi)$, so the scalar potential is

$$V = F^{*i}F_i + \frac{1}{2}\sum_a D^a D^a = |\partial W/\partial \phi_i|^2 + \sum_a g_a^2(\phi^* T^a \phi)^2. \tag{15}$$

The sum is over $a = 1, 2, 3$ for the three gauge couplings. The two terms are called "F-terms" and "D-terms". Remarkable, the unbroken supersymmetric theory gives a scalar potential bounded from below. On the one hand that is good since unbounded potentials are a problem, but it also implies that the Higgs mechanism cannot happen for unbroken supersymmetry since the potential will be minimized at the origin. In the above,

$$L_{chiral} = D^\mu \phi_i^* D_\mu \phi_i + \bar{\psi}_i \gamma^\mu D_\mu \psi_i \tag{16}$$

$$+(\frac{1}{2}M_{ij}\psi_i\psi_j + \frac{1}{2}y^{ijk}\phi_i\psi_j\psi_k + c.c.) + F_i^* F_i.$$

This completes the derivation of the unbroken supersymmetry Lagrangian.

10. NON-RENORMALIZATION THEOREM

For unbroken supersymmetry there is a very important result, called the non-renormalization theorem, that is very useful for building models to relate the theory to the real world. Because of this result the supersymmetry fields get a wave function renormalization only, so they have the familiar log running of couplings and masses, but no other renormalizations. Consequently the parameters of the superpotential W are not renormalized, in any order of perturbation theory. In particular, terms that were allowed in W by gauge invariance and Lorentz invariance are not generated by quantum corrections if they are not present at tree level, so no F-terms are generated if they are initially absent. If there is no μ-term in the superpotential (see below), none is generated. The non-renormalization theorem is difficult to probe without extensive formalism, so I just state it here. References and a pedagogical derivation are given in reference 7.

11. TOWARD SOFTLY-BROKEN SUPERSYMMETRY WITH A TOY MODEL

Consider the Wess-Zumino model, with,

$$W = \frac{m}{2}\phi\phi + \frac{g}{6}\phi\phi\phi, \tag{17}$$

and

$$L = (\partial\phi)^2 + i\Psi^\dagger\bar{\sigma}^\mu\partial_\mu\Psi - F_\phi^* F_\phi + (\frac{1}{2}W_{\phi\phi}\Psi\Psi - W_\phi F_\phi + c.c.). \tag{18}$$

This is written in two component notation. $W_\phi = -F_\phi^* = m\phi + \frac{g}{2}\phi\phi$ is the derivative of the superpotential with respect to ϕ,and $W_{\phi\phi}$ the second derivative. We put $\phi = (A + iB)/2$ and $F_\phi = (F + iG)/2$, where A, B, F, G are real scalars, and switch to four component notation. Under the supersymmetry transformations, with ε a constant spinor,

$$\delta A = \bar{\varepsilon}\gamma_5\Psi\,, \tag{19}$$

$$\delta B = -\bar{\varepsilon}\Psi\,, \tag{20}$$

$$\delta\Psi = F\varepsilon - G\gamma_5\varepsilon + \gamma^\mu\partial_\mu\gamma_5 A\varepsilon + \gamma^\mu\partial_\mu B\varepsilon\,, \tag{21}$$

$$\delta F = -\bar{\varepsilon}\gamma^\mu\partial_\mu\Psi\,, \tag{22}$$

$$\delta G = -\bar{\varepsilon}\gamma_5\gamma^\mu\partial_\mu\Psi\,, \tag{23}$$

the Lagrangian changes by a total derivative, so the action is invariant with the usual assumptions.

Now substitute for W_ϕ and $W_{\phi\phi}$ etc. Then the Lagrangian is

$$\begin{aligned} L = {} & \frac{1}{2}(\partial A)^2 + \frac{1}{2}(\partial B)^2 + \frac{i}{2}\bar{\psi}\gamma^\mu\partial_\mu\psi + \frac{1}{2}m\bar{\psi}\psi \\ & + \frac{g}{\sqrt{2}}A\bar{\psi}\psi - \frac{ig}{\sqrt{2}}B\bar{\psi}\gamma_5\psi - \frac{1}{2}(F^2 + G^2) \\ & - \frac{m}{2}(2AF - 2BG) - \frac{g}{2\sqrt{2}}(F(A^2 - B^2) - 2GAB). \end{aligned} \tag{24}$$

Now the equations of motion for F, G are

$$F = -mA - \frac{g}{2\sqrt{2}}(A^2 - B^2), G = mB + \frac{g}{\sqrt{2}}AB. \tag{25}$$

Substituting these gives interaction vertices $\frac{mg}{2\sqrt{2}}A(A^2 - B^2)$.

With one coupling strength g and one mass m the full Lagrangian is supersymmetric. (Note that without supersymmetry there can be four different masses and four different couplings, so there are six relations predicted by supersymmetry which only allows one mass and one coupling.) But when supersymmetry is broken we expect the masses to differ. Suppose we allow four different masses, m_A, m_B, m_ψ, and m_g, where the last is the mass that is needed in some terms to give each term dimension four, so it multiplies g. It's clear how to rewrite the Lagrangian with these separate masses. There are four three-particle vertices, $A\bar{\psi}\psi$, $A^3, AB^2, B\bar{\psi}\psi$. Now if we write the expression for a tadpole graph,

$$\langle 0 \,|L|\, A\rangle = \frac{g}{\sqrt{2}} \left\{ 4m_\psi \int \frac{d^4p}{p^2 - m_\psi^2} - m_g \int \frac{d^4p}{p^2 - m_B^2} \right.$$
$$\left. - 3m_g \int \frac{d^4p}{p^2 - m_A^2} \right\}, \tag{26}$$

we see that in general this has a quadratic divergence, which cancels in the supersymmetry limit as expected. The fermion loop gives a minus sign, the factor of 4 in the first term arises from $Tr(\gamma^\mu p_\mu + m_\psi) = Tr m_\psi = 4m_\psi$, and the 3 in the last from the A^3. But — and here is the important point — the divergence still cancels if $m_A \neq m_B \neq m_g$, but not if $m_\psi \neq m_g$. Thus extra contributions to boson masses do not reintroduce quadratic divergences — they are called "soft" supersymmetry breaking. But extra contributions to fermion masses do lead to quadratic divergences, "hard" supersymmetry breaking. This result is true to all orders in perturbation theory, though this pedagogical argument does not show it. Some of the results are obvious since couplings proportional to masses will not introduce quadratic divergences, but it is still helpful to see the supersymmetry structure. After the supersymmetry breaking there are three masses and one coupling, so there are still four tests that the theory is a broken supersymmetric one.

To understand the general structure of supersymmetry breaking better, recall how symmetry breaking works in the SM. It is not possible to break the $SU(2) \times U(1)$ symmetry from within the SM. So a new sector, the Higgs sector is needed. Interactions are assumed in the Higgs sector that give a potential with a minimum away from the origin, so the Higgs field gets a vev which breaks the symmetry. To generate mass for W, Z, q, l an interaction is needed to transmit the breaking to the "visible" particles W, Z, q, l. For fermions this interaction is $L_{fermion} = g_e \bar{e}_L e_R h + cc \rightarrow g_e \mathrm{v} \bar{e} e$ after h gets a vev for the fermions, and we can identify $m_e = g_e \mathrm{v}$. Similarly, for the gauge bosons the Lagrangian term $(D^\mu h)(D_\mu h) \rightarrow g^2 hh W^\mu W_\mu \rightarrow g^2 \mathrm{v}^2 W^\mu W_\mu$ giving W, Z masses. The fundamental symmetry breaking is spontaneous (h gets a vev), but the effective Lagrangian appears to have explicit breaking.

The situation is very similar for supersymmetry. It is not possible to break supersymmetry in the "visible" sector, i.e. the sector containing the superpartners of the SM particles. A separate sector is needed where supersymmetry is broken. Originally it was called the "hidden" sector, but that is not a good name since it need not be really hidden. Then there must be some interaction(s) to transmit the breaking to the visible sector. Since the particles of both sectors interact gravitationally, gravity can always transmit the breaking. Other interactions may as well. We will have to find out how the breaking is transmitted from data on the superpartners, their masses and decays and phases and flavor rotations. Different ways of transmitting the breaking give different patterns of the soft parameters that we discuss below. A significant complication is that the effects of the supersymmetry breaking are mixed up with effects of the transmission. All the effects of the supersymmetry breaking and of the way it is transmitted, for any theory, are embedded in the soft-breaking Lagrangian that we turn to studying.

12. THE SOFT-BREAKING LAGRANGIAN

The (essentially) general form of L_{soft} is[8]

$$L_{soft} = \frac{1}{2}(M_\lambda \lambda^a \lambda^a + c.c.) + m_{ij}^2 \phi_j^* \phi_i \tag{27}$$

$$+(\frac{1}{2} b_{ij} \phi_i \phi_j + \frac{1}{6} a_{ijk} \phi_i \phi_j \phi_k + c.c.)$$

This obviously breaks supersymmetry since only scalars and gauginos get mass, not their superpartners. It is soft as in our example above because it can be proved to not introduce any quadratic divergences. Models for supersymmetry breaking, however they originate, in string theory or supergravity or dynamically, all lead to this form. We will write it for the SSM shortly.

If all fields carry gauge quantum numbers there are terms that could be added to this without generating quadratic divergences, such as $\phi_i^* \phi_j \phi_k$, but such terms seldom arise in models so they are usually ignored.[9] If such terms are truly absent once measurements are analyzed, their absence may be a clue to how supersymmetry is broken and transmitted.

13. THE MINIMAL SUPERSYMMETRIC STANDARD MODEL

To write the supersymmetric SM we first take all of the quarks and leptons and put them in chiral superfields with superpartners. For each set of quantum numbers, such as up quarks or electrons, the scalar, fermion, and auxiliary fields (ϕ, ψ, F) form a supermultiplet in the same sense as (n, p) form a strong isospin doublet or (ν_e, e) form an electroweak doublet. All superpartners are denoted with a tilde, and there is a superpartner for each spin state of each fermion — that is important since the SM treats fermions of different chirality differently. The gauge bosons are put in vector superfields with their fermionic superpartners. Since W is analytic in the scalar fields, we cannot include the complex conjugate of the scalar field as in the SM to give mass to the down quarks, so there must be two Higgs doublets (or more) in supersymmetry, and each has its superpartners. The requirement that the trace anomalies vanish so that the theories stay renormalizable, $TR(Y^3) = TR(T_{3L}^2 Y) = 0$, also implies the existence of the same two Higgs doublets. (The relevance of anomalies may seem unclear since we are only writing an effective theory, while anomaly conditions only need to be satisfied for the full theory. But if the anomaly conditions are not satisfied it may introduce a sensitivity to higher scales that the effective theory should not have.)

We proceed by first constructing the superpotential so we can calculate the F-terms, and then writing the Lagrangian, following equation 6 and summing over all the particles. The most general superpotential, if we don't extend the SM and don't include RH neutrinos, is

$$W = \bar{u} Y_u Q H_u - \bar{d} Y_d Q H_d - \bar{e} Y_e L H_d + \mu H_u H_d. \tag{28}$$

All the fields are chiral superfields. The bars over u, d, e are in the sense of Martin's notation, specifying the conjugate fields. The signs are conventional so that masses later are positive. Indices are suppressed — for example, the fourth and first terms are

$$\mu (H_u)_\alpha (H_d)_\beta \varepsilon_{\alpha\beta} \text{ and } \bar{u}_{ai} (Y_u)_{ij} Q^a_{j\alpha} (H_u)_\beta \varepsilon_{\alpha\beta}. \tag{29}$$

The Yukawa couplings Y_u etc. are dimensionless 3×3 family matrices that determine the masses of quarks and leptons, and the angles and phase of the CKM matrix after H_u^0 and H_d^0 get vevs. They also contribute to the squark-quark-higgsino couplings etc. since the fields in W are superfields containing all the components. This is the most general superpotential for the SSM if we assume baryon and lepton number are conserved (we'll return to this question). To see the structure more explicitly we can use the approximations

$$Y_u \approx \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & Y_t \end{pmatrix}, \quad Y_d \approx \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & Y_b \end{pmatrix}, \quad Y_e \approx \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & Y_\tau \end{pmatrix}, \tag{30}$$

which gives

$$W = Y_t(\bar{t}tH_u^0 - \bar{t}bH_u^+) - Y_b(\bar{b}tH_d^- - \bar{b}bH_d^0) \tag{31}$$

$$-Y_\tau(\bar{\tau}\nu_\tau H_d^- - \bar{\tau}\tau H_d^0) + \mu(H_u^+ H_d^- - H_u^0 H_d^0)$$

There are also other interactions from W such as vertices $H_u^0 t_R^* t_L$, $\tilde{H}_u^0 t_R^* \tilde{t}_L$, $\tilde{H}_u^0 \tilde{t}_R t_L$, etc., all with the same strength Y_t. All of them are measurable, and it will be an important check of supersymmetry to confirm they are all present with the same strength. All are dimensionless, so supersymmetry-breaking will only lead to small radiative corrections to these coupling strengths. In general one goes from one to another of these by changing any pair of particles into superpartners.

Before we turn to writing the full soft-breaking Lagrangian, we first look at two significant issues that depend on how supersymmetry is embedded in a more basic theory.

14. THE μ OPPORTUNITY

The term $\mu H_u H_d$ in the superpotential leads to a term in the Lagrangian

$$L = \ldots\ldots + \mu(\tilde{H}_u^+ \tilde{H}_d^- - \tilde{H}_u^0 \tilde{H}_d^0) + \ldots\ldots \tag{32}$$

which gives mass terms for higgsinos in the chargino and neutralino mass matrices, so μ enters there. This term also contributes to the scalar Higgs potential from the F-terms,

$$V = \ldots\ldots |\mu|^2 \left(\left|H_u^0\right|^2 + \left|H_d^0\right|^2 + \ldots\right) + \ldots\ldots \tag{33}$$

so these terms affect the Higgs mass, and F-terms also give contributions to the Lagrangian that affect the squark and slepton mass matrices,

$$L = ...\mu^* (\tilde{\bar{u}} Y_u \tilde{u} H_d^{0*} + ...) + ... \tag{34}$$

Thus phenomenologically μ must be of order the weak scale to maintain the solutions of the hierarchy problem, gauge coupling unification, and radiative electroweak symmetry breaking. The naive scale for any term in the superpotential is one above where the supersymmetry is broken, e.g. the string scale or unification scale, and since μ occurs in W one would naively expect μ to be of order that scale, far above the weak scale. In the past that has been called the "μ problem". But actually it is a clue to the correct theory and is an opportunity to learn what form the underlying theory must take. For example, in a string theory we expect all the mass terms to vanish since the SM particles are the massless modes of the theory, so in a string theory μ, which is a mass term, would naturally vanish. That could be a clue that the underlying theory is indeed a string theory. In the following we will view $\mu = 0$ as a "string boundary condition". Older approaches added symmetries to require $\mu = 0$. Note that because of the non-renormalization theorem once μ is set to zero in W it is not generated by loop corrections.

We also know phenomenologically that the μ contribution to the chargino and neutralino masses and the Higgs mass cannot vanish, or some of them would be so light they would have been observed, so we know that somehow a piece that plays the same role as μ is generated. We will call it μ_{eff}, but whenever there is no misunderstanding possible we will drop the subscript and just write μ for μ_{eff}. Different ways of generating μ_{eff} give different relations to the other soft-breaking parameters, a different phase for μ_{eff}, a characteristic size for μ_{eff}, etc. Once it is measured we will have more clues to the underlying theory. Any top-down approach must generate μ_{eff} and its phase correctly.

15. R-PARITY CONSERVATION

The μ opportunity looks like the μ problem if one views supersymmetry as an effective low energy theory without seeing it as embedded in a more fundamental high scale theory. Similarly, if we view supersymmetry as only a low energy effective theory there is another complication that arises. There are additional terms that one could write in W that are analytic, gauge invariant, and Lorentz invariant, but violate baryon and/or lepton number conservation. No such terms are allowed in the SM, which accidentally conserves B and L to all orders in perturbation theory, though it does not conserve them non-perturbatively. These terms are

$$W_R = \lambda_{ijk} L_i L_j e_k + \lambda'_{ijk} L_i Q_j \bar{d}_k + \lambda''_{ijk} \bar{u}_i \bar{d}_j d_k. \tag{35}$$

The couplings $\lambda, \lambda', \lambda''$ are matrices in family space. Combining the second and third one can get very rapid proton decay, so one or both of them must be required to be absent. That is not the way one wants to have a theory behave. Rather, B and L conservation consistent with observation should arise naturally from the symmetries of the theory. Most, but not all, theorists expect that an underlying symmetry will be present in the broader case to forbid all of the terms in W_R.

There are two approaches to dealing with W_R. We can add a symmetry to the effective low energy theory, called R-parity or a variation called matter parity, which we assume will arise from a string theory or extended gauge group. R-parity is multiplicatively conserved,

$$R = (-1)^{3(B-L)+2S} \tag{36}$$

where S is the spin. Then SM particles and Higgs fields are even, superpartners odd. This is a discrete Z_2 symmetry. Such symmetries that treat superpartners differently from SM particles and therefore do not commute with supersymmetry are called R-symmetries. Equivalently, one can use "matter parity",

$$P_m = (-1)^{3(B-L)}. \tag{37}$$

A term in W is only allowed if $P_m = +1$. Gauge fields and Higgs are assigned $P_m = +1$, and quark and lepton supermultiplets $P_m = -1$. P_m commutes with supersymmetry and forbids W_R. Matter parity could be an exact symmetry, and such symmetries do arise in string theory. If R-parity or matter parity holds there are major phenomenological consequences,

- At colliders, or in loops, superpartners are produced in pairs.
- Each superpartner decays into one other superpartner (or an odd number).
- The lightest superpartner (LSP) is stable. That determines supersymmetry collider signatures, and makes the LSP a good candidate for the cold dark matter of the universe.

The second approach is very different, and does not have any of the above phenomenological consequences. One arbitrarily sets λ' or $\lambda'' = 0$ so there are no observable violations of baryon number or lepton number conservation. Other terms are allowed and one sets limits on them when their effects are not observed, term by term. In the MSSM itself R-parity must be broken explicitly if it is broken at all. If it were broken spontaneously by a sneutrino vev there would be a Goldstone boson associated with the spontaneous breaking of lepton number (called a Majaron), and some excluded Z decays would have been observed.

We will not pursue this ad hoc approach, because we do not like arbitrarily setting some terms to zero, and we do not like giving up the LSP as cold dark matter if we are not forced to. Further, large classes of theories conserve R-parity or matter parity.[10] Often theories have a gauged $U(1)_{B-L}$ symmetry that is broken by scalar vevs and leaves P_m automatically conserved. String theories often conserve R-parity or P_m. Often theories conserve R-parity at the minimum of the Higgs potential. Baryogenisis via leptogenesis probably requires R-parity conservation because the usual $B + L$ violation plus L violation would allow the needed asymmetries to be erased. The lepton number needed for ν seesaw masses violates L by two units and does not violate R-parity conservation. In general, when supersymmetry is viewed as embedded in a more fundamental theory, R-parity conservation is very likely and easily justified. Ultimately, of course, experiment will decide, but we will assume R-parity conservation in the rest of these lectures.

16. DEFINITION OF MSSM

At this stage we can define the effective low energy supersymmetry theory, which we call the MSSM, as the theory with the SM gauge group and particles, and the superpartners of the SM particles, and conserved R-parity, and two Higgs doublets. Perhaps it would be better to include right handed neutrinos and their superpartners as well, but that is not yet conventional.

17. THE MSSM SOFT-BREAKING LAGRANGIAN

We can now write the general soft-breaking Lagrangian for the MSSM,

$$
\begin{aligned}
-L_{soft} = {} & \tfrac{1}{2}(M_3\tilde{g}\tilde{g} + M_2\widetilde{W}\widetilde{W} + M_1\widetilde{B}\widetilde{B} + c.c.) \\
& + \widetilde{Q}^\dagger m_Q^2 \widetilde{Q} + \widetilde{\bar{u}}^\dagger m_{\bar{u}}^2 \widetilde{\bar{u}} + \widetilde{\bar{d}}^\dagger m_{\bar{d}}^2 \widetilde{\bar{d}} + \widetilde{L}^\dagger m_L^2 \widetilde{L} + \widetilde{\bar{e}}^\dagger m_{\bar{e}}^2 \widetilde{\bar{e}}
\end{aligned}
$$

$$+(\tilde{\bar{u}}^{\dagger} a_u \tilde{Q} H_u - \tilde{\bar{d}}^{\dagger} a_d \tilde{Q} H_d - \tilde{\bar{e}}^{\dagger} a_e \tilde{L} H_d + c.c.)$$

$$+m_{H_u}^2 H_u^* H_u + m_{H_d}^2 H_d^{2*} + (bH_u H_d + c.c.). \tag{38}$$

For clarity a number of the indices are suppressed. $M_{1,2,3}$ are the complex bino, wino, and gluino masses, e.g. $M_3 = |M_3|\, e^{i\phi_3}$, etc. In the second line m_Q^2, etc, are squark and slepton hermitean 3×3 mass matrices in family space. The $a_{u,d,e}$ are complex 3×3 family matrices, usually called trilinear couplings. b is sometimes written as $B\mu$ or as m_3^2 or as m_{12}^2. Additional parameters come from the gravitino complex mass and from $\mu_{eff} = \mu e^{i\phi_\mu}$; we will usually risk writing the magnitude of μ_{eff} as just μ assuming the context will distinguish this from the original μ of the superpotential. This may seem to involve a lot of parameters, but all the physical parameters are observable from direct production and study of superpartners and their effects. The absence of observation of superpartners and their effects already gives us useful information about some of the parameters. It is important to understand that all of these parameters are masses or flavor rotation angles or phases or Higgs vevs, just as for the SM. If we had no measurements of the quark and lepton masses and interactions there would be even more parameters for the SM than here.

With this Lagrangian we can do general, useful, reliable phenomenology, as we will see. For example, in the SM we did not know the top quark mass until it was measured. Nevertheless, for any chosen value of the top mass we could calculate its production cross section at any collider, all of its decay BR, its contribution to radiative corrections, etc. Similarly, for the superpartners we can calculate expected signals, study any candidate signal and evaluate whether it is consistent with the theory and with other constraints or data, and so on. A possible signal might have too small or large a cross section to be consistent with any set of parameters, or decay BR that could not occur here. Many examples can be given. We can also study whether superpartners can be studied at any proposed future facility. Further, most processes depend on only a few of the parameters — we will see several examples of this in the following.

Now let us count the parameters of the broken supersymmetric theory relative to the SM. There are no new gauge or Yukawa couplings, and still only one strong CP angle $\bar{\vartheta}$, so that is already rather economical. Then

- m_Q^2, etc are 5 3×3 hermitean matrices → 9 real parameters each → 45
- $a_{u,d,e}$ are 3 3×3 complex matrices → 18 real parameters each → 54
- $M_{1,2,3}$, μ, b are complex → 10

• $m^2_{H_{u,d}}$ are real by hermiticity $\to 2$
giving a total of 111 parameters. As for the CKM quark matrix it is possible to redefine some fields and absorb some parameters. Baryon and lepton number are conserved, and there are two U(1) symmetries that one can see by looking at the Lagrangian. One arises because if μ and b are zero there is a symmetry where $H_{u,d} \to e^{i\alpha} H_{u,d}$ and the combinations $L\bar{e}, Q\bar{u}, Q\bar{d} \to e^{-i\alpha} L\bar{e}, Q\bar{u}, Q\bar{d}$. For example, one can take $Q \to e^{-i\alpha} Q$, $L \to e^{-i\alpha} L$, and $\bar{e}, \bar{u}, \bar{d}$ invariant. Such a symmetry is called a Peccei-Quinn symmetry if it holds for $\mu = 0$ but is broken when $\mu \neq 0$. The other arises because if $M_i, a_i, b = 0$ there is a continuous R-symmetry, e.g. the Higgs fields can have charge 2, the other matter fields charge 0, and the superpotential charge 2. Symmetries are called R-symmetries whenever members of a supermultiplet are treated differently.

With these four symmetries, four parameters can be absorbed. Also, the SM has two parameters in the Higgs potential, $\mu^2\phi^2 + \lambda\phi^4$, so to count the number beyond the SM we subtract those 2. Then there are 111-4-2=105 new parameters. The SM itself has 3 gauge couplings, 9 quark and charged lepton masses, 4 CKM angles, 2 Higgs potential parameters, and one strong CP phase $\to$ 19. So there are 124 parameters altogether. When massive neutrinos are included one has RH ν masses, and the angles of the flavor rotation matrix (which has 3 real angles and 3 phases for the ν case since the Majorana nature of the neutrinos prevents absorbing two of the phases). In the following we will discuss how to measure many of the parameters. All are measurable in principle. Once they are measured they can be used to test any theory. In practice, as always historically, some measurements will be needed to formulate the underlying theory (e.g. to learn how supersymmetry is broken and to compactify) and others will then test approaches to doing that.

Only 32 of these parameters are masses of mass eigenstates! There are four neutralinos, two charginos, four Higgs sector masses, three LH sneutrinos, six each of charged sleptons, up squarks, and down squarks, and the gluino. We will examine the connections between the soft masses and the mass eigenstates below. Of the 32 masses, only the gluino occurs directly in L_{soft} — the rest are all related in complicated ways to L_{soft}! One could add the gravitino with its complex mass to the list of parameters. Even the gluino mass gets significant corrections that depend on squark masses.

Some of the ways these parameters contribute is to determining the breaking of the EW symmetry and therefore to the Higgs potential, and the masses and cross sections and decays of Higgs bosons, to the relic density and annihilation and scattering of the LSP, to flavor changing transitions because the rotations that diagonalize the fermion masses will not in general diagonalize the squarks and sleptons, to baryogenesis (which cannot be explained with only the CKM phase), to superpartner masses and signatures at colliders, rare decays with superpartner loops (e.g. $b \to s + \gamma$), electric dipole and magnetic dipole moments, and more.

18. CONNECTING HIGH AND LOW SCALES

Two of the most important successes of supersymmetry depend on connecting the unification and EW scales. We will not study this topic in detail here since Martin covers it thoroughly, but we will look at the aspects we need, particularly for the Higgs sector. The connection is through the logarithmic renormalization and running of masses and couplings, with RGEs. In general we imagine the underlying theory to be formulated at a high energy scale, while we need to connect with experiment at the EW scale. We can imagine running the theory down (top-down) or running an effective Lagrangian determined by data up (bottom-up). It is necessary to calculate for all the parameters of the superpotential and of the soft-breaking Lagrangian. The RGEs are known for gauge couplings and for the superpotential couplings to three loops, and to two loops for other parameters, for the MSSM and its RHν extension. We will only look at one-loop results since we are mainly focusing on pedagogical features. An interesting issue is that calculations must be done with regularization and renormalization procedures that do not break supersymmetry, and that is not straightforward. How to do that is not a solved problem in general, but it is understood through two loops and more loops in particular cases, so in practice there is no problem.

Since our ability to formulate a deeper theory will depend on deducing from data the form of the theory at the unification scale, learning how to convert EW data first into an effective theory at the weak scale, and then into an effective theory at the unification scale, is in a sense the major challenge for particle physics in the coming years. There are of course ambiguities in running to the higher scales. Understanding the uniqueness of the resulting high scale theory, and how to resolve ambiguities as well as possible, is very important.

For the Higgs sector we need to examine the running of several of the soft masses, whose RGEs follow. The quantity t is $\ln(Q/Q_0)$, where Q is the scale and Q_0 a reference scale.

$$16\pi^2 dM_{H_u}^2/dt \approx 3X_t - 6g_2^2\left|M_2\right|^2 - \frac{6}{5}g_1^2\left|M_1\right|^2 \tag{39}$$

$$16\pi^2 dM_{H_d}^2/d \approx 3X_b + X_\tau - 6g_2^2\left|M_2\right|^2 - \frac{6}{5}\left|M_1\right|^2 \tag{40}$$

where

$$X_{t,b} \approx 2\left|Y_{t,b}\right|^2 (M_{H_{u,d}}^2 + m_{Q_3}^2 + m_{\bar{u}_3,\bar{d}_3}^2) + 2\left|a_{t,b}\right|^2 \tag{41}$$

Note that $X_{t,b}$ are positive so $M_{H_{u,d}}^2$ decrease as they evolve toward the EW scale from a high scale, and unless $\tan\beta$ is very large, X_t is larger than X_b. We also need to look at just the leading behavior of the squark running,

$$16\pi^2 dM_{Q_3}^2/dt = X_t + X_b + ... \tag{42}$$

$$16\pi^2 dM_{\bar{u}_3}^2/dt = 2X_t + ... \tag{43}$$

$$16\pi^2 dM_{\bar{d}_3}^2/dt = 2X_b + ... \tag{44}$$

Think back to the SM, where the coefficient (usually called μ^2 there but remember that μ is not the same as our μ) of ϕ^2 in the Higgs potential must be negative to lead to spontaneous symmetry breaking with the minimum of the potential away from the origin. Here $M_{H_u}^2$ plays the role, effectively, of the SM μ^2. We see that because of the large X_t the right hand side of the equation for $M_{H_u}^2$ is indeed the largest, and not only does $M_{H_u}^2$ decrease as it runs but the other quantities run slower so they do not get vevs at the same time. Thus the theory naturally can lead to a derivation of the Higgs mechanism! This is extremely important. The theory could easily have had a form where no Higgs vev formed, or where a Higgs vev could only form if some squark also got a vev, which would violate charge and color conservation. The precise conditions for REWSB are somewhat more subtle in supersymmetry — $M_{H_u}^2$ does not actually need to be negative, just smaller than $M_{H_d}^2$, as we will see next.

19. RADIATIVE ELECTROWEAK SYMMETRY BREAKING (REWSB)

The Higgs sector is the natural domain of supersymmetry. The Higgs mechanism[11] occurs as the scale decreases from the more symmetric high scale, with vacuum expectation values becoming non-zero somewhat above the EW scale. As we will see, the Higgs mechanism is intricately tied up with supersymmetry and with supersymmetry breaking — there is no Higgs mechanism unless supersymmetry is broken. This should be contrasted with the other big issue of flavor physics, the origin of the number of families and the differences between the flavor and mass eigenstates, which is already in the structure of the theory at the unification scale, as discussed above. Supersymmetry accommodates the flavor issues, and allows data to constrain them, but supersymmetry can explain the Higgs physics with string boundary conditions (we'll be more precise about that later).

Once we have the superpotential and L_{soft} we can calculate the scalar potential that determines the Higgs physics — that is very different from the SM case where one adds the scalar potential in by hand. The result is for the electrically neutral fields,

$$V = |\mu_{eff}|^2 \left(|H_u|^2 + |H_d|^2\right) \qquad F \qquad (45)$$

$$+\frac{1}{8}(g_1^2 + g_2^2)(|H_u|^2 - |H_d|^2) \qquad D \qquad (46)$$

$$+m_{H_u}^2 |H_u|^2 + m_{H_d}^2 |H_d|^2 - (bH_uH_d + c.c.). \qquad soft \quad (47)$$

From now on again we will just write μ for μ_{eff}. Now we want to minimize this. If it has a minimum away from the origin vevs will be generated. If we had included the charged scalars we could use gauge invariance to rotate away any vev for (say) H_u^+. Then we would find that the minimization condition $\partial V/\partial H_d^- = 0$ implied that $\langle H_d^- \rangle = 0$, so at the minimum electromagnetism is an unbroken symmetry. The only complex term in V is b. We can redefine the phases of H_u, H_d to absorb the b phase, so we can take b as real and positive. Then by inspection we will have a minimum when the term with b subtracts the most it can, so $\langle H_u\rangle \langle H_d\rangle$ will be real and positive. Since H_u, H_d have hypercharge $\pm\frac{1}{2}$, we can use a hypercharge gauge transformation to take the two vevs separately real and positive. Therefore at the tree level CP is conserved in the Higgs sector and we can choose the mass eigenstates to have definite CP.

Writing $\partial V/\partial H_u = \partial V/\partial H_d = 0$ one finds that the condition for a minimum away from the origin is

$$b^2 > (|\mu|^2 + M_{H_u}^2)(|\mu|^2 + M_{H_d}^2). \tag{48}$$

So $M_{H_u}^2 < 0$ helps to generate EWSB but is not necessary. There is no EWSB if b is too small, or if $|\mu|^2$ is too large. For a valid theory we must also have the potential bounded from below, which was automatic for the unbroken theory but is not when the soft terms are included. The quartic piece in V guarantees V is bounded from below except along the so-called D-flat direction $\langle H_u \rangle = \langle H_d \rangle$, so we need the quadratic terms positive along that direction, which implies

$$2b < 2\,|\mu|^2 + M_{H_u}^2 + M_{H_d}^2. \tag{49}$$

Remarkably, the two conditions cannot be satisfied if $M_{H_u}^2 = M_{H_d}^2$, so the fact that $M_{H_u}^2$ runs more rapidly than $M_{H_d}^2$ is essential. They also cannot be satisfied if $M_{H_u}^2 = M_{H_d}^2 = 0$, i.e. if supersymmetry is unbroken!

We write $\langle H_{u,d} \rangle = \mathrm{v}_{u,d}$. Requiring the Z mass be correct gives

$$\mathrm{v_u^2 + v_d^2 = v^2 = \frac{2M_Z^2}{g_1^2 + g_2^2}} \approx (174\mathrm{GeV})^2 \tag{50}$$

and it is convenient to write

$$\tan\beta = \mathrm{v_u/v_d}. \tag{51}$$

Then $\mathrm{v}_u = \mathrm{v}\sin\beta$, $\mathrm{v}_d = \mathrm{v}\cos\beta$, and with our conventions $0 < \beta < \pi/2$.

With these definitions the minimization conditions can be written

$$|\mu|^2 + M_{H_d}^2 = b\tan\beta - \frac{1}{2}M_Z^2\cos 2\beta \tag{52}$$

$$|\mu|^2 + M_{H_u}^2 = b\cot\beta + \frac{1}{2}M_Z^2\cos 2\beta.$$

These satisfy the EWSB conditions. They can be used (say) to eliminate b and $|\mu|^2$ in terms of $\tan\beta$ and M_Z^2. Note the phase of μ is not determined. These two equations have a special status because they are the only two equations of the entire theory that relate a measured quantity (M_Z^2) to soft parameters. If the soft parameters are too large, these equations would require very precise cancellations to keep the Z mass correct.

We have two Higgs fields, each an SU(2) doublet of complex fields, so 8 real scalars. Three of them are Nambu-Goldstone bosons that are eaten by $W^{\pm}, Z$ to become the longitudinal states of the vector bosons, just as in the SM, so 5 remain as physical particles. They are usually classified as 3 neutral ones, h, H, A, and a charged pair, $H^{\pm}$. The mass matrix is calculated from V with $M_{ij}^2 = \frac{1}{2}\partial^2 V/\partial\phi_i\partial\phi_j$ where $\phi_{i,j}$ run over the 8 real scalars. Then the eigenvalue equation det $\left|\lambda - M_{ij}^2\right| = 0$ determines the mass eigenstates. This splits into block diagonal 2×2 factors. The factors for the charged states and the neutral one in the basis $(ImH_u,\ ImH_d)$ each have one zero eigenvalue, the Nambu-Goldstone bosons. The two CP even neutrals can mix, with mixing matrix

$$\begin{pmatrix} h \\ H \end{pmatrix} = \sqrt{2}\begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix}\begin{pmatrix} ReH_u - \mathrm{v_u} \\ ReH_d - \mathrm{v_d} \end{pmatrix}. \tag{53}$$

The resulting tree level masses are

$$m_{h,H}^2 = \frac{m_A^2 + M_Z^2}{2} \mp \frac{1}{2}\sqrt{(m_A^2 + M_Z^2)^2 - 4m_A^2 M_Z^2 \cos^2 2\beta}, \tag{54}$$

$$m_A^2 = 2b/\sin 2\beta, \tag{55}$$

$$m_{H^\pm}^2 = m_A^2 + M_{W^\pm}^2. \tag{56}$$

From eq. 54, one can see that if $m_A^2 \to 0$ then $m_h^2 \to 0$, and if m_A^2 gets large then $m_h^2 \to 0$, so m_h^2 has a maximum. A little algebra shows the maximum is

$$m_h^{tree} \leq |\cos 2\beta|\, M_Z, \tag{57}$$

where we have emphasized that this maximum does not include radiative corrections. This important result leads to the strongest quantitative test of the existence of supersymmetry, that there must exist a light Higgs boson. If the gauge theory is extended to larger gauge groups there are additional contributions to the tree level mass, but they are bounded too.

There are also significant radiative corrections.[12] The Higgs potential has contributions to the h^4 term from loops involving top quarks and top squarks. These are not small because the top Yukawa coupling is of order unity and the top-Higgs coupling is proportional to the top mass. To include the effect one has to calculate the contribution to the Higgs potential, reminimize, and recalculate the mass matrix eigenvalues. The result is

$$m_h^2 \lesssim \cos^2 2\beta M_Z^2 + \frac{3\alpha_2}{2\pi}\frac{m_t^4}{m_W^2}\ln\frac{\tilde{m}_t^2}{M_Z^2} \approx M_Z^2(1+\frac{1}{4}\ln\frac{\tilde{m}_t^2}{M_Z^2}) \tag{58}$$

where the last equality uses $|\cos 2\beta| = 1$, which is true for $\tan\beta \gtrsim 4$. The contributions from two loops have mainly been calculated and are small but not negligible. This result shows that if $m_h \approx 115$ GeV, it is necessary that the tree level term give essentially the full M_Z contribution, so $\cos^2 2\beta \approx 1$.

If $\tan\beta$ is large the REWSB situation is more complicated. Then the top and bottom Yukawa couplings are approximately equal, so from the RGEs [equations 39-41] we see that $M_{H_u}^2, M_{H_d}^2$ run together, and both can go negative, or the conditions [equations 48,49] may not be satisfied. The EWSB conditions can be rewritten [using equation 55] so one condition is that

$$2m_A^2 \approx M_{H_d}^2 - M_{H_u}^2 - M_Z^2. \tag{59}$$

Experimentally, $m_A^2 \gtrsim M_Z^2$ (or A would have been observed at LEP or the Tevatron), so the EWSB condition is that $M_{H_u}^2$ must be smaller than $M_{H_d}^2$ by an amount somewhat larger than M_Z^2. That allows a narrow window, and preferably the theory would not have to be finely adjusted to allow the REWSB to occur. Also, in this situation the other condition can be written

$$b \approx \frac{M_{H_d}^2 - M_{H_u}^2}{\tan\beta} \sim \frac{M_Z^2}{\tan\beta} \ll M_Z^2 \tag{60}$$

when $\tan\beta$ is large, and this is a clear fine tuning[32] since the natural scale for b is of order the typical soft term, presumably of order or somewhat larger than M_Z^2. So REWSB is possible with large $\tan\beta$ but it is necessary to explain why this apparent fine tuning occurs. The actual effects of increasing $\tan\beta$ are complicated. The b and τ Yukawas get larger, so the top and stop and $m_{H_{u,d}}^2$ RGEs change. $m_{H_{u,d}}^2$ get driven more negative, but the larger Yukawas decrease the stop masses, which makes $m_{H_u}^2$ less negative, etc.

If $\tan\beta$ is large, theories with $M_{H_u}^2$ and $M_{H_d}^2$ split are then favored. That could occur in the unification scale formulation of the theory. One possible way to get a splitting even if $M_{H_u}^2, M_{H_d}^2$ start degenerate is via D-terms from extending the gauge theory.[13] D-terms arise whenever a U(1) symmetry is broken. Under certain circumstances their magnitude may be of order the weak scale even though the U(1) symmetry is broken at a high scale, and they can contribute if the superpartners are charged under that U(1) symmetry. If one looks at SO(10) breaking to SU(5)×U(1) and the breaking of this U(1), the soft masses are

$$\begin{aligned} m_Q^2 = m_{\tilde{e}}^2 = m_{\tilde{u}}^2 &= m_{10}^2 + m_D^2 \\ m_L^2 = m_{\tilde{d}}^2 &= m_5^2 - 3m_D^2 \\ m_{H_{d,u}}^2 &= m_{10}^2 \pm 2m_D^2. \end{aligned}$$

The main point for us is that $m_{H_u}^2$ and $m_{H_d}^2$ are split. The splitting affects the other masses, so in principle m_D^2 is accessible experimentally if sufficiently many scalar masses can be measured.

Note that because b is in L_{soft} it is not protected by a non-renormalization theorem. So to have b small at the weak scale does not mean it is small at the unification scale. It's RGE is

$$16\pi^2 db/dt = b(3Y_t^2 - 3g_2^2 + ...) + \mu(6a_t Y_t + 6g_2^2 M_2 + ...)$$

so if it starts out at zero it is regenerated from the second term, or alternatively cancellations can make it small at the weak scale. Such cancellations would look accidental or fine tuned if one did not know the high scale theory, but the appropriate way to view them would be as a clue to the high scale theory. Similarly, large $\tan\beta$ would presumably mean that one vev is approximately zero at tree level and a small value is generated for it by radiative corrections. No theory is currently known that does that, but if an appropriate symmetry can be found that does it will be a clue to the high scale theory.

Before we leave Higgs physics we will derive one Feynman rule to illustrate how that works. From above we write

$$H_d = \mathrm{v}\cos\beta + \frac{1}{\sqrt{2}}(-\mathrm{h}\sin\alpha + \mathrm{H}\cos\alpha + \mathrm{i}\mathrm{A}\sin\beta) \tag{61}$$

$$H_u = \mathrm{v}\sin\beta + \frac{1}{\sqrt{2}}(\mathrm{h}\cos\alpha + \mathrm{H}\sin\alpha + \mathrm{i}\mathrm{A}\cos\beta).$$

Then from the covariant derivative term there is the Lagrangian contribution

$$\frac{g_2^2}{\cos^2\theta_W}(|H_u|^2 + |H_d|^2)Z^\mu Z_\mu \tag{62}$$

so substituting this gives the hZZ vertex

$$\frac{g_2^2 \mathrm{v}}{2\cos^2\theta_W} Z^\mu Z_\mu h(\sin\beta\cos\alpha - \cos\beta\sin\alpha) = \frac{g_2 M_Z}{\cos\theta_W}\sin(\beta-\alpha) Z^\mu Z_\mu h\,. \tag{63}$$

Similar manipulations give the couplings

$$\begin{array}{cccc} & h & H & A \\ \bar{t}t,\ \bar{c}c,\ \bar{u}u & \cos\alpha/\sin\beta & \sin\alpha/\cos\beta & \cot\beta \\ \bar{b}b,\ \bar{\tau}\tau ... & \text{-}\sin\alpha/\cos\beta & \cos\alpha/\sin\beta & \tan\beta \\ WW, ZZ & \sin(\beta-\alpha) & \cos(\beta-\alpha) & 0 \\ ZA & \cos(\beta-\alpha) & \sin(\beta-\alpha) & 0 \end{array} \tag{64}$$

The ZAh and ZHA vertices are non-zero, while the Zhh and ZHH vertices vanish; there is no tree level $ZW^\pm H^\mp$ vertex.

Finally, we note that in the supersymmetric limit where the soft parameters become zero one has

$$V = |\mu|^2\left(|H_u|^2 + |H_d|^2\right) + \frac{g_1^2 + g_2^2}{2}\left(|H_u|^2 - |H_d|^2\right) \tag{65}$$

So the minimum is at $\mu = 0$, $H_u = H_d$; the latter implies $\tan\beta = 1$.

20. YUKAWA COUPLINGS, tan β , AND THEORETICAL AND EXPERIMENTAL CONSTRAINTS ON tan β

It's important to understand how $\tan\beta$ originates, and what is known about it. At high scales the Higgs fields do not have vevs, so $\tan\beta$ does not exist. The superpotential contains information about the quark and lepton masses through the Yukawa couplings. As the universe cools, at the EW phase transition vevs become non-zero and one can define $\tan\beta = \mathrm{v}_u/\mathrm{v}_d$. Then quark and lepton masses become non-zero, $m_{q,l} = Y_{q,l}\mathrm{v}_{u,d}$.

There are two values for $\tan\beta$ that are in a sense natural. As pointed out just above, the supersymmetric limit corresponds to $\tan\beta = 1$. Typically in string theories some Yukawa couplings are of order gauge couplings, and others of order zero. The large couplings for each family are interpreted as the top, bottom, and tau couplings. If $Y_t \approx Y_b$ then $\tan\beta \sim m_t/m_b$. Numerically this is of order 35, but a number of effects could make it rather larger or smaller, e.g. the values of m_t and m_b change considerably with scale, and with RGE running so $m_t(M_Z)/m_b(M_Z) \sim 50$. Finally $\tan\beta$

is determined at the minimum of the Higgs potential, and can be driven smaller.

There are theoretical limits on $\tan\beta$ arising from the requirement that the theory stay perturbative at high scales (remember, the evidence that the entire theory stays perturbative is both the gauge coupling unification and the radiative EWSB). Requiring that $Y_t = g_2 m_t/\sqrt{2}M_W \sin\beta$ not diverge puts a lower limit on $\sin\beta$ which corresponds to $\tan\beta \gtrsim 1.2$ when done in the complete theory, and similarly $Y_b = g_2 m_b/\sqrt{2}M_W \cos\beta$ leads to $\tan\beta \lesssim 60$. This upper limit is probably reduced by REWSB.

There are no measurements of $\tan\beta$, and as I emphasize below it is not possible to measure $\tan\beta$ at a hadron collider in general. Perhaps we will be lucky and find ourselves in a part of parameter space where such a measurement is possible, or more likely, a combination of information from (say) $g_\mu - 2$ and superpartner masses will lead to at least useful constraints on $\tan\beta$. LEP experimental groups have claimed lower limits on $\tan\beta$ from the absence of superpartner signals, but those are quite model dependent and do not hold if phases are taken into account. Similarly, there is a real lower limit on $\tan\beta$ from the absence of a Higgs boson below 115 GeV, as explained above and in Section 22. That limit is about 4 if phases are not included, but lower when they are.

21. IN WHAT SENSE DOES SUPERSYMMETRY EXPLAIN EWSB?

Understanding the mechanism of EWSB, and its implications, is still the central problem of particle physics. Does supersymmetry indeed explain it? If so, the explanation depends on broken supersymmetry, and we have seen that in the absence of supersymmetry breaking the EW symmetry is not broken. That's OK. An explanation in terms of supersymmetry moves us a step closer to the primary theory. Historically we have learned to go a step at a time, steadily moving toward more basic understanding. If we think of supersymmetry as an effective theory at the weak scale only, then we would expect the sense in which it explains EWSB to be different from that we would find if we think of low energy supersymmetry as the low energy formulation of a high scale theory. That is, top-motivated bottom-up is different from bottom-up. It should be emphasized that one could have supersymmetry breaking without EWSB, but not EWSB without supersymmetry breaking.

It may clarify the issues to first ask what needs explanation. We can explicitly list

(1) Why are there Higgs scalar fields, i.e. scalars that carry SU(2)×U(1) quantum numbers, at all?
(2) Why does the Higgs field get a non-zero vev?
(3) Why is the vev of order the EW scale instead of a high scale?
(4) Why does the Higgs interact differently with different particles, in particularly different fermions?

Let us consider these questions.

At least scalars are naturally present in supersymmetric theories, and generally carry EW quantum numbers, whereas in the SM scalars do not otherwise occur. If we connect to a high scale theory, some (most) explicitly have SM-like Higgs fields, e.g. in the E_6 representation of heterotic string theories. Basically as long as we view supersymmetry as embedded in a high scale theory we will typically have Higgs scalars present, though not in all possible cases. That in turn can point to the correct high scale theory.

We have seen that the RGE running naturally does explain the origin of the Higgs vev if the soft-breaking terms and μ_{eff} are of order the weak scale, and if one Yukawa coupling is of order the gauge couplings. If we view the theory as a low energy effective theory we have seen that we do not know why μ in the superpotential is zero, but if we view the theory as embedded in a string theory then it is natural to have $\mu = 0$ in the superpotential. We referred to this as string boundary conditions. Then how μ_{eff} is generated points toward the correct high scale theory. If μ_{eff} is of order the weak scale then it is appropriate to explain the Higgs mechanism *and* gauge coupling unification. Similarly, the mechanism of supersymmetry breaking has to give soft masses of order the weak scale if supersymmetry explains (or, as some prefer to say, predicts) gauge coupling unification.

In a string theory, for example, we expect some Yukawa couplings to be of order the gauge couplings. We identify one of those with the top quark. Then the running of $M^2_{H_u}$ is fast and it is driven negative, or decreases sufficiently, to imply the non-zero Higgs vev. The relevant soft-breaking terms, particularly $M^2_{H_u}$ and $M^2_{H_d}$ must be of order the weak scale. The theory accommodates different couplings for all the fermions. It does not explain the numerical values of the masses, but allows them to be different — that is non-trivial.

So a complete explanation requires thinking of supersymmetry as embedded in a deeper theory such as string theory (so scalar fields exist in the theory, and $\mu \approx 0$, and the top Yukawa is of order 1), and requires that the soft terms are of order the weak scale after supersymmetry is broken. If we only think of supersymmetry as a low energy effective theory not all of these elements are present, so the explanation is possible but incomplete. It is not circular to impose soft-breaking parameters of order the weak scale to explain the EWSB since one is using supersymmetry breaking to explain EW breaking, which is important progress — that is how physics has increased understanding for centuries.

It is also very important to note that the conditions on the existence of Higgs and on μ and on the soft parameters are equally required for the gauge coupling unification — if they do not hold in a theory then it will not exhibit gauge coupling unification. The explanation of EWSB requires in addition to the conditions for gauge coupling unification only that there is a Yukawa coupling of order the gauge couplings, i.e. a heavy top quark, which is a fact.

Perhaps it is amusing to note that two families are needed to have both a heavy fermion so the EW symmetry is broken, and light fermions that make up the actual world we are part of. No reasons are yet known why a third family is needed — it is clear that CP violation could have arisen from soft phases with two families, and does not require the three family SM.

Now that we have developed some foundations we turn to applications in several areas.

22. CURRENT AND FORTHCOMING HIGGS PHYSICS

There are two important pieces of information about Higgs physics that both independently suggest it will not be too long before a confirmed discovery. But of course it is such an important question that solid data is needed.

The first is the upper limit on m_h from the global analysis of precision LEP (or LEP + SLC +Tevatron) data.[14] Basically the result is that there are a number of independent measurements of SM observables, and every parameter needed to calculate at the observed level of precision is measured except m_h. So one can do a global fit to the data and determine the range of values of m_h for which the fit is acceptable. The result is that at 95% C.L. m_h should be below about 200 GeV. The precise value does not matter for us, and because the data really determines $\ln m_h$ the sensitivity is exponential so it moves around with small changes in input. What is important is that there is an upper limit. The best fit is for a central value of order 100 GeV, but the minimum is fairly broad. The analysis is done for a SM Higgs but is very similar for a supersymmetric Higgs over most of the parameter space.

In physics an upper limit does not always imply there is something below the upper limit. Here the true limit is on a contribution to the amplitude, and maybe it can be faked by other kinds of contributions that mimic it. But such contributions behave differently in other settings, so they can be separated. If one analyzes the possibilities[15] one finds that there is a real upper limit of order 450 GeV on the Higgs mass, if (and only if) additional new physics is present in the TeV region. That new physics or its effects could be detected at LHC and/or a 500 GeV linear electron collider, and/or a higher intensity Z factory ("giga-Z") that accompanies a linear collider. So the upper limit gives us very powerful new information. If no other new physics (besides supersymmetry) occurs and conspires in just the required way with the heavier Higgs state, the upper limit really is about 200 GeV.

The second new information is a possible signal from LEP[16] in its closing weeks for a Higgs boson with m_h=115 GeV. The ALEPH detector was the only group to do a blind analysis, and it is technically a very strong detector, so its observation of about a 3σ signal is important information. It was not possible to run LEP to get enough more data to confirm this signal. Fortunately, its properties are nearly optimal for early confirmation at the Tevatron, since its mass is predicted, and cross section and branching ratio to $b\bar{b}$ are large. Less is required to confirm a signal in a predicted mass bin than to find a signal of unknown mass, so less than 10 fb^{-1} of integrated luminosity will be required if the LEP signal is indeed correct. If funding and the collider and the detectors all work as planned confirming evidence for h could come in 2004.

Suppose the LEP h is indeed real. What have we learned?[17] Most importantly, of course, that a point-like, fundamental Higgs boson exists. It is point-like because its production cross section is not suppressed by structure effects. It is a new kind of matter, different from the century old matter particles and gauge bosons. It completes the SM, and points to how to extend the SM. It confirms the Higgs mechanism, since it is produced with the non-gauge-invariant ZZh vertex, which must originate in the gauge-invariant $ZZhh$ vertex with one h having a vev.

The mass of 115 GeV also tells us important information. First, the Higgs boson is not a purely SM one, since the potential energy would be unbounded from below at that mass. Basically the argument is that one has to write the potential with quantum corrections, and the corrections from fermion loops dominate because of the heavy top and can be negative if m_h is too small. The SM potential is

$$V(h) = -\mu^2 h^2 + \left\{ \lambda + \frac{3M_Z^4 + 6M_W^4 + m_h^4 - 12m_t^4}{64\pi^2 \mathrm{v}^4} \ln(\ \ ) \right\} h^4, \qquad (66)$$

where the argument of the ln is some function of the masses larger than one. In the usual way $\lambda = m_h^2/2\mathrm{v}^2$. The second term in the brackets is negative, so λ and therefore m_h has to be large enough. The full argument has to include higher loops, thermal corrections, a metastable universe rather than a totally stable one, etc., and requires m_h to be larger than about 125 GeV if h can be a purely SM Higgs boson.

Second, 115 GeV is an entirely reasonable value of m_h for supersymmetry, but only if $\tan\beta$ is constrained to be larger than about 4. That is because as described above, the tree level contribution is proportional to $|\cos 2\beta|$ and to get a result as large as 115 it is necessary that $|\cos 2\beta|$ be essentially unity, giving a lower limit on $\tan\beta$ of about 4 . Even then the tree level can only contribute a maximum of M_Z to m_h. The rest comes from the radiative corrections, mainly the top loop. Numerically one gets

$$m_h^2 \approx (91)^2 + (40)^2 \left\{ \ln \frac{m_{\tilde{t}}^2}{m_t^2} + ... \right\} \qquad (67)$$

where $m_{\tilde{t}}^2$ is an appropriate average of the two stop mass eigenstates. The second term must supply about 25 GeV, which is quite reasonable.

The LEP signal, assuming it is correct, can only provide us a limited amount of information since it only supplies two numbers, m_h and $\sigma \times BR$. The full Higgs potential depends on at least 7 parameters,[18] so none of them can be explicitly measured. Because the potential depends on the stop loops, it depends on the hermitean stop mass matrix (equation 69 below).

Since the elements are complex, in general the loop contributions to the Higgs potential will be complex, so the potential will have to be re-minimized taking into account the possibility of a relative phase between the Higgs vevs. One can write

$$H_d = \frac{1}{\sqrt{2}} \begin{pmatrix} \mathrm{v_d + h_d + ia_d} \\ h_d^- \end{pmatrix}, \qquad H_u = \frac{e^{i\theta}}{\sqrt{2}} \begin{pmatrix} h_u^+ \\ \mathrm{v_u + h_u + ia_u} \end{pmatrix}. \qquad (68)$$

At the minimum of the potential it turns out that θ cannot be set to zero or absorbed by redefinitions. The resulting θ is a function of the phase of μ, ϕ_μ, and of the phase(s) in a_t (and of course of other parameters). Thus

m_h and $\sigma_h \times BR(b\bar{b})$ are functions of the magnitudes of μ and a_t, m_Q^2, m_u^2, b, $\tan\beta$, and the physical phase(s) $\phi_\mu + \phi_{a_t}$ at least. Since some of these are matrices they can involve more than one parameter. Also, if $\tan\beta$ is large there will be important sbottom loops, and chargino and neutralino loops can contribute. So only in special cases can data about the Higgs sector be inverted to measure $\tan\beta$ and the soft parameters, and only then if there are at least 7 observables.

If θ is significant then even and odd CP states mix and there are 3 mixed neutral states which could all show up in the $b\bar{b}$ or $\gamma\gamma$ spectrum, and those spectra could show different amounts of the three mass eigenstates. Both cross section and BR for the lightest state can be different from the SM and from the CP conserving supersymmetry case.

One can check that the phase can be very important. For example, if a Higgs is observed at LEP and the Tevatron one can ask what region of parameter space is consistent with a given mass and $\sigma_h \times BR(b\bar{b})$. The answer is significantly different, for example for $\tan\beta$, if the phase is included. Or if no Higgs is observed one can ask what region of parameters is excluded. If the phase is included the actual limit on m_h is about 10% lower than the published limits from LEP, below 100 GeV. Similarly, lower values of $\tan\beta$ are allowed if phases are included than those reported by LEP experimenters.

23. THE STOP MASS MATRIX

Arranging the stop mass terms from the Lagrangian in the form

$$\begin{pmatrix} \tilde{t}_L^* & \tilde{t}_R^* \end{pmatrix} m_{\tilde{t}}^2 \begin{pmatrix} \tilde{t}_L \\ \tilde{t}_R \end{pmatrix},$$

the resulting Hermitean stop mass matrix is

$$m_{\tilde{t}}^2 = \begin{pmatrix} m_{Q_3}^2 + m_t^2 + \Delta_u & \mathrm{v}(a_t \sin\beta - \mu Y_t \cos\beta) \\ & m_{\bar{u}_3}^2 + m_t^2 + \Delta_{\bar{u}} \end{pmatrix}. \tag{69}$$

The $\Delta's$ are D-terms, from the $(\phi^* T \phi)^2$ piece of the Lagrangian — $\Delta_u = (\frac{1}{2} - \frac{2}{3}\sin^2\theta_W)\cos 2\beta M_Z^2$, $\Delta_{\bar{u}} = \frac{2}{3}\sin^2\theta_W \cos 2\beta M_Z^2$. These EW D-terms are proportional to the T_3 and hypercharge charges. The pieces proportional to $\sin^2\theta_W$ come from the breaking of the U(1) symmetry and vanish if $\sin^2\theta_W \to 0$. The m_t^2 comes from the F-terms in the scalar potential, $Y_t^2 H_u^{0*} H_u^0 \tilde{t}_L^* \tilde{t}_L$ and a similar term for $\tilde{t}_R$, when the Higgs get vevs.

F-terms in V also give the term -$\mu Y_t \tilde{t}^* \tilde{t} H_d^{0*}$ which gives the second term in the 12 position when H_d^0 gets a vev. The soft term $a_u \tilde{t}^* \tilde{Q}_3 H_u^0$ gives the first 12 term when the Higgs gets a vev. Similar mass matrices are written for all the squarks and sleptons. For the lighter ones the Yukawas and possibly the trilinears are small, and the fermion masses are small, so only the diagonal elements are probably large. Each of the elements above is a 3×3 matrix, so $m_{\tilde{t}}^2$ is a 6×6 matrix. a_t and μ and even v are in general complex.

24. WHAT CAN BE MEASURED IN THE HIGGS SECTOR?

Assuming the LEP signal is indeed valid, as suggested particularly by the ALEPH blind experiment, and it is confirmed at the Tevatron, what can we eventually learn? I will focus on the Tevatron and LHC since they will be our only direct sources of Higgs information in the next decade. The Tevatron can use the WWh, ZZH channels. In addition once m_h is known the inclusive channel, with about a pb cross section, can be used. If the total cross section at the Tevatron for Higgs production is 1.5 pb, and each detector gets 15 fb^{-1} of integrated luminosity, the total number of Higgs bosons produced is about 45,000 in a known mass bin. At some level it will be possible to measure $g_{WWh} g_{b\bar{b}h}$ and $g_{ZZh} g_{b\bar{b}h}$ from σxBR for the WWh and ZZh channels, so their ratio tests whether h couples to gauge bosons proportional to mass. Once m_h is known it will be possible to see $h \rightarrow \tau\bar{\tau}$ in both inclusive production and associated production with a W, and test if the coupling to fermions is proportional to mass. A similar test comes from not seeing $h \rightarrow \mu\bar{\mu}$ (or seeing a few events of this mode since it should occur a bit below the 10^{-3} level). The inclusive production is dominantly via a top loop so it measures g_{tth} indirectly, and this is complicated since superpartner loops contribute as well as SM ones. It may be possible to see the $t\bar{t}h$ final state directly.[19] Since $BR(\gamma\gamma)$ is at the 10^{-3} level an observation or useful limit will be possible here if the resolution is good enough. All of these can give very important tests of what the Higgs sector is telling us.

It is also interesting to ask if data can distinguish a SM Higgs from a supersymmetric one, though most likely there will be signals of superpartners as well as a Higgs signal so there will not be any doubt. If $\tan\beta$ is large the ratio of $b\bar{b}$ to $\tau\bar{\tau}$ is sensitive to supersymmetric-QCD effects and can vary considerably from its tree level value.[20] The ratio of top to bottom couplings is sensitive to ways in which the supersymmetric Higgs sector varies from the SM one. If $\tan\beta$ is large and m_A is less than about 150 GeV it is possible A can be observed at the Tevatron. Altogether, the Tevatron may be a powerful Higgs factory if it takes full advantage of its opportunities. It is still unlikely that there will be enough independent measurements at the Tevatron to invert the equations relating the soft parameters and $\tan\beta$ to observables. The lighter stop mass eigenstate $\tilde{t}_1$ may be observable at the Tevatron, and provide another observable for the Higgs sector.

At LHC it is very hard to learn much about the lightest Higgs h if its mass is of order 115 GeV. It will most likely be observed in the inclusive production and decay to $\gamma\gamma$, but observation in the $\gamma\gamma$ mode does not tell us much about the Higgs physics once the Higgs boson has been discovered, which will have occurred if indeed $m_h \approx 115$ GeV. The $\gamma\gamma$ mode does not demonstrate the Higgs mechanism is operating since it occurs for any scalar boson. The SM does have a definite prediction for BR($\gamma\gamma$) from the top and W loops, and superpartner loops can be comparable, so a measurement would be very interesting. Note that one cannot assume the $\gamma\gamma$ BR is known.

Maybe it will be possible to detect the $\tau\bar{\tau}$ mode at LHC using WW fusion to produce h and tagging the quarks.[21] This mode also confirms the non-gauge-invariant WWh vertex. Considerable additional information about the Higgs sector may come from observing the heavier Higgs masses and $\sigma\times BR$, and the heavier stop. Since $A \to \gamma\gamma$ but not to ZZ, WW it may be possible to see A if it is not above the $t\bar{t}$ threshold. Decays of the heavy Higgs to $\tau's$ are enhanced if $\tan\beta$ is large. Note that one cannot assume only SM decays of h in analysis since channels such as h$\to LSP + LSP$ are potentially open and can have large BR since they are not suppressed by factors such as m_b^2/M_W^2. The combined data from the Tevatron and LHC may provide enough observables to invert the Higgs sector, at least under certain reasonable and checkable assumptions.

25. NEUTRALINOS AND CHARGINOS

The lightest superpartners are likely to be the neutralinos and charginos, possibly the lighter stop, and the gluino. Their mass matrices have entries from the higgsino-gaugino mixing once the SU(2)×U(1) symmetry is broken, so the mass eigenstates are mixtures of the symmetry eigenstates. When phases are neglected these matrices are described in detail in many places so I will not repeat that here. However, it is worth looking at the most general case including phases for several instructive reasons. The chargino mass matrix follows from the L_{soft}, in the wino-higgsino basis:

$$M_{\tilde{C}} = \begin{pmatrix} M_2 e^{i\phi_2} & \sqrt{2} M_W \sin\beta \\ \sqrt{2} M_W \cos\beta & \mu e^{i\phi_\mu} \end{pmatrix}. \tag{70}$$

The situation is actually more subtle — this is a submatrix of the actual chargino mass matrix, but this contains all the information — and the reader should see Martin or earlier reviews for details. Also, the off-diagonal element can be complex too since it arises from the last term in eq.14 when the Higgs gets a vev, and the vev can be complex as explained above; I will just keep the phases of M_2 and μ here. The masses of the mass eigenstates are the eigenvalues of this matrix. To diagonalize it one forms the hermitean matrix $M^\dagger M$. The easiest way to see the main points are to write the sums and products of the mass eigenstates,

$$M^2_{\tilde{C}_1} + M^2_{\tilde{C}_2} = Tr M^\dagger_{\tilde{C}} M_{\tilde{C}} = M_2^2 + \mu^2 + 2M_W^2, \tag{71}$$

$$M^2_{\tilde{C}_1} M^2_{\tilde{C}_2} = \det M^\dagger_{\tilde{C}} M_{\tilde{C}} \tag{72}$$

$$= M_2^2 \mu^2 + 2M_W^4 \sin^2 2\beta - 2M_W^2 M_2 \mu \sin 2\beta \cos(\phi_2 + \phi_\mu)$$

Experiments measure the masses of the mass eigenstates. One thing to note is that the masses depend on the phases ϕ_2 and ϕ_μ, even though there is no CP violation associated with the masses. Often it is implicitly assumed that phases can only be measured by observing CP-violating effects, but we see that is not so. The combination $\phi_2 + \phi_\mu$ is a physical phase, invariant under any reparameterization of phases, as much a basic parameter as $\tan\beta$ or any soft mass.

If one wants to measure the soft masses, μ, $\tan\beta$, $\phi_2 + \phi_\mu$ it is necessary to invert such equations. Since there are fewer observables than parameters to measure, additional observables are needed. One can measure the production cross sections of the mass eigenstates. But then additional parameters enter since exchanges of sneutrinos (at an electron collider) or squarks (at a hadron collider) contribute. One can decide to neglect the additional contributions, but then one is not really doing a measurement. If one "measures" $\tan\beta$ from the above equations by setting the phase to zero, as is usually done, the result is different from that which would be obtained if the phase were not zero. When the phases are present the phenomenology, and any deduced results, can be quite different. We saw that for the Higgs sector above. It is studied for the chargino sector in Ref. 22. Similar arguments apply for the neutralino mass matrix.

One implication of this analysis is that $\tan\beta$ is not in general measurable at a hadron collider — there are simply not enough observables.[23] One can count them, and the equations never converge. Depending on what can be measured, by combining observables from the chargino and neutralino sectors, and the Higgs sector, it may be possible to invert the equations. This is a very strong argument[23] for a lepton collider with a polarized beam, where enough observables do exist if one is above the threshold for lighter charginos and neutralinos, because measurements with different beam polarizations (not possible at a hadron collider) double the number of observables, and measurements with different beam energies (not possible at a hadron collider) double them again. The precise counting has to be done carefully, and quadratic (and other) ambiguities and experimental errors mean that one must do a thorough simulation[24] to be sure of what is needed, but there appear to be sufficient observables to measure the relevant parameters. The issue of observing the fundamental parameters of L_{soft} is of course broader, as discussed in Section 17. There are 33 masses in the MSSM including the gravitino, but 107 new parameters in L_{soft} (including the gravitino). The rest are flavor rotation angles and phases. Many can be measured by combining data from a linear electron collider above the threshold for a few superpartners and hadron colliders. It is also necessary to include flavor changing rare decays to measure the off-diagonal elements of the sfermion mass matrices and the trilinear couplings.

26. NEUTRALINOS

In a basis $\Psi^0 = (\widetilde{B}, \widetilde{W}_3, \widetilde{H}_d, \widetilde{H}_u)$ terms in the Lagrangian can be rearranged into $-\frac{1}{2}(\Psi^0)^T M_{\tilde{N}} \Psi^0$ with the symmetric

$$M_{\tilde{N}} = \begin{pmatrix} M_1 e^{i\phi_1} & 0 & -\frac{g_1}{\sqrt{2}} H_d^{0*} & \frac{g_1}{\sqrt{2}} H_u^{0*} \\ & M_2 e^{i\phi_2} & \frac{g_2}{\sqrt{2}} H_d^{0*} & -\frac{g_2}{\sqrt{2}} H_u^{0*} \\ & & 0 & -\mu e^{i\phi_\mu} \\ & & & 0 \end{pmatrix}.$$

Although the elements are complex, this matrix can still be diagonalized by a unitary transformation. Its form in a basis $\Psi' = (\tilde{\gamma}, \widetilde{Z}, \tilde{h}_s, \tilde{h}_a)$ is sometimes useful:

$$M_{\tilde{N}} = \begin{pmatrix} M_1 s_W^2 + M_2 c_W^2 & (M_1 - M_2) s_W c_W & 0 & 0 \\ & M_1 s_W^2 + M_2 c_W^2 & M_Z & 0 \\ & & \mu \sin 2\beta & -\mu \cos 2\beta \\ & & & -\mu \sin 2\beta \end{pmatrix}.$$

If $M_1 \approx M_2$ and/or if $\tan\beta$ is large (so $\sin 2\beta \approx 0$) this takes a simple form.

The lightest neutralino is the lightest eigenvalue of this, and may be the LSP. Its properties then determine the relic density of cold dark matter (if the LSP is indeed the lightest neutralino). It also largely determines the collider signatures for supersymmetry. It will be a linear combination of the basis states,

$$\widetilde{N}_1 = \alpha \widetilde{B} + \beta \widetilde{W}_3 + \gamma \widetilde{H}_d + \delta \widetilde{H}_u$$

with $|\alpha|^2 + |\beta|^2 + |\gamma|^2 + |\delta|^2 = 1$.

An interesting limit that is at least pedagogically instructive arises if we take $M_1 \approx M_2$ (at the EW scale) and $\tan\beta \approx 1$, and $\mu < M_Z$. Then $\widetilde{N}_1 \approx \tilde{h}$, where $\tilde{h} = \tilde{h}_d \sin\beta + \tilde{h}_u \cos\beta$, and $M_{\widetilde{N}_1} \approx \mu$. $\widetilde{N}_2 \approx \tilde{\gamma}$, with $M_{\widetilde{N}_2} \approx M_2$, $\alpha \approx -\beta \approx -45°$ so $\cos(\alpha - \beta)$ and $\cos 2\beta \to 0$. At tree level the $Z\tilde{\gamma}\tilde{h}$, $h\tilde{\gamma}\tilde{h}$, and $Z\tilde{h}\tilde{h}$ vertices vanish, and the dominant decay of the second neutralino is $\widetilde{N}_2 \to \widetilde{N}_1 + \gamma$. $M_{\tilde{C}_1} \gtrsim M_{\widetilde{N}_2}$.

27. EFFECTS OF PHASES

The effects of phases have been considered much less than the masses. As we saw above for charginos and the Higgs sector they affect not only CP-violating observables but essentially all observables. They can have significant impacts in a variety of places, including $g_\mu - 2$, electric dipole moments (EDMs), CP violation in the K and B systems, the baryon asymmetry of the universe, cold dark matter, superpartner production cross sections and branching ratios, and rare decays. We do not have space to give a complete treatment, but only to make some points about the importance of the observations and what they might teach us about physics beyond the SM in general; while we focus to some extent on the phases because they are usually not discussed, our concern is relating them to the entire L_{soft}.

There are some experiments that suggest some of the phases are small, mainly the neutron and electron EDMs. On the other hand, we know that the baryon asymmetry cannot be explained by the quark CKM phase, so some other phase(s) are large, and the soft phases are good candidates.

Recently it has been argued that very large phases are needed if baryogenesis occurs at the EW phase transition;[25] see also Ref. 26. Further, there is no known symmetry or basic argument that the soft phases in general should be small. If the outcome of studying how to measure them was to demonstrate that some were large that could be very important because both compactification and supersymmetry breaking would have to give such large phases. The phase structure of the effective soft Lagrangian at the weak scale and at the unification scale are rather closely related, so it may be easier to deduce information about the high scale phases from data than about high scale parameters in general. If the outcome of studying how to measure the phases was to demonstrate that the phases were small that would tell us different but very important results about the high scale theory. It would also greatly simplify analyzing weak scale physics, but that is not sufficient reason to assume the phases are small.

28. $g_\mu - 2$

In early 2001 it was reported that the anomalous magnetic moment of the muon was larger than the SM prediction by a significant amount. The experiment is now analyzing several times more data than the original report was based on, and the SM theory is being reexamined.

Even if the effect disappears, it is worth considering $g-2$ experiments, because in a supersymmetric world the entire anomalous moment of any fermion vanishes if the supersymmetry is unbroken, so magnetic moments are expected to be very sensitive to the presence of low energy supersymmetry, and particularly of broken supersymmetry. The analysis can be done in a very general and model independent manner,[27] and illustrates nicely how one can say a great deal with supersymmetry even though it seems to have a number of parameters. So it is also pedagogically interesting. There are only two supersymmetric contributions, a chargino-sneutrino loop and a smuon-neutralino loop. One can see that starting from the complete theory, with no assumptions beyond working in the MSSM, there are only 11 parameters that can play a role out of the original set of over 100,

$$|M_2|\,, |M_1|\,, |\mu|\,, |A_\mu|\,, m_{\tilde{\mu}_L}, m_{\tilde{\mu}_R}, m_{\tilde{\nu}}, \tan\beta, \phi_2+\phi_\mu, \phi_1+\phi_\mu, \phi_A+\phi_\mu.$$

In the general case all 11 of them can be important, and the experimental result will give a complicated constraint among them. But if we ask about putting an upper limit on superpartner masses, which would be of great interest, we can say more. For larger masses one can see that the chargino-sneutrino diagram dominates, and in addition that it is proportional to $\tan\beta$; The $\tan\beta$ factor arises from the needed chirality flip on a chargino line. Thus only the magnitudes of M_2, and μ, $\tan\beta$, $m_{\tilde{\nu}}$ and the phase enter in this limit. If we illustrate the result by assuming a common superpartner mass $\tilde{m}$ (just for pedagogical reasons, not in the actual calculations), we find that

$$a_\mu^{susy}/a_\mu^{SM} \approx \left(\frac{100\ GeV}{\tilde{m}}\right)^2 \tan\beta\cos(\phi_2+\phi_\mu). \qquad (73)$$

Further, to put upper limits on the masses we can take the phase to be zero since it turns out to enter only in the above form under these assumptions (for the general case see Ref. 28). And if we express results in terms of the lighter chargino mass rather than M_2 and μ we can eliminate one parameter; for a given chargino mass there will be ranges of M_2 and μ. So we are down to three parameters, with no uncontrolled approximations or assumptions. We will not focus on details of the data here since the new data in 2002 will in any case require a new analysis. If the effect persists there will be significant upper limits on the superpartner masses. Note the relevant physical phase here is $\phi_2+\phi_\mu$.

It is interesting to consider the supersymmetry limit so the supersymmetric SM contribution vanishes. In that limit the two lighter neutralino masses vanish, and their contribution cancels the photon contribution, the two heavier neutralino masses become M_Z and their contribution cancels that of the Z, and the two charginos have M_W and cancel the W contribution. Since the chargino has a sign opposite to that of the W in the supersymmetric limit but the same sign for the broken supersymmetry physical situation it is important to check that indeed the piece proportional to $\tan\beta$ does change sign as needed.

29. ELECTRIC DIPOLE MOMENTS

In the SM electric dipole moments are unobservably small, of order $10^{-33}e$ cm. That is basically because they are intrinsically CP-violating quantities, and for CP violation to occur in the SM it is necessary for all three families to affect the quantity in question. Otherwise one could rotate the CKM matrix in such a way that the phase did not occur in the elements that contributed. So it must be at least a two-loop suppression. There must also be a factor of the electron or quark mass because of a chirality flip, with the scale being of order M_W. In addition there is a GIM suppression.

Interpreting results will be complicated because the neutron, and any nuclear EDM, can have a contribution from strong CP violation, while the electron can only feel effects from EW interactions.

Naively, the EDM is the imaginary part of a magnetic moment operator, and the real part is the magnetic moment. So EDMs can arise from the same diagrams as $g_\mu - 2$, but for the electron and for quarks (in neutrons). It is more complicated in reality because the part of the amplitude that has an imaginary part may not give the dominant contribution to the magnetic moment. It has been known for a long time that if the soft phases were of order unity and if all contributions were independent, then the supersymmetry contributions to EDMs are too large by a factor of order 50. However, over a significant part of parameter space various contributions can cancel. Some of that cancellation is generic, e.g. between chargino and neutralino in the electron EDM because of the relative minus sign in eq.32. The smallness of EDMs may be telling us that the soft phases are small. Then we need to find out why they are small. Or it may be telling us that cancellations do occur. Cancellations look fine tuned from the point of view of the low energy theory, but small phases look fine tuned too. Relations among soft parameters in the high scale theory will look fine-tuned in the low scale theory if we do not know the origin of those relations. If $\tan\beta$ is very large cancellations become unlikely since the chargino contribution will dominate the eEDM just as it does for $g_\mu - 2$, but if $\tan\beta$ is of order 4-5 the situation has to be studied carefully.

30. MEASURING PHASES AT HADRON COLLIDERS

Phases, as well as soft masses, can affect distributions at colliders. We briefly illustrate that here for an oversimplified model.[29] Consider gluino production at a hadron collider. The Lagrangian contains a term

$$M_3 e^{i\phi_3} \lambda_{\tilde{g}} \lambda_{\tilde{g}} + c.c. \tag{74}$$

It is convenient to redefine the fields so the phase is shifted from the masses to the vertex, so one can write $\psi_{\tilde{g}} = e^{i\phi_3/2}\lambda_{\tilde{g}}$. Then writing the Lagrangian in terms of ψ the vertices $q\tilde{q}\tilde{g}$ get factors $e^{\pm i\phi_3/2}$. The production cross sections for gluinos, for example from $q + \bar{q} \to \tilde{g} + \tilde{g}$ by squark exchange, have factors $e^{+i\phi_3/2}$ at one vertex and $e^{-i\phi_3/2}$ at the other, so they do not depend on the phase. That is clear from general principles since ϕ_3 is not by itself a physical, reparameterization-invariant phase. But gluinos always decay, and for example in the decay $\tilde{g} \to q + \bar{q} + \tilde{B}$ mediated by squarks there is a factor of $e^{i\phi_3/2}$ at the $q\tilde{q}\tilde{g}$ vertex and a factor $e^{i\phi_1/2}$ at the $q\tilde{q}\tilde{B}$ vertex, so the rate depends on the physical relative phase $\phi_3 - \phi_1$. In general it is more complicated with all the relative phases of the neutralino mass matrix entering. In this simple example the experimental distribution in Bino energy is

$$d\sigma/dx \sim m_{\tilde{g}}^4 \left(\frac{1}{\tilde{m}_L^4} + \frac{1}{\tilde{m}_R^4}\right) \times \tag{75}$$

$$[x - 4x^2/3 - 2y^2/3 + y(1 - 2x + y^2)\cos(\phi_3 - \phi_1)]$$

where $x = E_{\tilde{B}} m_{\tilde{g}}$ and $y = m_{\tilde{B}} m_{\tilde{g}}$. Other distributions are also affected. If one tries to obtain information from gluino decay distributions without taking phases into account the answers will be misleading if the phases are not small. The same result is of course true for many superpartner decays. It is important to realize that the same phases are appearing here as appear for example in studying ε and ε' in the kaon system or in $b \to s + \gamma$.

31. LSP COLD DARK MATTER

If it is stable, the LSP is a good candidate for the cold dark matter of the universe. Historically, it is worth noting that this was noticed before we knew that non-baryonic dark matter was needed to understand large scale structure. It is a prediction of supersymmetry. We discussed above why we expected R-parity or a similar symmetry to hold, with the stability of the LSP as one of its consequences. Then the basic argument is simple. As the universe cools, soon after the EW phase transition all particles have decayed except photons, $e^{\pm}, u^{\pm}, d^{\pm}$, neutrinos, and LSPs. The quarks form baryons, which join with electrons to make atoms. The relic density of all but LSPs is known to be $\Omega_{SM} < 0.05$, while $\Omega_{matter} \approx 1/3$. The LSPs annihilate as the universe cools, with a typical annihilation cross section $\sigma_{ann} \sim \rho_{LSP} G_F^2 E^2$, and in the early universe $E \sim T$. The expansion rate is governed by the Hubble parameter $H \sim T^2/M_{Pl}$. The LSPs freeze out and stop annihilating when their mean free collision path is of order the horizon, so $\sigma_{ann} \sim H$. This gives a density $\rho_{LSP} \sim 1/M_{Pl} G_F^2 \sim 10^{-9}$ GeV^3. At freeze-out $T \sim 1$ GeV, and ρ_γ is of order the entropy $S \sim T^3 \sim 1$ GeV^3, so $\rho_{LSP}/\rho_\gamma \sim 10^{-9}$, similar to the density of baryons. Thus $\Omega_{LSP} \sim (M_{LSP}/M_{proton})\Omega_{baryon}$. Quantitative calculations in many models confirm this.

But the actual calculations of the relic density depend on several soft parameters such as masses of sleptons and gauginos, and also on $\tan\beta$

and on soft phases. In the absence of measurements or a theory that can convincingly determine all of these, we cannot in fact say more than that qualitatively the LSP is a good candidate, even if WIMPs are apparently discovered. Since we have argued above that in practice it is unlikely that $\tan\beta$ will be measured accurately at hadron colliders (though we may be lucky with $g_\mu - 2$ plus hadron colliders), it may be difficult to compute Ω_{matter} accurately even after LSPs are detected. It should be emphasized that detection of LSPs is not sufficient to argue they are actually providing the cold dark matter[30] — LSPs could be detected in direct experiments scattering off nuclei, and in space based searches, and at colliders even if $\Omega_{LSP} \lesssim 0.05$. Alternatively, they could make up the CDM even though they were not detected in direct and space based experiments.

Further, in recent years it has come to be understood that LSPs may be produced dominantly by processes that are not in thermal equilibrium rather than the equilibrium process described above. In that case the relic density is not so simply connected to the LSP nature.

32. COMMENTS ON RELATING CP VIOLATION AND STRING THEORY; COULD THE CKM PHASE BE SMALL?

Where does CP violation originate? Can the pattern of CP phenomena give us important clues to formulating and testing string theory? Very little work has been done about the fundamental origins of CP violation. In 1985 Strominger and Witten discussed how to define CP transformations in string theory, and in 1993 Dine, Leigh, and McIntyre argued that CP was a gauge symmetry in string theory, for both strong and EW CP violation. As a gauge symmetry it could not be broken explicitly, perturbatively. or non-perturbatively. More recently Bailin et al, Dent, Geidt, and Lebedev have discussed aspects of this question. Little thought has been given to CP violation in D-brane worlds, Type IIB theories with SM particles as Type I open strings, and so on.

From the point of view of connecting to the observable world, however the CP violation originates it will appear as phases in either the Yukawa couplings in the superpotential, or as phases in L_{soft}. Any theory for CP violation will produce characteristic patterns of such phases. So if we could measure those phases perhaps we would have rather direct information about such questions as moduli dependence of Yukawas, supersymmetry breaking and transmission, vevs of moduli and the dilaton, and the compactification manifold.

If one begins with a string theory including proposed solutions to how to compactify, and to break supersymmetry, the connection to the observable world is first made by writing down the Kahler potential, gauge kinetic function, and superpotential, $W = Y_{\alpha\beta\gamma}\phi_\alpha\phi_\beta\phi_\gamma$. Then L_{soft} is calculated for the assumed approach to supersymmetry breaking, etc. The trilinear terms, for example, are linear combinations of the Yukawas and derivatives of the Yukawas with respect to moduli fields. So if the Yukawas have large phases it seems likely the trilinear terms will also have large phases. On the other hand, phases could enter the trilinears through the Kahler potential even if they were not present in the Yukawas. Recalling that the quark CKM phase is unable to provide the CP violation needed for the baryon asymmetry, it is interesting to consider the possibility that all CP violation originates in the soft phases. It is possible to describe CP violation in the kaon and B systems with only soft phases.[31]

Phenomenologically there are a number of ways that soft phases could be shown to be large. One is observing an eEDM. The nEDM is not so simple to interpret since it could arise from strong CP violation,.but perhaps the relative size of the nEDM and HgEDM could show the effect of soft phases. The Higgs sector could show phase effects, as could superpartner masses, production cross sections, and decay BR. The size of $K_L \to \pi^0 \nu\bar{\nu}$ could deviate from the SM prediction. It is much harder to demonstrate that $\delta_{CKM} \neq 0$.

33. PHASES (AND FLAVOR STRUCTURE) OF L_{soft}

The soft-breaking Lagrangian has, as we have seen, many phases, and interesting and potentially important flavor structure. Few top-down models, e.g. string based models, have studied or even looked at the phase and flavor structure. There is and will be much more data on these topics, and there should be much more theoretical analysis of them. We have looked a little at string motivated models that give interesting phase structure. There is some work on this by Bailin et al for the heterotic string. Following the framework of Ibanez, Munoz, and Rigolin,[39] we have looked at how the phases emerge in some D-brane models.[33]

If one embeds the MSSM on one brane, usually the gaugino masses M_i all have the same phase, and using the freedom from a U(1) symmetry as one can rotate that phase away. An interesting structure emerges if one embeds the SM gauge groups on two intersecting branes. We studied the simplest case with SU(2) on one brane, and SU(3)×U(1) on the other. While we did not try to derive such a structure from an actual compactification, it is known that explicit compactifications of intersecting branes exist, and that open strings connecting D-branes intersecting at non-vanishing angles lead to theories with chiral fermions, so it is plausible that such a model can exist. We follow Ibanez et al in assuming the supersymmetry breaking occurs in a hidden sector, and is transmitted by complex F-term moduli vevs to the superpartners. Then this model gives for soft terms

$$M_1 = M_3 = -A_t \sim e^{i\alpha_1}, \qquad M_2 \sim e^{-i\alpha_2}, \tag{76}$$

and all the other soft terms are real. One important lesson is that such a theory has only 9 parameters — the many parameters of L_{soft} have been reduced by the theory down to this number. They are

$$\alpha_2 - \alpha_1, m_{3/2}, \tan\beta, |\mu|, |A_t|, \phi_\mu, X_1, X_2, X_3. \tag{77}$$

Here only the relative phase of the moduli vevs enters, $m_{3/2}$ is the gravitino mass and sets the overall mass scale, and the X_i are measures of the relative importance of different moduli. The X_i could be measured, in which case they would tell us about the structure of the theory, and/or they could be computed in a good theory. Measuring the string-based parameters here would teach us about formulating and testing string theory. Any theory will have relations among the soft parameters so the actual number of parameters is far smaller than the full number of L_{soft}. This number could be reduced further by some assumptions. Also, not all of them will contribute in any given process, as we have seen. The resulting theory can be used to simultaneously study collider physics and LSP cold dark matter as is usual, and also CP violation. An extended version of the model[34] can also address flavor issues.

In this model one can illustrate how results of the low energy theory can appear fine-tuned and somewhat arbitrary because they are not apparently due to a symmetry when they originate in dynamics that occur at the high scale and are hidden at the low energy scale. If the gluino-gluon box diagram indeed explains direct CP violation in the kaon system, then one needs a certain phase relation to hold,

$$\arg(\phi_{A_{sd}} M_3^*) \approx 10^{-2}, \tag{78}$$

which seems fine-tuned. But as we saw in eq.76, M_3 and the elements of A have the same phase in this D-brane based theory, and so the quantity in eq.78 is zero at the high scale. Since the phases of M_3 and of A run differently, a small phase is generated at the low scale. While we are not arguing this is the actual explanation for ε'_K, it does nicely illustrate how such phases could be related by an underlying theory yet not follow from any low energy symmetry.

34. DIRECT EVIDENCE FOR SUPERPARTNERS?—AT THE TEVATRON?

So far all the evidence for low energy supersymmetry is indirect. Although the evidence is strong, it could in principle be a series of coincidences. More indirect evidence could come soon from improved $g_\mu - 2$, other rare decays, b-factories, proton decay, CDM detection. But finally it will be necessary to directly observe superpartners, and to show they are indeed superpartners. That could first happen at the Tevatron collider.

Indeed, as we discussed earlier, if supersymmetry is really the explanation for EWSB then the soft masses should be of order M_Z, and the cross sections for their production are typical EW ones, or larger for gluinos, so superpartners should be produced in significant quantities at the Tevatron collider that has just begun to take data after a six year upgrade in luminosity. Assuming the luminosity and the detectors are good enough to separate signals from backgrounds, if direct evidence for superpartners does not emerge at the Tevatron then either nature does not have low energy supersymmetry or there is something completely missing from our understanding of low energy supersymmetry. There is no other hint of such a gap in our understanding. Thus if superpartners do not appear at the Tevatron many people will wait until LHC has taken data to be convinced nature is not supersymmetric, but it is unlikely that superpartners could be found at the LHC if they are not first found at the Tevatron. So let us examine how they are likely to appear at the Tevatron.

Accepting that supersymmetry explains EWSB, we expect the gluinos, neutralinos, and charginos to be rather light. The lighter stop may be light as well. Sleptons may also be light though there is somewhat less motivation for that. We can list a number of channels and look at the signatures for each of them. Almost all cases require a very good understanding of the SM events that resemble the possible signals, both in magnitude (given the detector efficiencies) and the distributions. Missing transverse energy will be denoted by $\not{E}_T$. It is reasonable to expect the Tevatron to have an integrated luminosity of 2 fb^{-1} per detector by sometime in 2004, and 15

fb^{-1} by sometime in 2007. Until we know the ordering of the superpartner masses we have to consider a number of alternative decays of $\widetilde{N}_2$, $\widetilde{C}_1$, $\tilde{t}_1$, $\tilde{g}$, etc.[35]

• $\widetilde{N}_1 + \widetilde{N}_1$

This channel is very hard to tag at a hadron collider.

• $\widetilde{N}_1 + \widetilde{N}_{2,3}$

These channels can be produced through an s-channel Z or a t-channel squark exchange. The signatures depend considerably on the character of $\widetilde{N}_2$, $\widetilde{N}_3$. $\widetilde{N}_1$ escapes. If $\widetilde{N}_2$ has a large coupling to $\widetilde{N}_1 + Z$ (for real or virtual Z) then the $\widetilde{N}_1$ will escape and the Z will decay to e or μ pairs each 3% of the time, so the event will have missing energy and a prompt lepton pair. There will also be tau pairs and jet pairs, but those are somewhat harder to identify. Or, perhaps $\widetilde{N}_2$ is mainly photino and $\widetilde{N}_1$ mainly higgsino, in which case there is a large BR for $\widetilde{N}_2 \to \widetilde{N}_1 + \gamma$ and the signature of $\widetilde{N}_2$ is one prompt γ and missing energy. The production cross section can depend significantly on the wave functions of $\widetilde{N}_1, \widetilde{N}_2$. If the cross section is small for $\widetilde{N}_1 + \widetilde{N}_2$ it is likely to be larger for $\widetilde{N}_1 + \widetilde{N}_3$. Most cross sections for lighter channels will be larger than about 50 fb, which corresponds to 200 events (not including BR) for an integrated luminosity of 2 fb^{-1} per detector.

• $\widetilde{N}_1 + \widetilde{C}_1$

These states are produced through s-channel $W^{\pm}$ or t-channel squarks. The $\widetilde{N}_1$ escapes, so the signature comes from the $\widetilde{C}_1$ decay, which depends on the relative sizes of masses, but is most often $\widetilde{C}_1 \to l^{\pm} + \not{E}_T$. This is the signature if sleptons are lighter than charginos ($\widetilde{C}_1 \to \tilde{l}^{\pm} + \nu$, followed by $\tilde{l}^{\pm} \to l^{\pm} + \widetilde{N}_1$), or if sneutrinos are lighter than charginos by a similar chain, or by a three-body decay ($\widetilde{C}_1 \to \widetilde{N}_1 + virtual\ W$, $W \to l^{\pm} + \nu$). But it is not guaranteed — for example if stops are lighter than charginos the dominant decay could be $\widetilde{C}_1 \to \tilde{t} + b$. In the case where the lepton dominates the event signature is then $l^{\pm} + \not{E}_T$, so it is necessary to find an excess in this channel. Compared to the SM sources of such events the supersymmetry ones will have no prompt hadronic jets, and different distributions for the lepton energy and for the missing transverse energy.

• $\widetilde{N}_2 + \widetilde{C}_1$

If $\widetilde{N}_2$ decays via a Z to $\widetilde{N}_1 + l^+ + l^-$ and $\widetilde{C}_1$ decays to $\widetilde{N}_1 + l^{\pm}$, this channel gives the well-known "tri-lepton" signature, three charged leptons, $\not{E}_T$, and no prompt jets, which may be relatively easy to separate from SM background. But it may be that $\widetilde{N}_2 \to \widetilde{N}_1 + \gamma$, so the signature may be $l^{\pm} + \gamma + \not{E}_T$.

• $\tilde{l}^+ + \tilde{l}^-$

Sleptons may be light enough to be produced in pairs. Depending on masses, they could decay via $\tilde{l}^{\pm} \to l^{\pm} + \widetilde{N}_1$, $\widetilde{C}_1 + \nu$, $W + \tilde{\nu}$. If $\widetilde{N}_1$ is mainly higgsino decays to it are suppressed by lepton mass factors, so $\tilde{l}^{\pm} \to l^{\pm} + \widetilde{N}_2$ may dominate, followed by $\widetilde{N}_2 \to \widetilde{N}_1 + \gamma$.[36]

For a complete treatment one should list all the related channels, and combine those that can lead to similar signatures. The total sample may be dominated by one channel but have significant contributions from others, etc. It should also be emphasized that the so-called "backgrounds" are not junk backgrounds that cannot be calculated, but from SM events whose rates and distributions can be completely understood. Determining these background rates is essential to identifying a signal and to identifying new physics, and requires powerful tools in the form of simulation programs, which in turn require considerable expertise to use correctly. The total production cross section for all neutralino and chargino channels at the Tevatron collider is expected to be between 0.1 and 10 pb, depending on how light the superpartners are, so even in the worst case there should be several hundred events in the two detectors. If the cross sections are on the low side it will require combining inclusive signatures to demonstrate new physics has been observed.

• gluinos can be produced via several channels, $\tilde{g} + \tilde{g}$, $\tilde{g} + \widetilde{C}_1, \tilde{g} + \widetilde{N_1}$, etc.

If supersymmetry indeed explains EWSB it would be surprising if the gluino were heavier than about 500 GeV, as argued above. Then the total cross section for its production should be large enough to observe it at the Tevatron. If all its decays are three-body, e.g. $\tilde{g} \to \tilde{q} + \bar{q}$ followed by $\tilde{q} \to q + \widetilde{C}_1$, etc, then the signature has energetic jets, $\not{E}_T$, and sometimes charged leptons. There are two channels that are particularly interesting and not unlikely to occur — if $t + \tilde{t}$ or $b + \tilde{b}$ are lighter than $\tilde{g}$ then they will dominate because they are two-body. The signatures can then be quite different, with mostly b and c jets, and smaller multiplicity.

Gluinos and neutralinos are normally Majorana particles. Therefore they can decay either as particle or antiparticle. If, for example, a decay path $\tilde{g} \to \bar{t}(\to W^- \bar{b}) + \tilde{t}$ occurs, with $W^- \to e^- \nu$, there is an equal probability for $\tilde{g} \to e^+ + \ldots$. Then a pair of gluinos can with equal probability give same-sign or opposite sign dileptons! The same result holds for any way of tagging the electric charge — we just focus on leptons since their charge is easiest to identify. The same result holds for neutralinos. The SM allows no way to get prompt isolated same-sign leptons, so any observation of such events is a signal beyond the SM, and very likely a strong indication of supersymmetry.

• Stops can be rather light, so they should be looked for very seriously. They can be pair-produced via gluons, with a cross section that is about 1/8 of the top pair cross section. It is smaller because of a p-wave threshold suppression for scalars, and a factor of 4 suppression for the number of spin states. They could also be produced in top decays if they were lighter than $m_t - M_{\tilde{N}_1}$, and in gluino decays if they are lighter than $m_{\tilde{g}} - m_t$, which is not at all unlikely. Their obvious decay is $\tilde{t} \to \tilde{C} + b$, which will indeed dominate if $m_{\tilde{t}} > m_{\tilde{C}}$. If this relation does not hold, it may still dominate as a virtual decay, followed by $\tilde{C}$ real or virtual decay (say to $W + \tilde{N}_1$), in which case the final state is 4-body after W decays, and suppressed by 4-body phase space. That may allow the one-loop decay $\tilde{t} \to c + \tilde{N}_1$ to dominate stop decay. As an example of how various signatures may arise, if the mass ordering is $t > \tilde{C}_1 > \tilde{t} > \tilde{N}_1$ and $t > \tilde{t} + \tilde{N}_1$, then a produced $t\bar{t}$ pair will sometimes (depending on the relative branching ratio, which depends on the mass values) have one top decay to $W + b$ and the other to $c + \tilde{N}_1$, giving a $W + 2$ jets signature, with the jets detectable by b or charm tagging, and thus an excess of such events.

• An event was reported by the CDF collaboration from Tevatron Run 1, $p\bar{p} \to ee\gamma\gamma \not{E}_T$, that is interesting for several reasons, both as a possible signal and to illustrate some pedagogical issues. That such an event might be an early signal of supersymmetry was suggested in 1986. It can arise[36,37] if a selectron pair is produced, and if the LSP is higgsino-like, in which case the decay of the selectron to $e + \tilde{N}_1$ is suppressed by a factor of m_e. Then $\tilde{e} \to e + \tilde{N}_2$ dominates, followed by $\tilde{N}_2 \to \tilde{N}_1 + \gamma$. The only way to get such an event in the SM is production of $WW\gamma\gamma$ with both $W \to e + \nu$, with an overall probability of order 10^{-6} for such an event in Run 1. Other checks on kinematics, cross section for selectrons, etc., allow a supersymmetry interpretation, and the resulting values of masses do not imply any that must have been found at LEP or as other observable channels at the Tevatron, though over some of the parameter space some associated signal could have been seen. There are many consistency conditions that must be checked if such an interpretation is allowed, and a number of them could have failed but did not. Indeed, a related interpretation that had the decay of the selectron to electron plus very light gravitino is excluded by the absence of a signal at LEP for events with two photons and large missing energy. If this event were a signal additional ones would soon occur in Run 2. Because of the needed branching ratios there would be no trilepton signal at the Tevatron, since $\tilde{N}_2$ decays mainly into a photon instead of l^+l^-, and the decay of $\tilde{N}_3$ would be dominated by $\tilde{\nu}\nu$.

Although it might look easy to interpret any non-standard signal or excess as supersymmetry, in fact a little thought shows it is very difficult. As illustrated in the above examples, a given signature implies an ordering of superpartner masses, which implies a number of cross section and decay branching ratios. All must be right. All the couplings in the Lagrangian are determined, so there is little freedom once the masses are fixed by the kinematics of the candidate events. To prove a possible signal is indeed consistent with supersymmetry one has also to check that relations among couplings are indeed satisfied. Such checks will be easy at lepton colliders, but difficult at hadron colliders, so we do not focus on them here. There can of course be alternative interpretations of any new physics, but in all cases it will be possible to show the supersymmetry one is preferred (if it is indeed correct) — that is a challenge we would love to have.

35. AFTER THE FIRST CELEBRATION

Once a signal is found, presumably at the Tevatron, there will of course be a lot of checking required to confirm it because it will not be dramatic, as discussed above, but rather excesses in a few channels that slowly increase with integrated luminosity. Deducing even the masses of mass eigenstates may be difficult if more than one channel contributes significantly to a topological excess. Nevertheless, it will be possible to very quickly deduce some general results about supersymmetry breaking and how it is transmitted.

For example, one of the key questions is the nature of the LSP.[38] That can immediately exclude some ways to transmit supersymmetry breaking and favor others, and constrain ideas about how supersymmetry breaking occurs. From the discussion above we can make a table whose columns are various forms the LSP can take and whose rows are qualitative signatures that do not require complete studies, though they still require an understanding of the backgrounds:

	$\widetilde{B}$	$\tilde{\text{h}}$	$\widetilde{G}$	unstable
prompt $\gamma's$	no	some	yes	no
trileptons	yes	no	no	no
large $\not{E}_T$	yes	yes	yes	no

The table can be extended to other and more detailed LSP descriptions such as wino LSP, degenerate LSP and NLSP, etc. It can be extended to a number of additional signatures and made more quantitative. There are some caveats that can be added — e.g. for the gravitino case it can happen that long lifetimes for the lightest neutralino change the signature. But the basic point that qualitative features of the excess events will tell us a considerable amount remains. An unstable LSP implies that R-parity (or matter parity) is not conserved, a gravitino LSP implies gauge mediation for the way the supersymmetry breaking is transmitted, and the bino and higgsino LSP's suggest gravity mediation, with different consequences for cold dark matter.

36. EXTENSIONS OF THE MSSM

I want to emphasize that it may be very important to not restrict analysis of data by over constraining the MSSM with additional assumptions. Also I have focused on the MSSM for pedagogical simplicity, but nature could define simplicity differently. Surely the neutrino sector must be added, and that affects RGE's for the sectors we have examined. There is good motivation for extra U(1) symmetries, which may lead to extra D-terms and to extra neutralinos that mix to affect the neutralino mass eigenstates behavior and the CDM physics. There will be Planck-scale suppressed operators that may be crucial for flavor physics and for understanding the fermion masses and for precise calculations of gauge coupling unification. There may be extra scalars related to inflation, and axions, which affect cosmology and CDM physics. By using the MSSM without assuming relations among parameters many of these affects can be allowed for, while if parameters are related by ad hoc assumptions the extensions could only appear if inconsistencies appeared in the analysis — that is hard to see because of the initially large experimental uncertainties. For example, extra D-terms shift various scalar masses and separate $M^2_{H_u}$ and $M^2_{H_d}$, so assuming all the scalars masses are degenerate does not allow the D-term contributions to appear.

37. CONCLUDING REMARKS

These lectures have emphasized how to construct a supersymmetric description of nature at the weak scale based on forthcoming data from colliders, rare decays, static properties, cold dark matter studies, and more, and how to connect that to a unification scale description, so that we can eventually learn a complete effective Lagrangian near the Planck scale. That is the most that can be achieved by the traditional approach of science. If we also understand string theory (and we do not distinguish here between string theory and M-theory) well enough, possibly we will be able to bridge the gap to the Planck scale in 10 dimensions and formulate a fundamental theory. If so a number of features of the effective theory will be able to test ideas about the fundamental theory. The most important features of the experimental discovery of supersymmetry will be threefold: we will understand the natural world better; we will know we are on the right track to make more progress; and we will be opening a window to see physics at the Planck scale, which makes immensely more likely that we will be able to formulate and test a fundamental theory at the Planck scale.

Sometimes I am asked "what is left to compute" by students or postdocs looking for interesting projects, and interested workers in related areas. Much is indeed already known about supersymmetry from over two decades of work by a number of good people. But in fact we have just gotten to the stage where the most important problems can be addressed!. Little is known about how to relate data to the parameters of L_{soft} in practice, little is known about the flavor properties of L_{soft} and how to compute them theoretically or extract them from data uniquely, and little is known about how to relate data at the weak scale to an effective Lagrangian at the string scale. There is much to understand and to compute. The third of those issues will be the main focus of supersymmetry research once superpartners are being directly studied.

There are several practical features that should be emphasized. Unless we are missing important basic ideas, a Higgs boson and superpartners will be produced at the Tevatron collider. Supersymmetry signals have two escaping LSP's, so they are never dramatic or obvious or easy to interpret. They will appear as excesses in several channels, where channels are labeled by numbers of leptons and jets, and missing transverse energy. Once superpartners are found, the entire challenge to experimenters is to measure the parameters of L_{soft}, which has been the main subject of these lectures. The relations of the parameters of L_{soft} to data is complicated, and it is easy to get the wrong answers if care is not taken. Although there seem to be a large number of parameters, any given measurement depends only on a few, and most parameters enter in a number of places, so using information from one place in other analysis will greatly facilitate progress. Interpreting the data and learning its implications will be challenging, and it is a challenge we are eager to have.

ACKNOWLEDGMENTS

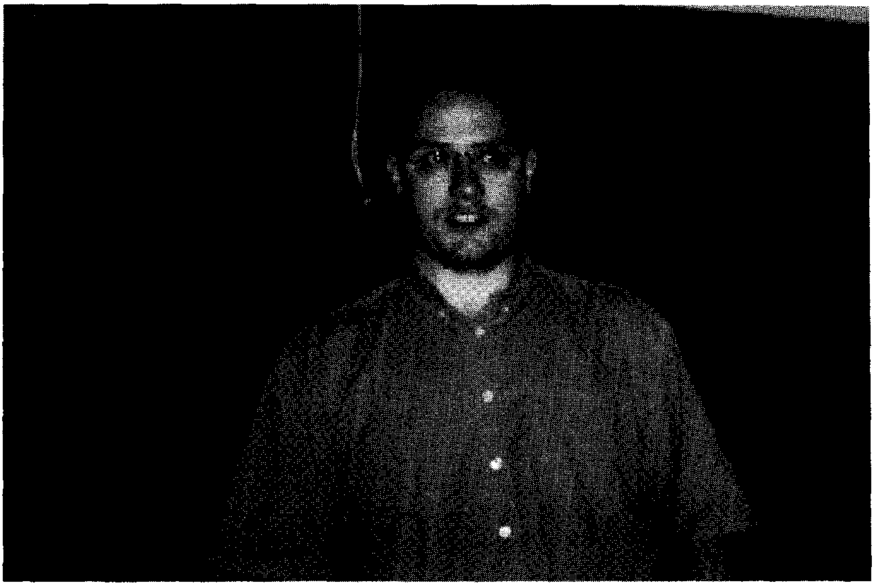
These lectures have benefited greatly from my collaborations and discussions with Mike Brhlik, Dan Chung, Lisa Everett, Steve King, and Lian-Tao Wang.

References

1. A Supersymmetry Primer, S.P.Martin, hep-ph/9709356
2. *Perspectives on Supersymmetry*, G.L.Kane (ed.), Singapore, World Scientific (1998)
3. I.S.Towner and J.C.Hardy, nucl-th/9809087
4. O.Lebedev and W.Loinaz, hep-ph/0106056; G.Altarelli, F.Caravaglios, G.F.Giudice, P.Gambino, and G.Ridolfi, hep-ph/0106029; G.-C.Cho and K.Hagiwara, hep-ph/0105037
5. L.Everett, G.L.Kane, and L.-T. Wang, in preparation
6. S.P.Martin and P.Ramond, Phys.Rev. D51(1995)6515; hep-ph/9501244; G.Kribs, Nucl. Phys. B535(1998)41; hep-ph/9803259; K.Dienes, Physics Reports 287 (1997) 447
7. The Supersymmetric Soft-Breaking Lagrangian: Theory and Applications, D.J.H.Chung, L.Everett, G.L.Kane, S.F.King, J. Lykken, and L.-T.Wang, in preparation.
8. S.Dimopoulos and H.Georgi, Nucl. Phys. B193(1981)150; L.Girardello and M.Grisaru, Nucl. Phys. B194(1982)65
9. The literature can be traced from I.Jack and D.R.T.Jones, Phys. Lett. B457(1999)101; hep-ph/9903365; J.L.Diaz-Cruz, Proceedings Merida 1999 "Particles and Fields", 299.
10. L.Ibanez and G.G.Ross, Nucl. Phys. B368(1992)3; S.P.Martin, Phys. Rev. D54(1996)2340; hep-ph/9602349
11. L.Ibanez and G.G.Ross, Phys. Lett. B110(1982)215; L. Alvarez-Gaume, M.Claudsen, and M.Wise, Nucl. Phys. B221(1983)495; K.Inoue, A.Kakuto, H.Komatsu, and S.Takeshita, Prog. Theor. Phys. 68(1982)927
12. See H.Haber in *Perspectives on Supersymmetry,* G.L.Kane (ed.), Singapore, World Scientific (1993)
13. For some review and references see Martin, [1].
14. LEP Electroweak Working Group, LEPEWWG/2001-01
15. G.L.Kane and J.Wells, hep-ph/0003249; see also the elaboration by M. Peskin and J. Wells, hep-ph/0101342
16. LEP Higgs Working Group for Higgs boson searches hep-ex/0107029; ALEPH-2001-066; ALEPH Collaboration hep-ph/0201014
17. G.L.Kane, S.F.King, and L.-T. Wang, Phys. Rev. D64(2001)095013; hep-ph/0010312
18. G.L.Kane and L.-T. Wang, Phys. Lett. B488(2000)383, hep-ph/0003198; M.Carena, J.Ellis, A.Pilaftsis, and C.Wagner, Phys. Lett. B495(2000)155, hep-ph/0009212; T.Ibrahim and P.Nath, Phys. Rev. D63(2001)035009, hep-ph/0008237
19. J.Goldstein, C.Hill, J.Incandela, S.Parke, D.Rainwater, D.Stuart, Phys. Rev. Lett. 86(2001)1694; L.Reina, S.Dawson, and D.Wackeroth, hep-ph/0109066; W.Beenakker,S.Dittmaier, M.Kramer, B.Plumper, M.Spira, P.Zerwas, hep-ph/0107081
20. The literature can be traced from M.Carena, H.Haber, H.Logan, S.Mrenna, hep-ph/0106116

21. D.Rainwater, D.Zeppenfeld, K.Hagiwara, Phys. Rev. D59(1999)014037; hep-ph/9808468
22. N.Ghodbane, S.Katsanevas, I.Laktineh, and J.Rosiek, hep-ph/0012031
23. M.Brhlik and G.L.Kane, Phys. Lett. B437(1998)331; hep-ph/9803391
24. S.Y.Choi, J.Kalinowski, G.Moortgat-Pick, P.M.Zerwas, hep-ph/0202039
25. J.Cline, M.Joyce, and K.Kainulainen, hep-ph/0110031
26. M.Carena, J.M.Morena, M.Quiros, M.Seco, and C.Wagner, Nucl. Phys. B599(2001) 158; hep-ph/0011055
27. Lisa L.Everett, G.L.Kane, S.Rigolin, and L.-T. Wang, Phys. Rev. Lett 86(2001)3484; hep-ph/0102145
28. T.Ibrahim and P.Nath, hep-ph/9908443
29. S.Mrenna, G.L.Kane, and L.-T.Wang, Phys. Lett. B483(2000)175; hep-ph/9910477
30. M.Brhlik, D.Chung, and G.L.Kane, Int. J. Mod. Phys. D10(2001)367; hep-ph/0005158
31. M.Brhlik, L.Everett, G.L.Kane, S.F.King, and O.Lebedev, Phys. Rev. Lett. 84(3041)2000; hep-ph/9909480
32. A.Nelson and L.Randall, Phys. Lett. B316 (1993) 516; hep-ph/9308277
33. M.Brhlik, L.Everett, G.L.Kane, and J.Lykken, Phys. Rev. Lett. 83(1999)2124, hep-ph/9905215; Phys. Rev. D62(2000)035005, hep-ph/9908326
34. L.Everett, G.L.Kane, S.King, S.Rigolin, and L.-T.Wang, hep-ph/0202100; See also S.Abel, S.Khalil, and O.Lebedev, hep-ph/0103031 and references therein
35. See chapters by J.-F.Grivaz; M.Carena, R.L.Culbertson, S.Eno, H.J.Frisch, and S.Mrenna; J.F.Gunion and H.Haber in ref.2.
36. See chapter by G.L.Kane in ref.2.
37. S.Ambrosanio, G.L.Kane, G.Kribs, S.Martin, and S.Mrenna, Phys. Rev. Lett. 76(1996)3498
38. G.L.Kane, Proceedings of Supersymmetry 1997, ed. M.Cvetic and P.Langacker.
39. L.Ibanez, C.Munoz, and S.Rigolin, hep-ph/9812397; Nucl. Phys. B553(1999)43.

This page intentionally left blank



ANDREW R. LIDDLE

This page intentionally left blank

RECENT DEVELOPMENTS IN COSMOLOGY

ANDREW R. LIDDLE
Astronomy Centre, University of Sussex
Brighton BN1 9QJ, United Kingdom
E-mail:a.liddle@sussex.ac.uk

1. Overview

The last few years in cosmology have been thrilling ones, as dramatic improvements in observational technology have begun to impose stringent constraints on theoretical ideas in cosmology built up over the preceding two decades. For the purposes of this article I'll focus on the following set of overall goals:

- To obtain a physical description of the Universe, including its global dynamics and matter content.
- To measure the cosmological parameters describing the Universe, and to develop a fundamental understanding of as many of those parameters as possible.
- To understand the origin and evolution of cosmic structures.
- To understand the physical processes which took place during the extreme heat and density of the early Universe.

Over recent years, much progress has been made on all of these topics, to the extent that it is widely believed amongst cosmologists that we may stand on the threshold of the first precision cosmology, in which the parameters necessary to describe our Universe have been identified and will soon be, in most cases at least, measured to a satisfying degree of precision. Whether this optimism has any grounding in reality remains to be seen, though so far the signs are promising in that the basic picture of cosmology, centred around the Hot Big Bang, has time and again proven the best framework for interpreting the constantly improving observational situation.

In particular, the process of cosmological parameter estimation is well underway, thanks to observations of distant Type Ia supernovae, of galaxy clustering, and of the cosmic microwave background. These have established a standard cosmological model, where the Universe is dominated by dark energy, contains substantial dark matter, and with the baryons from which we are made comprising only around 4%. Overall this model can be described by around ten parameters (e.g. see Ref. 1), and the viable region of parameter space is starting to shrink under pressure from observations. However, it is worth bearing in mind that we seek high precision determinations at least in part because they ought to shed light on fundamental physics, and there progress has been less rapid. Some parameters are likely to have no particular fundamental importance (for instance, there would probably be little fundamental significance were the Hubble constant to turn out to be $63\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ rather than say $72\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, though accurate determination of this parameter is essential if we are to pin down other parameters), but the 10% or so measured accuracy of the baryon density is to be set against the lack of even an order-of-magnitude theoretical understanding thus far.

This article does not attempt to cover the complete range of moden cosmology, but is intended as a status report on a subset of topics which I've chosen as being potentially of the most interest to theoretical particle physicists. The main descriptive sections concern structure formation in the Universe and the inflationary cosmology, and the final section is a mixed bag of especially topical subjects.

2. Structure Formation in the Universe

2.1. *Gravitational instability*

One of the most powerful tools in cosmology is the development of structures. By 'structure' I mean anything corresponding to inhomogeneity within the Universe, be it galaxies, variations in the gravitational potential, or anisotropies in the cosmic microwave background. The evolution of structures proves sensitive to all the main cosmological parameters, and hence is well suited to constraining them. Different types of observation naturally probe different physical regimes, for instance small verses large scales, and also different stages of the Universe's evolution, with the microwave background probing the Universe when it was around one thousandth of its present size.

The young universe was much closer to uniformity than the present state; for instance the irregularities in the cosmic microwave background are only around one part in 10^5, while the present matter distribution features highly overdense galaxies with voids in between. The main driving force in this evolution, at least in its initial stages, is simply gravity; any initial overdensity will exert an unbalanced gravitational force upon neighbouring material and will tend to accrete material, amplifying the original perturbations. At least until well after the cosmic microwave background radiation is released, the perturbation evolution is well described on all scales by linear perturbation theory, though ultimately linear theory for the density field breaks down on short scales as virialized galaxies begin to form. On sufficiently large scales linear theory remains adequate even today.

The Hot Big Bang model, supplemented by gravitational instability in order to form structures, gives an excellent broad-brush description of our Universe. However, like any theory or model in physics, its predictions depend on some input parameters not specified by the theory. A key goal is to measure those parameters to a satisfying degree of accuracy. For example, the detailed process of gravitational instability depends on

- The expansion rate of the Universe (the Hubble parameter).
- The density of the material providing the gravitational attraction.
- The physical properties of the material; for example does it only experience gravitational attraction, or are other interactions important?
- The form of the initial perturbations that get the whole structure formation process going.

Current ideas in cosmology suggest that around 10 parameters may be sufficient to describe our Universe. At present, however, we don't even know the complete set of important parameters, far less have accurate values for them all. The hope is that over the next few years we will both identify the important parameters and measure them to high accuracy, in many cases at the percent level.

2.2. *Quantifying microwave background anisotropies*

Although the strongest tests of cosmological models will always come from the combination of all available data, cosmic microwave background (CMB) anisotropies have received much attention lately (and are likely to be the single most important tool for constraining inflation, as discussed in the next section), and so it is worth spending some time defining the necessary terminology.

We observe the temperature $T(\theta, \phi)$ coming from different directions. We write this as a dimensionless perturbation and expand in spherical harmonics

$$\frac{T(\theta, \phi) - \bar{T}}{\bar{T}} = \sum_{\ell, m} a_{\ell m} Y_m^\ell(\theta, \phi) . \tag{1}$$

There is no unique prediction for the coefficients $a_{\ell m}$, but in the simplest inflationary cosmologies they are drawn from a gaussian distribution whose mean square is independent of m and given by the **radiation angular power spectrum**

$$C_\ell = \left\langle |a_{\ell m}|^2 \right\rangle_{\text{ensemble}} \tag{2}$$

The ensemble average represents the theorist's ability to average over all possible observers in the Universe (or indeed over different quantum mechanical realizations), whereas an observer's highest ambition is to estimate it by averaging over the multipoles of different m as seen at our own location. The radiation angular power spectrum depends on all the cosmological parameters, and so it can be used to constrain them. To extract the full information polarization also has to be measured; this gives three additional power spectra, describing two independent modes of polarization, and the cross-correlation between the temperature anisotropies and one polarization mode (other cross-correlations vanish assuming absence of parity violation).

Computation of the power spectra requires a lot of physics: gravitational collapse, photon–electron interactions (and their polarization dependence), neutrino free-streaming etc. But as long as the perturbations are small, linear perturbation theory can be used which makes accurate calculations possible. A major step forward for the field was the public release of Seljak & Zaldarriaga's code CMBFAST[2] which can compute the spectra within one percent accuracy for a given cosmological model in around one minute. An example spectrum is shown in Figure 1.

2.3. *Recent CMB results*

During 2000 and 2001 studies of microwave anisotropies took a huge leap forward with the first results from a new generation of instruments. First out with results was the Boomerang collaboration,[3] followed closely by the Maxima collaboration;[4] these made the first accurate mapping of the first peak in the angular power spectrum, corresponding to the first gravitational

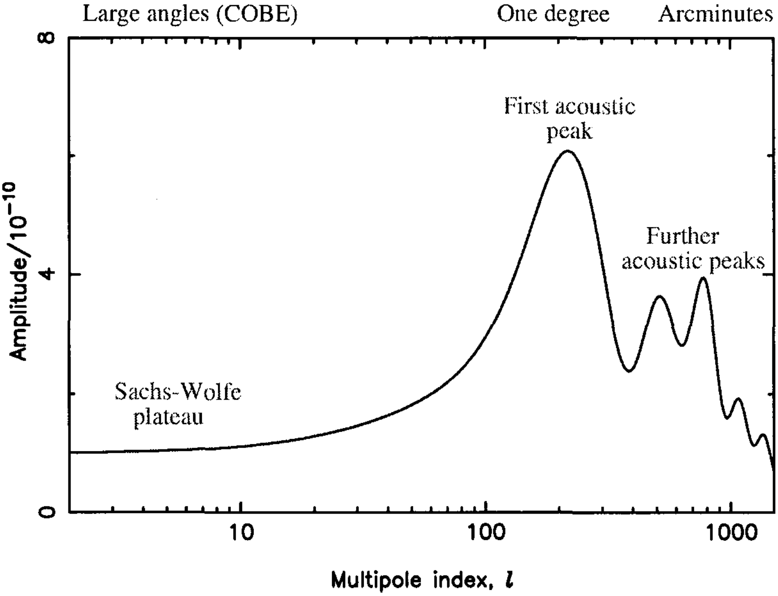


Figure 1. The radiation angular power spectrum for a particular cosmological model. The annotations name the different features.

compression of the primordial fluid. The location of this peak is fixed primarily by the propagation of light to us after last-scattering, and is a sensitive probe of the geometry of the Universe. These results are consistent with a flat geometry, with only a small margin for error, and provided a convincing exclusion of a low-density open Universe with $\Omega_0 \sim 0.3$ which had up until then been regarded as a viable cosmology.

The first Boomerang and Maxima results gave tentative, but inconclusive, indication of further features to small angular scales. The situation improved further in mid 2001, with new results from the DASI experiment[5] and a reanalysis of the Boomerang data[6] including a much larger fraction of the total dataset. These results are shown in Figure 2, alongside a best-fitting theoretical model. These latest results show the first clear evidence for further oscillations in the angular power spectrum, as predicted in Figure 1. This observation is of particular qualitative significance for the inflationary cosmology, as discussed in the next section, and of quantitative significance for constraining the baryon density as described in the following subsection.

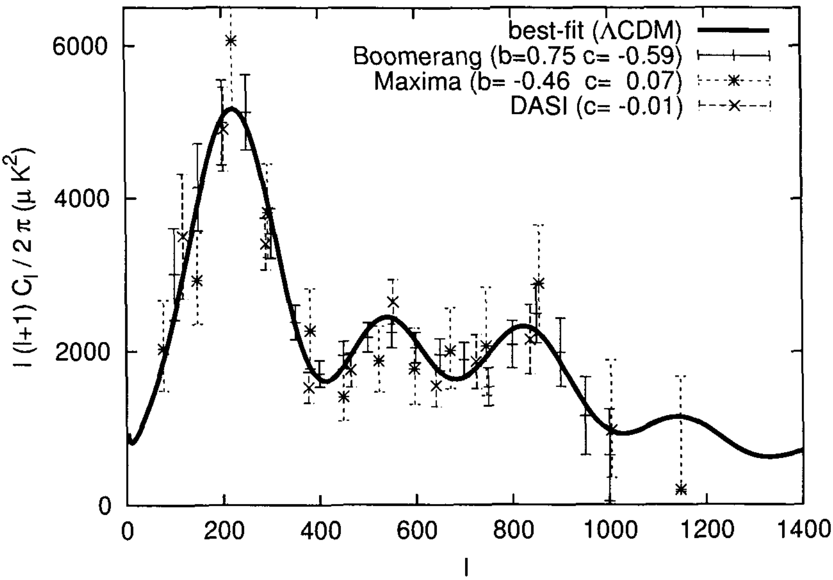


Figure 2. A best-fit cosmology with a cosmological constant is shown in comparison to recent CMB anisotropy results. The data sets of Boomerang, Maxima and DASI are shown with the best-fit calibration values for those experiments presuming this best-fit model is correct. [Figure courtesy of Julien Lesgourgues.]

2.4. *The standard cosmological model*

The observations of the last few years have led to the establishment of a standard cosmological model, with ingredients as follows.

Cosmological constant	$\sim 66\%$
Cold dark matter	$\sim 30\%$
Baryons	$\sim 4\%$
Photons and neutrinos	$\sim 10^{-4}$

Spatial flatness.

Hubble constant around $70\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$.

Initial conditions seeded by slow-roll inflation.

This model is in remarkable agreement with observational data.

Baryons

There are now three independent powerful ways of estimating the baryon density of the Universe. Listing the uncertainties at 95% confidence, we have

Nucleosynthesis: $\Omega_{\text{baryon}} h^2 = 0.019 \pm 0.002$.
It is widely, though not universally, thought that the measurement of the deuterium abundance in high-redshift absorption systems gives a highly-accurate probe of the baryon density during nucleosynthesis.

Microwave background: $\Omega_{\text{baryon}} h^2 = 0.02 \pm 0.01$.
The baryon density can be inferred from the CMB spectrum, as it governs the relative heights of the first and second peaks (corresponding to compressions and rarefactions of the cosmic fluid respectively). While the Boomerang 2000 results gave a suspiciously high value for this, reanalysis in 2001 plus new results from DASI have brought the value into excellent agreement with nucleosynthesis.

Cluster baryon fraction: $\Omega_{\text{baryon}}/(\Omega_{\text{cdm}} + \Omega_{\text{baryon}}) = 0.12 \pm 0.05$.
Clusters are observed via X-ray emission from hot intracluster gas. This gas is in hydrostatic equilibrium against gravity which is principally supplied by the dark matter. For the standard cosmological model this agrees excellently with nucleosynthesis.

Cosmological constant

Famously, in 1998 two teams studying distant supernovae discovered that they were fainter than expected, and having eliminated other possible causes concluded that this was due to the expansion history of the Universe, and required a presently-accelerating cosmology.[7] This can be brought about by a cosmological constant Λ, and if one additionally restricts to a flat geometry as motivated by the CMB, this leads to the cosmological constant density of the Standard Cosmological Model.

Now, if that was the sole evidence for a cosmological constant I wouldn't believe it for a second. However the circumstantial evidence is extremely powerful; for instance:

(1) Microwave anisotropies show the Universe is flat (or close to flat), provided the initial perturbations are adiabatic.
(2) Nucleosynthesis plus the cluster baryon fraction imply $\Omega_{\text{cdm}} + \Omega_{\text{baryon}} \sim 0.3$ which implies $\Omega_\Lambda \sim 0.7$.

(3) The correct galaxy power spectrum is reproduced if $(\Omega_{\rm cdm} + \Omega_{\rm baryon})h \simeq 0.2$ (where h is the Hubble constant in the usual units); this concurs well with direct measures of h.

As a result of this and other arguments, the so-called ΛCDM model presently has no serious rivals.

Table 1.

Current	NASA's Map satellite was launched in mid-2001 and is currently making an all-sky survey of the CMB (results late 2003??).
2002	Maxima and Boomerang make the first serious attempts to measure CMB polarization anisotropies.
2001–2004	Main operations phase of the Sloan Digital Sky Survey,[8] seeking to redshift a million galaxies.
2002–2005	First systematic surveys for high-redshift galaxy clusters using X-rays and the Sunyaev–Zel'dovich effect.
2007	ESA's *Planck* satellite launched, for high-resolution all-sky mapping of CMB temperature and polarization.
2010??	Launch of the LISA satellites, capable of probing a stochastic gravitational wave background (though not the inflationary one except in exceptional models).

The cosmological constant poses the twin problems of its unexpectedly small magnitude (in fundamental physics terms) and the mystery of why it should only come to dominate the Universe at the present epoch (around redshift 0.3). To address these, instead of a pure cosmological constant, one might prefer an effective one, for example a slowly-rolling potential-dominated scalar field as described in the next section for early Universe inflation. Such scenarios are known as quintessence. Current observations force such scenarios to be quite close to the pure cosmological constant, and though differences may yet be unveiled by improved experiments it appears only quite limited information will be available. It is actually becoming quite hard to construct simple quintessence models capable of matching all observations while employing plausible initial conditions.

2.5. *What's coming up?*

Table 1 shows a selection (far from complete) of things to look out for in coming years which will drive further moves to precision cosmology.

3. The Inflationary Cosmology

3.1. *Overview and models*

This section focusses on the last two of the goals listed at the start of this article. The claim is that during the very early Universe, a physical process known as **inflation** took place, which still manifests itself in our present Universe via the perturbations it left behind which later led to the development of structure in the Universe. By studying those structures, we hope to shed light on whether inflation occurred, and by what physical mechanism. An extensive account of inflation appears in my textbook with David Lyth.[9]

I begin by defining inflation. The scale factor of the Universe at a given time is measured by the scale factor $a(t)$. In general a homogeneous and isotropic Universe has two characteristic length scales, the curvature scale and the Hubble length. The Hubble length is more important, and is given by

$$cH^{-1} \quad \text{where} \quad H \equiv \frac{\dot{a}}{a} \,. \tag{3}$$

Typically, the important thing is how the Hubble length is changing with time as compared to the expansion of the Universe, i.e. what is the behaviour of the **comoving Hubble length** H^{-1}/a?

During any standard evolution of the Universe, such as matter or radiation domination, the comoving Hubble length increases. It is then a good estimate of the size of the observable Universe. **Inflation** is defined as any epoch of the Universe's evolution during which the comoving Hubble length is decreasing

$$\frac{d\left(H^{-1}/a\right)}{dt} < 0 \Longleftrightarrow \ddot{a} > 0 \,, \tag{4}$$

and so inflation corresponds to any epoch during which the Universe has accelerated expansion. During this time, the expansion of the Universe outpaces the growth of the Hubble radius, so that physical conditions can become correlated on scales much larger than the Hubble radius, as required to solve the horizon and flatness problems.

As discussed in the last section, there is very good evidence from observations of Type Ia supernovae that the Universe is *presently* accelerating.[7] This is usually attributed to the presence of a cosmological constant. This is clearly at some level good news for those interested in the possibility of inflation in the early Universe, as it indicates that inflation is possible in principle, and certainly that any purely theoretical arguments which suggest inflation is not possible should be treated with some skepticism.

If the Universe contains a fluid, or combination of fluids, with energy density ρ and pressure p, then

$$\ddot{a} > 0 \Longleftrightarrow \rho + 3p < 0\,, \tag{5}$$

(where the speed of light c has been set to one). As we always assume a positive energy density, inflation can only take place if the Universe is dominated by a material which can have a negative pressure. Such a material is a scalar field, usually denoted ϕ. A homogeneous scalar field has a kinetic energy and a potential energy $V(\phi)$, and has an effective energy density and pressure given by

$$\rho = \frac{1}{2}\dot{\phi}^2 + V(\phi) \quad ; \quad p = \frac{1}{2}\dot{\phi}^2 - V(\phi)\,. \tag{6}$$

The condition for inflation can be satisfied if the potential dominates.

A model of inflation typically amounts to choosing a form for the potential, perhaps supplemented with a mechanism for bringing inflation to an end, and perhaps may involve more than one scalar field. In an ideal world the potential would be predicted from fundamental particle physics, but unfortunately there are many proposals for possible forms. Instead, it has become customary to assume that the potential can be freely chosen, and to seek to constrain it with observations. A suitable potential needs a flat region where the potential can dominate the kinetic energy, and there should be a minimum with zero potential energy in which inflation can end. Simple examples include $V = m^2\phi^2/2$ and $V = \lambda\phi^4$, corresponding to a massive field and to a self-interacting field respectively. Modern model building can get quite complicated — see Ref. 10 for a review.

3.2. *Inflationary cosmology: perturbations*

By far the most important aspect of inflation is that it provides a possible explanation for the origin of cosmic structures. The mechanism is fundamentally quantum mechanical; although inflation is doing its best to make the Universe homogeneous, it cannot defeat the uncertainty principle which ensures that residual inhomogeneities are left over. These are stretched to astrophysical scales by the inflationary expansion. Further, because these are determined by fundamental physics, their magnitude can be predicted independently of the initial state of the Universe before inflation. However, the magnitude does depend on the model of inflation; different potentials predict different cosmic structures.

One way to think of this is that the field experiences a quantum 'jitter' as it rolls down the potential. The observed temperature fluctuations in the cosmic microwave background are one part in 10^5, which ultimately means that the quantum effects should be suppressed compared to the classical evolution by this amount.

Inflation models generically predict two independent types of perturbation:

Density perturbations $\delta_{\mathrm{H}}^2(k)$**:** These are caused by perturbations in the scalar field driving inflation, and the corresponding perturbations in the space-time metric.

Gravitational waves $A_{\mathrm{T}}^2(k)$**:** These are caused by perturbations in the space-time metric alone.

They are sometimes known as scalar and tensor perturbations respectively, because of the way they transform. Density perturbations are responsible for structure formation, but gravitational waves can also affect the microwave background.

We do not expect to be able to predict the precise locations of cosmic structures from first principles (any more than one can predict the precise position of a quantum mechanical particle in a box). Rather, we need to focus on statistical measures of clustering. Simple models of inflation predict that the amplitudes of waves of a given wavenumber k obey gaussian statistics, with the amplitude of each wave chosen independently and randomly from a gaussian. What it does predict is how the width of the gaussian, known as its amplitude, varies with scale; this is known as the **power spectrum**.

With current observations it is a good approximation to take the power spectra as being power laws with scale, so

$$\delta_{\mathrm{H}}^2(k) = \delta_{\mathrm{H}}^2(k_0)\left[\frac{k}{k_0}\right]^{n-1} \tag{7}$$

$$A_{\mathrm{T}}^2(k) = A_{\mathrm{T}}^2(k_0)\left[\frac{k}{k_0}\right]^{n_{\mathrm{T}}} \tag{8}$$

In principle this gives four parameters — two amplitudes and two spectral indices — but in practice the spectral index of the gravitational waves is unlikely to be measured with useful accuracy, which is rather disappointing as the simplest inflation models predict a so-called consistency relation relating n_T to the amplitudes of the two spectra, which would be a distinctive test of inflation. The assumption of power-laws for the spectra requires assessment both in extreme areas of parameter space and whenever observations significantly improve.

3.3. *The current status of inflation*

The best available constraints come from combining data from different sources; for two recent attempts see Wang et al.[1] and Efstathiou et al.[11] Suitable data include observations of the recent dynamics of the Universe using Type Ia supernovae, cosmic microwave anisotropy data, and galaxy correlation function data.

Currently inflation is a definite qualitative success, with striking agreement between the predictions of the simplest inflation models and observations. In particular, the locations of the microwave anisotropy power spectrum peaks are most simply interpreted as being due to an adiabatic initial perturbation spectrum in a spatially-flat Universe. The multiple peak structure strongly suggests that the perturbations already existed at a time when their corresponding scale was well outside the Hubble radius. No unambiguous evidence of nongaussianity has been seen.

Quantitatively, however, things have some way to go. At present the best that has been done is to try and constrain the parameters of the power-law approximation to the inflationary spectra. The gravitational waves have not been detected and so their amplitude has only an upper limit and their spectral index is not constrained at all. The current situation can be summarized as follows.

Amplitude δ_H: COBE determines this (assuming no gravitational waves) to about ten percent accuracy (at one-sigma) as approximately $\delta_H = 1.9 \times 10^{-5}\,\Omega_0^{-0.8}$ on a scale close to the present Hubble radius (see Ref. 12, 9 for accurate formulae).

Spectral index n: This is thought to lie in the range $0.8 < n < 1.05$ (at 95% confidence). It would be extremely interesting were the scale-invariant case, $n = 1$, to be convincingly excluded, as that would be clear evidence of dynamical processes at work, rather than symmetries, in creating the perturbations.

Gravitational waves r: Measured in terms of the relative contribution to large-angle microwave anisotropies, the tensors are currently constrained to be no more than about 30%.

3.4. *Inflation and CMB oscillations*

A key property of inflationary perturbations is that they were created in the early Universe and evolved freely from then. Although a general solution to the perturbation equations has two modes, growing and decaying, only the growing mode will remain by the time the perturbation enters the horizon. This leads directly to the prediction of an oscillatory structure in the microwave anisotropy power spectrum, as seen in Figure 1.[13] The existence of such a structure is a robust prediction of inflation; if it is not seen then inflation cannot be the sole origin of structure.

The most significant recent development in observations pertaining to inflation is the first clear evidence for multiple peaks in the spectrum, seen by the DASI [5] and Boomerang[6] experiments shown in Figure 2. This is a crucial qualitative test which inflation appears to have passed, and which could have instead provided evidence against the entire inflationary paradigm for structure formation. These observations lend great support to inflation, though it must be stressed that they are not able to 'prove' inflation, as it may be that there are other ways to produce such an oscillatory structure.[14]

3.5. *Prospects for the future*

It remains possible that future observations will slap us in the face and lead to inflation being thrown out. But if not, we can expect an incremental succession of better and better observations, culminating (in terms of currently-funded projects) with the *Planck* satellite.[15] Faced with observational data of exquisite quality, an initial goal will be to test whether the simplest models of inflation continue to fit the data, meaning models with a single scalar field rolling slowly in a potential $V(\phi)$ which is then to be constrained by observations. If this class of models does remain viable, we can move on to reconstruction of the inflaton potential from the data.

Planck, currently scheduled for launch in February 2007, should be highly accurate. In particular, it should be able to measure the spectral index to an accuracy better than ± 0.01, and detect gravitational waves even if they are as little as 10% of the anisotropy signal. In combination with other observations, these limits could be expected to tighten significantly further, especially the tensor amplitude. Such observations would rule out almost all currently known inflationary models. Even so, there will be considerable uncertainties, so it is important not to overstate what can be achieved.

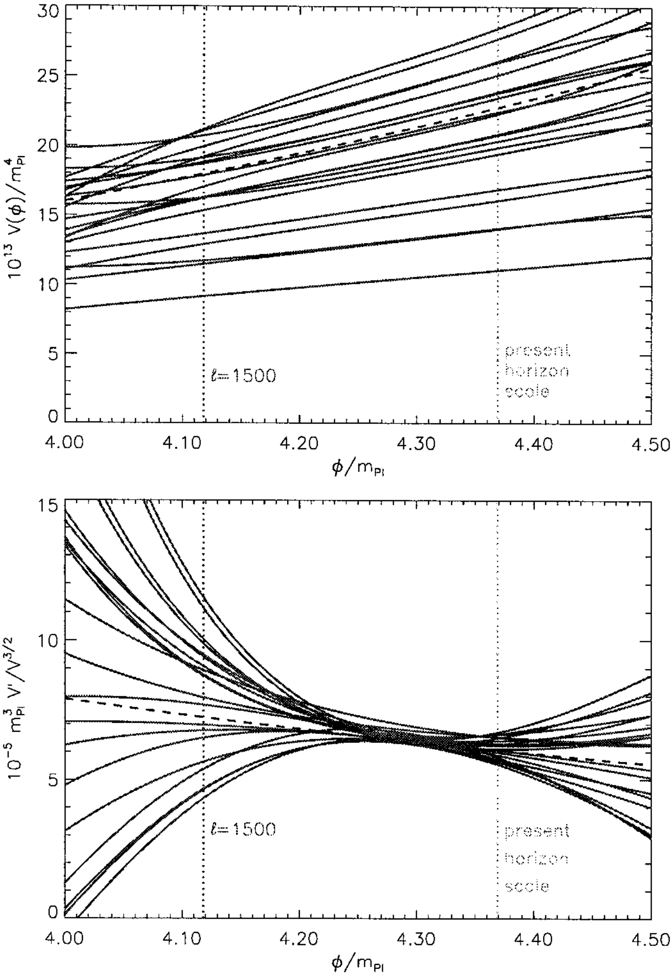


Figure 3. Sample reconstruction of a potential, where the dashed line shows the true potential and the solid lines are thirty Monte Carlo reconstructions (real life can only provide one). The upper panel shows the potential itself which is poorly determined. However some combinations, such as $(dV/d\phi)/V^{3/2}$ shown in the lower panel, can be determined at an accuracy of a few percent. See Ref. 16 for details.

Reconstruction can only probe the small part of the potential where the field rolled while generating perturbations on observable scales. We know enough about the configuration of the *Planck* satellite to be able to estimate how well it should perform. Ian Grivell and I recently described a numerical technique which gives an optimal construction.[16] Results of an example reconstruction are shown in Figure 3, where it was assumed that the true potential was $\lambda\phi^4$. The potential itself is not well determined here (the tensors are only marginally detectable), but certain combinations, such as $(dV/d\phi)/V^{3/2}$, are accurately constrained and would lead to high-precision constraints on inflation model parameters.

4. Selected topics

The previous sections overviewed the status of two areas of cosmology. In this section, I give a short account of some particular topics which have received a lot of attention lately.

4.1. *Does cold dark matter really work?*

Of the topics in this article, perhaps the one with the greatest potential significance for elementary particle physicists is this one: Is the dark matter really cold? Until lately the opinion of the astrophysics community was united behind this assumption, based on the remarkable success with which the cold dark matter model explains many observations in structure formation. However, more recently questions have opened up as to whether or not the cold dark matter assumption gives a good fit to observations on short scales, and in particular to the structure within galaxies.

High-resolution simulations of galaxy formation indicate that the dark matter retains considerable clumpiness when the small structures which first form are absorbed into larger structures. The persistence of substructure is a success for explaining the structure of galaxy clusters, where thousands of galaxies can retain their identity upon assimilation, but fails dismally in explaining our own galaxy where only a handful of dwarf satellites are observed.[17] Even if they were stripped of their visible baryon components, such knots of dark matter would be sufficient to destroy the observed thin disks of spiral galaxies.

Potentially related to this are two further problems:

- Dwarf galaxy cores: theory predicts that the density diverges towards the centre of halos, whereas in well-observed dwarf galaxies a uniform-density core is seen.
- Bulge constitution: enough microlensing events have been seen towards the Milky Way bulge to suggest that they explain *all* the dark matter in the central regions of our galaxy, leaving no room for particle dark matter.

It remains unclear whether these problems are really so robust that the cold dark matter paradigm is under serious threat. However they have been taken sufficiently seriously as to motivate a slew of papers on alternative sark matter properties, including warm dark matter, self-interacting dark matter, annihilating dark matter or condensated dark matter. Whether any of these could provide a unified solution to the problems listed above is unclear, but needless to say all have major consequences for dark matter search strategies and particle physics phenomenology.

4.2. *Do neutrinos play a role?*

What role can neutrinos play in cosmology? This is a topical question as evidence mounts up in favour of a non-zero neutrino mass from solar and atmospheric neutrino experiments.

Assuming standard interactions and negligible lepton asymmetry, theory predicts

$$\Omega_\nu = \frac{\sum_i m_{\nu_i}}{90\, h^2\, \mathrm{eV}} \tag{9}$$

Neutrinos could have a measurable effect on structure formation, through their free-streaming, even if Ω_ν were as little as 0.01, meaning $m_\nu \sim 0.5$ eV suggesting that neutrinos play only a very minor role in structure formation. However recent observations favour even smaller values. However if for some reason there were a substantial lepton asymmetry, there could be an effect at even smaller masses.

It's also worth noting that, even if massless, the behaviour of relic neutrinos does have to be taken into account to compute quantities such as the microwave anisotropy power spectrum. Observations of it therefore do offer an indirect confirmation that the relic neutrino population predicted by theory does in fact exist.

4.3. *Are existing treatments of inflation oversimplistic?*

Much of the information disseminated from the inflationary community to the broader physics and astronomy community is based around the simplest paradigm, where a single scalar field slow-rolls down a potential. While this continues to be in excellent agreement with observation, and is a powerful working hypothesis worthy of testing in its own right, much work has recently gone into studying more complicated situations. It is a continuing challenge to uncover the full phenomenology of models with more than one dynamical field. Such a situation changes many of the usual assumptions. There is no longer a unique trajectory for the inflation field, and predictions for the density perturbations may well become dependent on initial conditions. Perhaps more importantly, the perturbations are no longer guaranteed to be adiabatic, and isocurvature perturbations may well be non-gaussian and/or correlated with the adiabatic component. If the single-field paradigm fails, it will be important to understand whether there remain well-motivated inflationary models capable of explaining the data, particularly as efforts to determine cosmological parameters such as h and Ω_{B} will flounder if the initial perturbations cannot be accurately parametrized.

4.4. *The braneworld*

At a particle physics conference or school you can quite happily state that 'it is generally accepted that our Universe has more than three spatial dimensions' without much fear of contradiction, though it is probably not necessary even to step out of a physics building to find out how limited this general acceptance actually is. Nevertheless, extra dimensions have been with particle theorists consistently for a long time now, and undoubtedly they are an issue which may be of considerable importance for early Universe cosmology.

Until recently, it had been assumed that the failure to observe the predicted extra dimensions meant that the extra ones were "curled up" to be unobservably small. However, there is a now a new idea, the **braneworld**, which proposes that at least one of these extra dimensions might be relatively large, with us constrained to live on a three-dimensional **brane** running through the higher-dimensional space. Gravity is able to propagate in the full higher-dimensional space, which is known as the **bulk**.

This radical idea has many implications for cosmology, both in the present and early Universe, and so far we have only scratched the surface of possible new phenomena. Already many exciting results have been obtained; here there is only space to consider a few pertinent questions.

a) Are there modifications to the evolution of the homogeneous Universe? The answer appears to be yes; for example in a simple scenario (known as Randall–Sundrum Type II[18]) the Friedmann equation is modified at high energies so that, after some simplifying assumptions, it reads[19]

$$H^2 = \frac{8\pi G}{3}\left(\rho + \frac{\rho^2}{2\lambda}\right), \tag{10}$$

where λ is the tension of the brane. This recovers the usual cosmology at low energies $\rho \ll \lambda$, but otherwise we have new behaviour. This opens new opportunities for model building, see for example Ref. 20.

b) Are inflationary perturbations different?
Again the answer is yes — there are modifications to the formulae giving scalar and tensor perturbations.[21] Unfortunately the main effect of this is to introduce new degeneracies in interpreting observations, as a potential can always be found matching observations for any value of λ .[22] The initial perturbations therefore cannot be used to test the braneworld scenario.

c) Do perturbations evolve differently after they are laid down on large scales?
The answer here is less clear. It is possible that perturbation evolution is modified even at late times, e.g. perturbations in the bulk could influence the brane in a way that couldn't be predicted from brane variables alone. Whether there is a significant effect is unclear and is likely to be model dependent.

4.5. *The Ekpyrotic universe*

It has recently been proposed that the Big Bang is actually the result of the collision of two branes, dubbed the Ekpyrotic Universe.[23] It has been claimed that this scenario can provide a resolution to the horizon and flatness problems, essentially because causality arises from the higher-dimensional theory and allows a simultaneous Big Bang everywhere on our brane, though existing implementations solve the problem by hand in the initial conditions. As I write this, it remains unclear how to successfully describe the instant of collision between the two branes (the singularity problem), and considerable controversy surrounds whether or not the scenario can also generate nearly scale-invariant adiabatic perturbations.[24] Both aspects are required to make it a serious rival to inflation.[a]

Acknowledgments

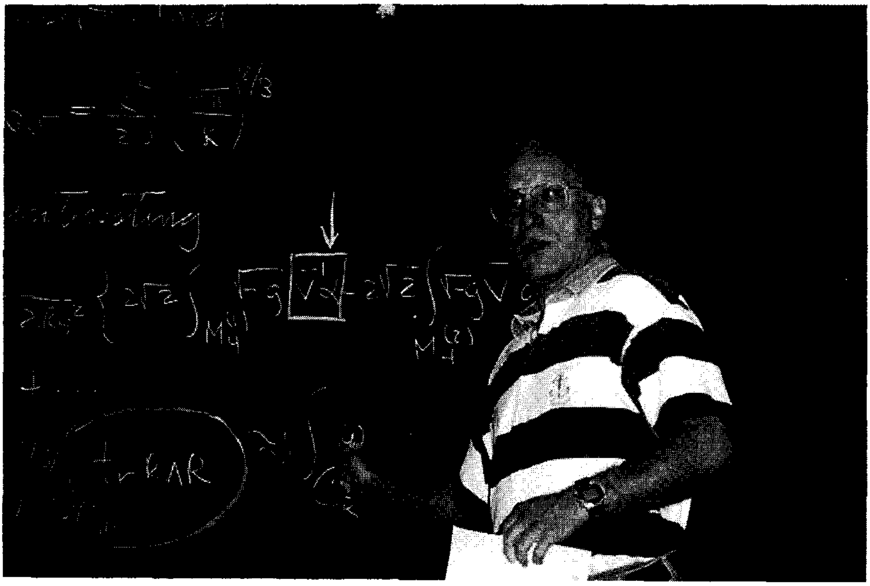
The author was supported in part by the Leverhulme Trust.

[a] Since the talks on which this article is based were given, there is also a newer incarnation of the scenario known as the Cyclic Universe[25] which further extends these ideas.

References

1. X. Wang, M. Tegmark and M. Zaldarriaga, astro-ph/0105091.
2. U. Seljak and M. Zaldarriaga, *Astrophys. J.* **469**, 1 (1996).
3. P. de Bernardis et al., 2000, *Nature*, 404, 955 (2000).
4. S. Hanany et al., *Astrophys. J. Lett.*, **545**, L5 (2000).
5. N. W. Halverson et al., astro-ph/0104489; C. Pryke et al., astro-ph/0104490.
6. C. B. Netterfield et al., astro-ph/0104460; P. de Bernardis et al., astro-ph/0105296.
7. S. Perlmutter, *Nature* **391**, 51 (1998); A. G. Riess et al., *Astronomical. J.* **116**, 1009 (1998); S. Perlmutter et al., *Astrophys. J.* **517**, 565 (1999).
8. Sloan Digital Sky Survey home page at `http://www.sdss.org`
9. A. R. Liddle and D. H. Lyth, *Cosmological inflation and large-scale structure*, Cambridge University Press (2000).
10. D. H. Lyth and A. Riotto, *Phys. Rep.* **314**, 1 (1999).
11. G. Efstathiou et al., astro-ph/0109152.
12. E. F. Bunn and M. White, *Astrophys. J.* **480**, 6 (1997); E. F. Bunn, A. R. Liddle and M. White, *Phys. Rev.* D**54**, 5917R (1996).
13. A. Albrecht, D. Coulson, P. Ferreira and J. Magueijo, *Phys. Rev. Lett.* **76**, 1413 (1996); A. Albrecht in *Critical Dialogues in Cosmology*, ed N. Turok, World Scientific 1997, `astro-ph/9612017`; W. Hu and M. White, *Phys. Rev. Lett.* **77**, 1687 (1996).
14. N. Turok, *Phys. Rev. Lett.* **77**, 4138 (1996).
15. *Planck* home page at `http://astro.estec.esa.nl/Planck/`
16. I. J. Grivell and A. R. Liddle, *Phys. Rev.* D**61**, 081301 (2000).
17. A. A. Klypin, A. V. Kravtsov, O. Valenzuela and F. Prada, *Astrophys. J.* **522**, 82 (1999); B. Moore, S. Ghigna, F. Governato, G. Lake, T. Quinn, J. Stadel and P. Tozzi, *Astrophys. J. Lett.* **524**, L19 (1999).
18. L. Randall and R. Sundrum, *Phys. Rev. Lett.* **83**, 4690 (1999).
19. C. Csaki, M. Graesser, C. Kolda and J. Terning, *Phys. Lett.* B**462**, 34 (1999); J. M. Cline, C. Grojean and G. Servant, *Phys. Rev. Lett.* **83**, 4245 (1999); P. Binétruy, C. Deffayet, U. Ellwanger and D. Langlois, *Phys. Lett.* B**477**, 285 (2000); T. Shiromizu, K. Maeda and M. Sasaki, *Phys. Rev.* D**62**, 024012 (2000).
20. E. J. Copeland, A. R. Liddle and J. E. Lidsey, *Phys. Rev.* D**64**, 023509 (2001).
21. R. Maartens, D. Wands, B. A. Bassett, and I. P. C. Heard, *Phys. Rev.* D**62**, 041301 (2000); D. Langlois, R. Maartens, and D. Wands, *Phys. Lett.* B**489**, 259 (2000).
22. A. R. Liddle and A. N. Taylor, *Phys. Rev.* D**65**, 041301 (2002).
23. J. Khoury, B. A. Ovrut, P. J. Steinhardt and N. Turok, *Phys. Rev.* D**64**, 123522 (2001) and hep-th/0108187 [see also R. Kallosh, L. Kofman and A. Linde, *Phys. Rev.* D**64**, 123523 (2001)].
24. D. H. Lyth, *Phys. Lett.* B**524**, 1 (2002); R. Brandenberger and F. Finelli, *JHEP* **0111**, 056 (2001); J. Khoury, B. A. Ovrut, P. J. Steinhardt and N. Turok, hep-th/0109050; J. Hwang, astro-ph/0109045.
25. P. J. Steinhardt and N. Turok, hep-th/0111030.

This page intentionally left blank



BURT A. OVRUT

This page intentionally left blank

LECTURES ON HETEROTIC M-THEORY

BURT A. OVRUT
Department of Physics
University of Pennsylvania
Philadelphia, PA 19104-6396
E-mail: ovrut@ovrut.hep.upenn.edu

We present three lectures on heterotic M-theory and a fourth lecture extending this theory to more general orbifolds. In Lecture 1, Hořava-Witten theory is briefly discussed. We then compactify this theory on Calabi-Yau threefolds, choosing the "standard" embedding of the spin connection in the gauge connection. We derive, in detail, both the five-dimensional effective action and the associated actions of the four-dimensional "end-of-the-world" branes. Lecture 2 is devoted to showing that this theory naturally admits static, $N = 1$ supersymmetry preserving BPS three-branes, the minimal vacuum having two such branes. One of these, the "visible" brane, is shown to support a three-generation E_6 grand unified theory, whereas the other emerges as the "hidden" brane with unbroken E_8 gauge group. Thus heterotic M-theory emerges as a fundamental paradigm for so-called "brane world" scenarios of particle physics. In Lecture 3 , we introduce the concept of "non-standard" embeddings. These are shown to permit a vast generalization of allowed vacua, leading on the visible brane to new grand unified theories, such as $SO(10)$ and $SU(5)$, and to the standard model $SU(3)_C \times SU(2)_L \times U(1)_Y$. It is demonstrated that non-standard embeddings generically imply the existence of five-branes in the bulk space. The physical properties of these bulk branes is discussed in detail. Finally, in Lecture 4 we move beyond Hořava-Witten theory and consider orbifolds larger than $S^1/\mathbf{Z}_2$. For explicitness, we consider M-theory orbifolds on $S^1/\mathbf{Z}_2 \times T^4/\mathbf{Z}_2$, discussing their anomaly structure in detail and completely determining both the untwisted and twisted sector spectra.

1. The Five-Dimensional Effective Theory

In this first lecture, we introduce our notation and briefly discuss the theory of the strongly coupled heterotic sting introduced by Hořava and Witten. In this theory, there is an eleven-dimensional bulk space bounded on either end of the x^{11}-direction by two "end-of-the-world" ten-dimensional nine-branes, each supporting an $N = 1$, E_8 supergauge theory. We then begin our construction of heterotic M-theory by compactifying the Hořava-Witten theory on a Calabi-Yau threefold. This leads to a five-dimesional bulk space bounded at the ends of the fifth dimesion by two end-of-the-world four-dimensional three-branes. Assuming, in this lecture, the "standard" embedding of the spin connection into one of the E_8 gauge connections we derive, in detail, both the five-dimensional bulk space effective action and the associated actions of the four-dimensional boundary branes. We end this lecture by discussing some of the properties of this effective theory and explicitly giving the $N = 2$ supersymmetry transformations of the bulk space quantum fields.

We begin by briefly reviewing the description of strongly coupled heterotic string theory as 11-dimensional supergravity with boundaries, as given by Hořava and Witten.[1,2] Our conventions are as follows. We will consider eleven-dimensional spacetime compactified on a Calabi-Yau space X, with the subsequent reduction down to four dimensions effectively provided by a double-domain-wall background, corresponding to an S^1/Z_2 orbifold. We use coordinates x^I with indices $I, J, K, \cdots = 0, \cdots, 9, 11$ to parameterize the full 11–dimensional space M_{11}. Throughout these lectures, when we refer to orbifolds, we will work in the "upstairs" picture with the orbifold S^1/Z_2 in the x^{11}–direction. We choose the range $x^{11} \in [-\pi\rho, \pi\rho]$ with the endpoints being identified. The Z_2 orbifold symmetry acts as $x^{11} \to -x^{11}$. Then there exist two ten–dimensional hyperplanes fixed under the Z_2 symmetry which we denote by $M_{10}^{(i)}$, $i = 1, 2$. Locally, they are specified by the conditions $x^{11} = 0, \pi\rho$. Barred indices $\bar{I}, \bar{J}, \bar{K}, \cdots = 0, \cdots, 9$ are used for the ten–dimensional space orthogonal to the orbifold. We use indices $A, B, C, \cdots = 4, \cdots 9$ for the Calabi–Yau space. All fields will be required to have a definite behaviour under the Z_2 orbifold symmetry in $D = 11$. We demand a bosonic field Φ to be even or odd; that is, $\Phi(x^{11}) = \pm\Phi(-x^{11})$. For a spinor Ψ the condition is $\Gamma_{11}\Psi(-x^{11}) = \Psi(x^{11})$ so that the projection to one of the orbifold planes leads to a ten–dimensional Majorana–Weyl spinor with positive chirality. Spinors in eleven dimensions will be Majorana spinors with 32 real components throughout the paper.

The bosonic part of the action is of the form

$$S = S_{\rm SG} + S_{\rm YM} \tag{1}$$

where $S_{\rm SG}$ is the familiar 11–dimensional supergravity

$$S_{\rm SG} = \tag{2}$$

$$-\frac{1}{2\kappa^2}\int_{M^{11}} \sqrt{-g}[R + \frac{1}{24}G_{IJKL}G^{IJKL} + \frac{\sqrt{2}}{1728}\epsilon^{I_1 \ldots I_{11}} C_{I_1 I_2 I_3} G_{I_4 \ldots I_7} G_{I_8 \ldots I_{11}}]$$

and $S_{\rm YM}$ are the two E_8 Yang–Mills theories on the orbifold planes explicitly given by

$$S_{\rm YM} = -\frac{1}{8\pi\kappa^2}\left(\frac{\kappa}{4\pi}\right)^{2/3}\int_{M_{10}^{(1)}}\sqrt{-g}\,\left\{{\rm tr}(F^{(1)})^2 - \frac{1}{2}{\rm tr}R^2\right\}$$
$$-\frac{1}{8\pi\kappa^2}\left(\frac{\kappa}{4\pi}\right)^{2/3}\int_{M_{10}^{(2)}}\sqrt{-g}\,\left\{{\rm tr}(F^{(2)})^2 - \frac{1}{2}{\rm tr}R^2\right\}\ . \qquad (3)$$

Here $F^{(i)}_{\bar{I}\bar{J}}$ are the two E_8 gauge field strengths and C_{IJK} is the 3–form with field strength $G_{IJKL} = 24\,\partial_{[I}C_{JKL]}$. In order for the above theory to be supersymmetric and anomaly free, the Bianchi identity for G should be modified such that

$$(dG)_{11\bar{I}\bar{J}\bar{K}\bar{L}} = -4\sqrt{2}\pi\left(\frac{\kappa}{4\pi}\right)^{2/3}\left\{J^{(1)}\delta(x^{11}) + J^{(2)}\delta(x^{11}-\pi\rho)\right\}_{\bar{I}\bar{J}\bar{K}\bar{L}} \qquad (4)$$

where the sources are given by

$$J^{(i)} = \frac{1}{16\pi^2}\left({\rm tr}F^{(i)}\wedge F^{(i)} - \frac{1}{2}{\rm tr}R\wedge R\right)\ . \qquad (5)$$

Under the Z_2 orbifold symmetry, the field components $g_{\bar{I}\bar{J}}$, $g_{11,11}$, $C_{\bar{I}\bar{J}11}$ are even, while $g_{\bar{I}11}$, $C_{\bar{I}\bar{J}\bar{K}}$ are odd.

The modification of the right hand side of equation (4) has important consequences. While the standard embedding of the spin connection of the Calabi–Yau threefold into the gauge connection

$${\rm tr}F^{(1)}\wedge F^{(1)} = {\rm tr}R\wedge R \qquad (6)$$

leads to vanishing source terms in the weakly coupled heterotic string Bianchi identity (which, in turn, allows one to set the antisymmetric tensor gauge field to zero), in the present case, one is left with non–zero sources $\pm{\rm tr}R\wedge R$ on the two hyperplanes. This follows from the fact that the sources in the Bianchi identity (4) are located on the orbifold planes with the gravitational part distributed equally between the two planes. The consequence is that not all components of the antisymmetric tensor field G can vanish. We find, for the standard embedding (6), that all components of G vanish with the exception of

$$G_{ABCD} = -\frac{1}{6}\alpha\,\epsilon_{ABCD}{}^{EF}\,\omega_{EF}\,\epsilon(x^{11}) \qquad (7)$$

where

$$\alpha = \frac{1}{8\sqrt{2}\pi v^{2/3}}\left(\frac{\kappa}{4\pi}\right)^{2/3}\int_{C_\omega}{\rm tr}R\wedge R. \qquad (8)$$

Here $\epsilon(x^{11})$ is the step function which is +1 (−1) for x^{11} positive (negative) and

$$v = \int_X \sqrt{\Omega} \tag{9}$$

where Ω_{AB} is a fixed Calabi–Yau metric and v is the associated volume of the Calabi–Yau threefold. The two–form ω_{AB} is the Kähler form associated with Ω_{AB} (that is, $\omega_{a\bar{b}} = i\Omega_{a\bar{b}}$ where a and $\bar{b}$ are holomorphic and anti-holomorphic indices) and $\mathcal{C}_\omega$ is the Poincare dual four-cycle of ω. Furthermore, in deriving this result, we have turned off all Calabi–Yau moduli with the exception of the radial breathing mode. This will be sufficient for all applications dealing with the universal moduli.

Phenomenologically, there is a regime where the universe appears five-dimensional. We would, therefore, like to derive an effective theory in the space consisting of the usual four space-time dimensions and the orbifold. We will, for simplicity, consider the universal zero modes only; that is, the five–dimensional graviton supermultiplet and the breathing mode of the Calabi–Yau space, along with its superpartners. These form a hypermultiplet in five dimensions. Furthermore, to keep the discussion as straightforward as possible, we will not consider boundary gauge matter fields. This simple framework suffices to illustrate our main ideas and was presented as such in Ref. 3. The general case was presented in Ref. 4. Our five-dimensional conventions are the following. Upon reduction on the Calabi-Yau space we have a five-dimensional spacetime M_5 labeled by indices $\alpha, \beta, \gamma, \cdots = 0, \cdots, 3, 11$. The orbifold fixed planes become four-dimensional with indices $\mu, \nu, \rho, \cdots = 0, \cdots, 3$. The 11-dimensional Dirac–matrices Γ^I with $\{\Gamma^I, \Gamma^J\} = 2g^{IJ}$ are decomposed as $\Gamma^I = \{\gamma^\alpha \otimes \lambda, \mathbf{1} \otimes \lambda^A\}$ where γ^α and λ^A are the five– and six–dimensional Dirac matrices, respectively. Here, λ is the chiral projection matrix in six dimensions with $\lambda^2 = 1$. In five dimensions we use symplectic-real spinors[5] ψ^i where $i = 1, 2$ is an $SU(2)$ index, corresponding to the automorphism group of the $N = 1$ supersymmetry algebra in five dimensions. We will follow the conventions given in Ref. 6.

We can perform the Kaluza-Klein reduction on the metric

$$ds_{11}^2 = V^{-2/3} g_{\alpha\beta} dx^\alpha dx^\beta + V^{1/3} \Omega_{AB} dx^A dx^B \ . \tag{10}$$

Since the compactification is on a Calabi–Yau manifold, the background corresponding to metric (10). preserves eight supercharges, the appropriate number for a reduction down to five dimensions. It might appear that we are simply performing a standard reduction of 11–dimensional supergravity on a Calabi–Yau space to five dimensions; for example, in the way described in Ref. 7. There are, however, two important ingredients that we have not yet included. One is obviously the existence of the boundary theories. We will return to this point shortly. First, however, let us explain a somewhat unconventional addition to the bulk theory that must be included.

Specifically, for the nonvanishing component G_{ABCD} in eq. (7) there is no corresponding zero mode field [a]. Therefore, in the reduction, we should take this part of G explicitly into account. In the terminology of Ref. 8, such an antisymmetric tensor field configuration is called a "non–zero mode". A more recent name for such a field configuration is a non-vanishing "G–flux". More generally, a non–zero mode is a background antisymmetric tensor field that solves the equations of motion but, unlike antisymmetric tensor field moduli, has nonvanishing field strength. Such configurations, for a p–form field strength, can be identified with the cohomology group $H^p(M)$ of the manifold M and, in particular, exist if this cohomology group is nontrivial. In the case under consideration, the relevant cohomology group is $H^4(X)$ which is nontrivial for a Calabi–Yau manifold X since $h^{2,2} = h^{1,1} \geq 1$. Again, the form of G_{ABCD} in eq. (7) is somewhat special, reflecting the fact that we are concentrating here on the universal moduli. In the general case, G_{ABCD} would be a linear combination of all harmonic $(2,2)$–forms.

The complete configuration for the antisymmetric tensor field that we use in the reduction is given by

$$\begin{array}{lll} C_{\alpha\beta\gamma}\ , & G_{\alpha\beta\gamma\delta} = 24\,\partial_{[\alpha}C_{\beta\gamma\delta]} & \\ C_{\alpha AB} = \frac{1}{6}\mathcal{A}_\alpha\omega_{AB}\ , & G_{\alpha\beta AB} = \mathcal{F}_{\alpha\beta}\omega_{AB}\ , & \mathcal{F}_{\alpha\beta} = \partial_\alpha\mathcal{A}_\beta - \partial_\beta\mathcal{A}_\alpha \\ C_{ABC} = \frac{1}{6}\xi\omega_{ABC}\ , & G_{\alpha ABC} = \partial_\alpha\xi\omega_{ABC} & \end{array} \tag{11}$$

and the non–zero mode is

$$G_{ABCD} = -\frac{\alpha}{6}\epsilon_{ABCD}{}^{EF}\,\omega_{EF}\,\epsilon(x^{11})\ , \tag{12}$$

where α was defined in eq. (8). Here, ω_{ABC} is the harmonic $(3,0)$ form on the Calabi–Yau space and ξ is the corresponding (complex) scalar zero mode. In addition, we have a five-dimensional vector field $\mathcal{A}_\alpha$ and 3–form

[a]This can be seen from the mixed part of the Bianchi identity $\partial_\alpha G_{ABCD} = 0$ which shows that the constant α in eq. (7) cannot be promoted as stands to a five–dimensional field. It is possible to dualize in five dimensions so the constant α is promoted to a five-form field, but we will not pursue this formulation here.

$C_{\alpha\beta\gamma}$, which can be dualized to a scalar σ. The total bulk field content of the five–dimensional theory is then given by the gravity multiplet

$$(g_{\alpha\beta}, \mathcal{A}_\alpha, \psi^i_\alpha) \tag{13}$$

together with the universal hypermultiplet

$$(V, \sigma, \xi, \bar{\xi}, \zeta^i). \tag{14}$$

Here ψ^i_α and ζ^i are the gravitini and the hypermultiplet fermions respectively and $i = 1, 2$ since they each form a doublet under the $SU(2)$ automorphism group of $N = 2$ supersymmetry in five dimesnions. From their relations to the 11–dimensional fields, it is easy to see that $g_{\mu\nu}$, $g_{11,11}$, $\mathcal{A}_{11}$, σ must be even under the Z_2 action whereas $g_{\mu 11}$, $\mathcal{A}_\mu$, ξ must be odd.

Examples of compactifications with non–zero modes in pure 11–dimensional supergravity on various manifolds including Calabi–Yau three–folds have been studied in Ref. 9. There is, however, one important way in which our non–zero mode differs from other non–zero modes in pure 11–dimensional supergravity. Whereas the latter may be viewed as an optional feature of generalized Kaluza-Klein reduction, the non–zero mode in Hořava–Witten theory that we have identified cannot be turned off. This can be seen from the fact that the constant α in expression (12) cannot be set to zero, unlike the case in pure 11–dimensional supergravity where it would be arbitrary, since it is fixed by eq. (8) in terms of Calabi–Yau data. This fact is, of course, intimately related to the existence of the boundary source terms, particularly in the Bianchi identity (4).

Let us now turn to a discussion of the boundary theories. In the five–dimensional space M_5 of the reduced theory, the orbifold fixed planes constitute four–dimensional hypersurfaces which we denote by $M_4^{(i)}$, $i = 1, 2$. Clearly, since we have used the standard embedding, there will be an E_6 gauge field $A_\mu^{(1)}$accompanied by gauginos and gauge matter fields on the orbifold plane $M_4^{(1)}$. For simplicity, we will set these gauge matter fields to zero in the following. The field content of the orbifold plane $M_4^{(2)}$ consists of an E_8 gauge field $A_\mu^{(2)}$ and the corresponding gauginos. In addition, there is another important boundary effect which results from the non–zero internal gauge field and gravity curvatures. More precisely, for the standard embedding defined in (6)

$$\begin{aligned}
&\int_X \sqrt{^6 g}\, \mathrm{tr} F^{(1)}_{AB} F^{(1)AB} = \int_X \sqrt{^6 g}\, \mathrm{tr} R_{AB} R^{AB} = 16\sqrt{2}\pi v \left(\frac{4\pi}{\kappa}\right)^{2/3} \alpha\ , \\
&\qquad F^{(2)}_{AB} = 0\ .
\end{aligned} \tag{15}$$

In view of the boundary actions (3), it follows that we will retain cosmological type terms with opposite signs on the two boundaries. Note that the size of those terms is set by the same constant α, given by eq. (8), which determines the magnitude of the non-zero mode.

We can now compute the five-dimensional effective action of Hořava–Witten theory. Using the field configuration (10)–(15) we find from the action (1)–(3) that

$$S_5 = S_{\rm grav} + S_{\rm hyper} + S_{\rm bound} \tag{16}$$

where

$$S_{\rm grav} = -\frac{1}{2\kappa_5^2}\int_{M_5}\sqrt{-g}\left[R + \frac{3}{2}\mathcal{F}_{\alpha\beta}\mathcal{F}^{\alpha\beta} + \frac{1}{\sqrt{2}}\epsilon^{\alpha\beta\gamma\delta\epsilon}\mathcal{A}_\alpha\mathcal{F}_{\beta\gamma}\mathcal{F}_{\delta\epsilon}\right] \tag{17a}$$

$$\begin{aligned} S_{\rm hyper} = -\frac{1}{2\kappa_5^2}\int_{M_5}\sqrt{-g}\,\Big[&\frac{1}{2}V^{-2}\partial_\alpha V\partial^\alpha V \\ &+2V^{-1}\partial_\alpha\xi\partial^\alpha\bar{\xi} + \frac{1}{24}V^2 G_{\alpha\beta\gamma\delta}G^{\alpha\beta\gamma\delta} \\ &+\frac{\sqrt{2}}{24}\epsilon^{\alpha\beta\gamma\delta\epsilon}G_{\alpha\beta\gamma\delta}\left(i(\xi\partial_\epsilon\bar{\xi} - \bar{\xi}\partial_\epsilon\xi) + 2\alpha\mathcal{A}_\epsilon\right) + \frac{1}{3}V^{-2}\alpha^2\Big] \end{aligned} \tag{17b}$$

$$\begin{aligned} S_{\rm bound} = &-\frac{1}{2\kappa_5^2}\left\{2\sqrt{2}\int_{M_4^{(1)}}\sqrt{-g}\,V^{-1}\alpha - 2\sqrt{2}\int_{M_4^{(2)}}\sqrt{-g}\,V^{-1}\alpha\right\} \\ &-\frac{1}{16\pi\alpha_{\rm GUT}}\sum_{i=1}^{2}\int_{M_4^{(i)}}\sqrt{-g}\,V{\rm tr}F_{\mu\nu}^{(i)\,2}\,. \end{aligned} \tag{17c}$$

In this expression, we have now dropped higher-derivative terms. The 4-form field strength $G_{\alpha\beta\gamma\delta}$ is subject to the Bianchi identity

$$(dG)_{11\mu\nu\rho\sigma} = -\frac{2\sqrt{2}\pi\kappa_5^2}{\alpha_{\rm GUT}}\left\{J^{(1)}\delta(x^{11}) + J^{(2)}\delta(x^{11}-\pi\rho)\right\}_{\mu\nu\rho\sigma} \tag{18}$$

which follows directly from the 11-dimensional Bianchi identity (4). The currents $J^{(i)}$ have been defined in eq. (5). The five-dimensional Newton constant κ_5 and the Yang-Mills coupling $\alpha_{\rm GUT}$ are expressed in terms of 11-dimensional quantities as

$$\kappa_5^2 = \frac{\kappa^2}{v}\,, \qquad \alpha_{\rm GUT} = \frac{\kappa^2}{2v}\left(\frac{4\pi}{\kappa}\right)^{2/3}\,. \tag{19}$$

We have checked the consistency of the truncation which leads to the above action by an explicit reduction of the 11-dimensional equations of motion to five dimensions. Note that the potential terms in the bulk and on the boundaries arise precisely from the inclusion of the non–zero mode and the gauge and gravity field strengths, respectively. Since we have compactified on a Calabi–Yau space, we expect the bulk part of the above action to have eight preserved supercharges and, therefore, to correspond to minimal $N = 1$ supergravity in five dimensions. Accordingly, let us compare the result (17) to the known $N = 1$ supergravity–matter theories in five dimensions.[6,10,11,12]

In these theories, the scalar fields in the universal hypermultiplet parameterize a quaternionic manifold with coset structure $\mathcal{M}_Q = SU(2,1)/SU(2) \times U(1)$. Hence, to compare our action to these we should dualize the three–form $C_{\alpha\beta\gamma}$ to a scalar field σ by setting (in the bulk)

$$G_{\alpha\beta\gamma\delta} = \frac{1}{\sqrt{2}} V^{-2} \epsilon_{\alpha\beta\gamma\delta\epsilon} \left(\partial^\epsilon \sigma - i(\xi \partial^\epsilon \bar{\xi} - \bar{\xi} \partial^\epsilon \xi) - 2\alpha\epsilon(x^{11}) \mathcal{A}^\epsilon \right) . \qquad (20)$$

Then the hypermultiplet part of the action (17b) can be written as

$$S_{\text{hyper}} = -\frac{v}{2\kappa^2} \int_{M_5} \sqrt{-g} \left[h_{uv} \nabla_\alpha q^u \nabla^\alpha q^v + \frac{1}{3} V^{-2} \alpha^2 \right] \qquad (21)$$

where $q^u = (V, \sigma, \xi, \bar{\xi})$. The covariant derivative ∇_α is defined as $\nabla_\alpha q^u = \partial_\alpha q^u - \alpha\epsilon(x^{11}) \mathcal{A}_\alpha k^u$ with $k^u = (0, -2, 0, 0)$. The sigma model metric $h_{uv} = \partial_u \partial_v K_Q$ can be computed from the Kähler potential

$$K_Q = -\ln(S + \bar{S} - 2C\bar{C}) , \quad S = V + \xi\bar{\xi} + i\sigma , \quad C = \xi . \qquad (22)$$

Consequently, the hypermultiplet scalars q^u parameterize a Kähler manifold with metric h_{uv}. It can be demonstrated that k^u is a Killing vector on this manifold. Using the expressions given in Ref. 13, one can show that this manifold is quaternionic with coset structure $\mathcal{M}_Q$. Hence, the terms in eq. (21) that are independent of α describe the known form of the universal hypermultiplet action. How do we interpret the extra terms in the hypermultiplet action depending on α? A hint is provided by the fact that one of these α-dependent terms modifies the flat derivative in the kinetic energy to a generalized derivative ∇_α. This is exactly the combination that we would need if one wanted to gauge the $U(1)$ symmetry on $\mathcal{M}_Q$ corresponding to the Killing vector k^u, using the gauge field $\mathcal{A}_\alpha$ in the gravity supermultiplet. In fact, investigation of the other terms in the action, including the fermions, shows that the resulting five-dimensional theory is precisely a gauged form of supergravity. Not only is a $U(1)$ isometry of $\mathcal{M}_Q$ gauged, but at the same time a $U(1)$ subgroup of the $SU(2)$ automorphism group is also gauged.

What about the remaining α-dependent potential term in the hypermultiplet action? From $D = 4$, $N = 2$ theories, we are used to the idea that gauging a symmetry of the quaternionic manifold describing hypermultiplets generically introduces potential terms into the action when supersymmetry is preserved (see for instance[14]). Such potential terms can be thought of as the generalization of pure Fayet-Iliopoulos terms. This is precisely what happens in our theory as well, with the gauging of the $U(1)$ subgroup inducing the α-dependent potential term in (21). The general gauged action was discussed in detail in Ref. 4.

The phenomenon that the inclusion of non-zero modes leads to gauged supergravity theories has already been observed in type II Calabi-Yau compactifications.[15,16] From the form of the Killing vector, we see that it is only the scalar field σ, dual to the 4–form $G_{\alpha\beta\gamma\delta}$, which is charged under the $U(1)$ symmetry. Its charge is fixed by α. We note that this charge is quantized since, suitably normalized, $\mathrm{tr}R \wedge R$ is an element of $H^{2,2}(X, \mathbf{Z})$.

To analyze the supersymmetry properties of the solutions shortly to be discussed, we need the supersymmetry variations of the fermions associated with the theory (16). They can be obtained either by a reduction of the 11–dimensional gravitino variation or by generalizing the known five–dimensional transformations[6,12] by matching onto gauged four–dimensional $N = 2$ theories. It is sufficient for our purposes to keep the bosonic terms only. Both approaches lead to

$$
\begin{aligned}
\delta\psi^i_\alpha &= D_\alpha\epsilon^i + \frac{\sqrt{2}i}{8}\left(\gamma_\alpha{}^{\beta\gamma} - 4\delta^\beta_\alpha\gamma^\gamma\right)\mathcal{F}_{\beta\gamma}\epsilon^i \\
&\quad -\frac{1}{2}V^{-1/2}\left(\partial_\alpha\xi\,(\tau_1 - i\tau_2)^i{}_j - \partial_\alpha\bar{\xi}\,(\tau_1 + i\tau_2)^i{}_j\right)\epsilon^j \\
&\quad -\frac{\sqrt{2}i}{96}V\epsilon_\alpha{}^{\beta\gamma\delta\epsilon}G_{\beta\gamma\delta\epsilon}(\tau_3)^i{}_j\epsilon^j + \frac{\sqrt{2}}{12}\alpha V^{-1}\epsilon(x^{11})\gamma_\alpha(\tau_3)^i{}_j\epsilon^j \qquad (23)\\
\delta\zeta^i &= \frac{\sqrt{2}}{48}V\epsilon^{\alpha\beta\gamma\delta\epsilon}G_{\alpha\beta\gamma\delta}\gamma_\epsilon\epsilon^i \\
&\quad -\frac{i}{2}V^{-1/2}\gamma^\alpha\left(\partial_\alpha\xi\,(\tau_1 - i\tau_2)^i{}_j + \partial_\alpha\bar{\xi}\,(\tau_1 + i\tau_2)^i{}_j\right)\epsilon^j \\
&\quad +\frac{i}{2}V^{-1}\gamma_\beta\partial^\beta V\epsilon^i + \frac{i}{\sqrt{2}}\alpha V^{-1}\epsilon(x^{11})(\tau_3)^i{}_j\epsilon^j
\end{aligned}
$$

where τ_i are the Pauli spin matrices.

In summary, we see that the relevant five-dimensional effective theory for the reduction of Hořava-Witten theory is a gauged $N = 1$ supergravity theory with bulk and boundary potentials.

2. The Domain Wall Solution and Generalizations

In the second lecture, we show that the effective five-dimensional bulk space theory does not have flat space for its static vacuum. Instead, the theory naturally admits static, $N = 1$ supersymmetry preserving BPS three-branes, the minmal vacuum consisting of two end-of-the-world three-branes. One of these branes, the one with the spin connection embedded in the gauge connection, supports a three generation E_6 grand unified theory and, hence, is called the "visible" or physical brane. The other brane is the "hidden" brane with an unbroken E_8 supergauge theory. Thus, heterotic M-theory emerges as a fundamental paradigm for so-called "brane world" scenarios of particle physics. In the second part of this lecture, we generalize the results of Lecture 1 to include, not just the universal hypermultiplet, but all (1,1)-moduli in the bulk space, as well as matter scalar multiplets on the boundary three-branes.

In order to re-construct the $D = 4$, $N = 1$ effective theory originally discussed we expect there to be a three–brane domain wall in five dimensions with a worldvolume lying in the four uncompactified directions. These solutions should break half the supersymmetry of the five–dimensional bulk theory and preserve Poincaré invariance in four dimensions.This domain wall can be viewed as the "vacuum" of the five–dimensional theory, in the sense that it provides the appropriate background for a reduction to the $D = 4$, $N = 1$ effective theory.

We notice that the theory (16) has all of the prerequisites necessary for such a three–brane solution to exist. Generally, in order to have a $(D-2)$–brane in a D–dimensional theory, one needs to have a $(D-1)$–form field or, equivalently, a cosmological constant. This is familiar from the eight–brane[30] in the massive type IIA supergravity in ten dimensions,[31] and has been systematically studied for theories in arbitrary dimension obtained by generalized (Scherk-Schwarz) dimensional reduction.[32] In our case, this cosmological term is provided by the bulk potential term in the action (16). From the viewpoint of the bulk theory, we could have multi three–brane solutions with an arbitrary number of parallel branes located at various places in the x^{11} direction. As is well known, however, elementary brane solutions have singularities at the location of the branes, needing to be supported by source terms. The natural candidates for those source terms, in our case, are the boundary actions. Given the anomaly-cancellation requirements, this restricts the possible solutions to those representing a pair of parallel three–branes corresponding to the orbifold planes.

From the above discussion, it is clear that in order to find a three-brane solution, we should start with the Ansatz

$$\begin{aligned} ds_5^2 &= a(y)^2 dx^\mu dx^\nu \eta_{\mu\nu} + b(y)^2 dy^2 \\ V &= V(y) \end{aligned} \tag{24}$$

where a and b are functions of $y = x^{11}$ and all other fields vanish. The general solution for this Ansatz, satisfying the equations of motion derived from action (16), is given by

$$\begin{aligned} a &= a_0 H^{1/2} \\ b &= b_0 H^2 \qquad\qquad H = -\frac{\sqrt{2}}{3}\alpha|y| + c_0 \\ V &= b_0 H^3 \end{aligned} \tag{25}$$

where a_0, b_0 and c_0 are constants. We note that the boundary source terms have fixed the form of the harmonic function H in the above solution. Without specific information about the sources, the function H would generically be glued together from an arbitrary number of linear pieces with slopes $\pm\frac{\sqrt{2}}{3}\alpha$. The edges of each piece would then indicate the location of the source terms. The necessity of matching the boundary sources at $y = 0$ and $\pi\rho$, however, has forced us to consider only two such linear pieces, namely $y \in [0, \pi\rho]$ and $y \in [-\pi\rho, 0]$. These pieces are glued together at $y = 0$ and $\pi\rho$ (recall here that we have identified $\pi\rho$ and $-\pi\rho$). Therefore, we have

$$\partial_y^2 H = -\frac{2\sqrt{2}}{3}\alpha(\delta(y) - \delta(y - \pi\rho)) \tag{26}$$

which shows that the solution represents two parallel three–branes located at the orbifold planes. We stress that this solution solves the five–dimensional theory (16) exactly, and is valid to all orders in κ.

Of course, we still have to check that our solution preserves half of the supersymmetries. When $g_{\alpha\beta}$ and V are the only non–zero fields, the supersymmetry transformations (23) simplify to

$$\begin{aligned} \delta\psi_\alpha^i &= D_\alpha \epsilon^i + \frac{\sqrt{2}}{12}\alpha\, \epsilon(y) V^{-1} \gamma_\alpha \, (\tau_3)^i{}_j \epsilon^j \\ \delta\zeta^i &= \frac{i}{2} V^{-1} \gamma_\beta \partial^\beta V \epsilon^i + \frac{i}{\sqrt{2}}\alpha\, \epsilon(y) V^{-1} \, (\tau_3)^i{}_j \epsilon^j \,. \end{aligned}$$

The Killing spinor equations $\delta\psi_\alpha^i = 0$, $\delta\zeta^i = 0$ are satisfied for the solution (25) if we require that the spinor ϵ^i is given by

$$\epsilon^i = H^{1/4}\epsilon_0^i \,, \quad \gamma_{11}\epsilon_0^i = (\tau_3)^i_j \epsilon_0^j \tag{27}$$

where ϵ_0^i is a constant symplectic Majorana spinor. This shows that we have indeed found a BPS solution preserving four of the eight bulk supercharges.

Let us discuss the meaning of this solution in some detail. First, we notice that it fits into the general scheme of domain wall solutions in various dimensions. It is, however, a new solution to the gauged supergravity action (16) in five dimensions which has not been constructed previously. In addition, its source terms are naturally provided by the boundary actions resulting from Hořava–Witten theory. Most importantly, it constitutes the fundamental vacuum solution of a phenomenologically relevant theory. The two parallel three–branes of the solution, separated by the bulk, are oriented in the four uncompactified space–time dimensions, and carry the physical low–energy gauge and matter fields. Therefore, from the low–energy point of view where the orbifold is not resolved the three–brane worldvolume is identified with four–dimensional space–time. In this sense the Universe lives on the worldvolume of a three–brane.

Thus far, we have limited the discussion to the universal hypermultiplet only, coupled to $N = 1$ five–dimensional gauged supergravity. This result can be extended in a straighforward fashion to include all the $(1,1)$ moduli of the Calabi–Yau threefold. We will not, however, explicitly include the $(2,1)$ sector as it is largely unaffected by the specific structure of Hořava–Witten theory. We now explain the generalized structure of the zero mode fields used in the reduction to five dimensions. We begin with the bulk space. Including the zero modes, the metric is given by

$$ds^2 = V^{-2/3} g_{\alpha\beta} dx^\alpha dx^\beta + g_{AB} dx^A dx^B \tag{28}$$

where g_{AB} is the metric of the Calabi–Yau space X. Its Kähler form is defined by $\omega_{a\bar{b}} = i g_{a\bar{b}}$ and can be expanded in terms of the harmonic $(1,1)$–forms ω_{iAB}, $i = 1, \cdots, h^{1,1}$ as

$$\omega_{AB} = a^i \omega_{iAB} \; . \tag{29}$$

The coefficients $a^i = a^i(x^\alpha)$ are the $(1,1)$ moduli of the Calabi–Yau space. The Calabi–Yau volume modulus $V = V(x^\alpha)$ is defined by

$$V = \frac{1}{v} \int_X \sqrt{{}^6 g} \tag{30}$$

where ${}^6 g$ is the determinant of the Calabi–Yau metric g_{AB} and v is defined in (9). The modulus V then measures the Calabi–Yau volume in units of v. The factor $V^{-2/3}$ in eq. (28) has been chosen such that the metric $g_{\alpha\beta}$ is the five–dimensional Einstein frame metric. Clearly V is not independent of the $(1,1)$ moduli a^i but it can be expressed as

$$V = \frac{1}{6}\mathcal{K}(a) \ , \quad \mathcal{K}(a) = d_{ijk}a^i a^j a^k \tag{31}$$

where $\mathcal{K}(a)$ is the Kähler potential and d_{ijk} are the Calabi–Yau intersection numbers.

Let us now turn to the zero modes of the antisymmetric tensor field. We have the potentials and field strengths,

$$\begin{aligned} C_{\alpha\beta\gamma} \ , &\quad G_{\alpha\beta\gamma\delta} \\ C_{\alpha AB} = \frac{1}{6}\mathcal{A}^i_\alpha \omega_{iAB} \ , &\quad G_{\alpha\beta AB} = \mathcal{F}^i_{\alpha\beta}\omega_{iAB} \\ C_{abc} = \frac{1}{6}\xi\omega_{abc} \ , &\quad G_{\alpha abc} = X_\alpha \omega_{abc} \ . \end{aligned} \tag{32}$$

The five–dimensional fields are therefore an antisymmetric tensor field $C_{\alpha\beta\gamma}$ with field strength $G_{\alpha\beta\gamma\delta}$, $h^{1,1}$ vector fields $\mathcal{A}^i_\alpha$ with field strengths $\mathcal{F}^i_{\alpha\beta}$ and a complex scalar ξ with field strength X_α that arises from the harmonic $(3,0)$ form denoted by ω_{abc}. In the bulk the relations between those fields and their field strengths are simply

$$\begin{aligned} G_{\alpha\beta\gamma\delta} &= 24\,\partial_{[\alpha}C_{\beta\gamma\delta]} \\ \mathcal{F}^i_{\alpha\beta} &= \partial_\alpha \mathcal{A}^i_\beta - \partial_\beta \mathcal{A}^i_\alpha \\ X_\alpha &= \partial_\alpha \xi \ . \end{aligned} \tag{33}$$

These relations, however, will receive corrections from the boundary controlled by the 11–dimensional Bianchi identity (4). We will derive the associated five–dimensional Bianchi identities later.

Next, we should set up the structure of the boundary fields. The starting point is the standard embedding of the spin connection in the first E_8 gauge group such that

$$\mathrm{tr}F^{(1)} \wedge F^{(1)} = \mathrm{tr}R \wedge R \ . \tag{34}$$

As a result, we have an E_6 gauge field $A^{(1)}_\alpha$ with field strength $F^{(1)}_{\mu\nu}$ on the first hyperplane and an E_8 gauge field $A^{(2)}_\mu$ with field strength $F^{(2)}_{\mu\nu}$ on the second hyperplane. In addition, there are $h^{1,1}$ gauge matter fields from the $(1,1)$ sector on the first plane. They are specified by

$$A^{(1)}_b = \bar{A}_b + \omega_{ib}{}^c T_{cp} C^{ip} \tag{35}$$

where $\bar{A}_b$ is the (embedded) spin connection. Furthermore, $p, q, r, \ldots = 1, \ldots, 27$ are indices in the fundamental $\mathbf{27}$ representation of E_6 and T_{ap} are the $(\mathbf{3}, \mathbf{27})$ generators of E_8 that arise in the decomposition under the subgroup $SU(3) \times E_6$. Their complex conjugate is denoted by T^{ap}. The C^{ip} are $h^{1,1}$ complex scalars in the **27** representation of E_6. Useful traces for these generators are $\mathrm{tr}(T_{ap}T^{bq}) = \delta_a^b \delta_p^q$ and $\mathrm{tr}(T_{ap}T_{bq}T_{cr}) = \omega_{abc} f_{pqr}$ where f_{pqr} is the totally symmetric tensor that projects out the singlet in $\mathbf{27}^3$.

So far, what we have considered is similar to a reduction of pure 11–dimensional supergravity on a Calabi–Yau space, as for example performed in Ref. 7, with the addition of gauge and gauge matter fields on the boundaries. An important difference arises, however, because the standard embedding (34), unlike in the case of the weakly coupled heterotic string, no longer leads to vanishing sources in the Bianchi identity (4). Instead, there is a net five-brane charge, with opposite sources on each fixed plane, proportional to $\pm \mathrm{tr} R \wedge R$. The nontrivial components of the Bianchi identity (4) are given by

$$(dG)_{11ABCD} = -\frac{1}{4\sqrt{2}\pi}\left(\frac{\kappa}{4\pi}\right)^{2/3} \{\delta(x^{11}) - \delta(x^{11} - \pi\rho)\} (\mathrm{tr} R \wedge R)_{ABCD} \,. \tag{36}$$

As a result, the components G_{ABCD} of the antisymmetric tensor field are nonvanishing. We find that

$$G_{ABCD} = -\frac{1}{4V}\alpha^i \, \epsilon_{ABCD}{}^{EF} \, \omega_{iEF} \, \epsilon(x^{11}) \tag{37}$$

where

$$\alpha_i = \frac{1}{8\sqrt{2}\pi}\left(\frac{\kappa}{4\pi}\right)^{2/3} \frac{1}{v^{2/3}} \int_{C_i} \mathrm{tr} R \wedge R \,. \tag{38}$$

Here, the four–cycles C_i are the Poincare duals of the harmonic $(1,1)$–forms ω_i. The index of the coefficient α^i in the second part of the first equation has been raised using the inverse of the metric

$$G_{ij}(a) = \frac{1}{2V} \int_X \omega_i \wedge (*\omega_j) \tag{39}$$

on the $(1,1)$ moduli space. Note that, while the coefficients α_i with lowered index are truly constants, as is apparent from eq. (38), the coefficients α^i depend on the $(1,1)$ moduli a^i since the metric (39) does. We can derive an expression for the boundary $\mathrm{tr} F^2$ and $\mathrm{tr} R^2$ terms in the action essential for the reduction of the boundary theories. We have

$$\mathrm{tr} R_{AB} R^{AB} = \mathrm{tr} F^{(1)}_{AB} F^{(1)AB} = 4\sqrt{2}\pi \left(\frac{4\pi}{\kappa}\right)^{2/3} V^{-1} \alpha^i \omega^{AB} \omega_{iAB} \tag{40}$$

while, of course

$$\mathrm{tr}F^{(2)}_{AB}F^{(2)AB} = 0\,. \tag{41}$$

The expression (37) for G_{ABCD} with α_i as defined in (38) is, as previously discussed, the new and somewhat unconventional ingredient in our reduction. This configuration for the antisymmetric tensor field strength is the generalized nonzero mode or G–flux. Generally, a nonzero mode is defined as a nonzero internal antisymmetric tensor field strength G that solves the equation of motion. In contrast, conventional zero modes of an antisymmetric tensor field, like those in eq. (33), have vanishing field strength once the moduli fields are set to constants. Since the kinetic term G^2 is positive for a nonzero mode it corresponds to a nonzero energy configuration. Given that nonzero modes, for a p–form field strength, satisfy

$$dG = d^*G = 0 \tag{42}$$

they correspond to harmonic forms of degree p. Hence, they can be identified with the pth cohomology group $H^p(X)$ of the internal manifold X. In the present case, we are dealing with a four–form field strength on a Calabi–Yau threefold X so that the relevant cohomology group is $H^4(X)$. The expression (37) is just an expansion of the nonzero mode in terms of the basis of $H^4(X)$. The appearance of all harmonic $(2,2)$ forms shows that it is necessary to include the complete $(1,1)$ sector into the low energy effective action in order to fully describe the nonzero mode. On the other hand, harmonic $(2,1)$ forms do not appear here and are, hence, less important in our context. We stress that the nonzero mode (37), for a given Calabi–Yau space, specifies a fixed element in $H^4(X)$ since the coefficients α_i are fixed in terms of Calabi–Yau properties. Thus we see that, correctly normalized, G is in the integer cohomology of the Calabi-Yau manifold. We emphasize that in heterotic M-theory, we are not free to turn off the non–zero mode. Its presence is simply dictated by the nonvanishing boundary sources.

Let us now summarize the field content which we have obtained above and discuss how it fits into the multiplets of five–dimensional $N = 1$ supergravity. We know that the gravitational multiplet should contain one vector field, the graviphoton. Thus, since the reduction leads to $h^{1,1}$ vectors, we must have $h^{1,1} - 1$ vector multiplets. This leaves us with the $h^{1,1}$ scalars a^i, the complex scalar ξ and the three-form $C_{\alpha\beta\gamma}$. Since there is one scalar in each vector multiplet, we are left with three unaccounted for real scalars (one from the set of a^i, and ξ) and the three-form. Together, these fields form the "universal hypermultiplet;" universal because it is present independently of the particular form of the Calabi-Yau manifold. From this, it is clear that it must be the overall volume breathing mode $V = \frac{1}{6}d_{ijk}a^i a^j a^k$ that is the additional scalar from the set of the a^i which enters the universal multiplet. The three-form may appear a little unusual, but recall that in five dimensions a three-form is dual to a scalar σ. Thus, the bosonic sector of the universal hypermultiplet consists of the four scalars $(V, \sigma, \xi, \bar{\xi})$, as presented previously.

The $h^{1,1}-1$ vector multiplet scalars are the remaining a^i. More properly, since the breathing mode V is already part of a hypermultiplet it should be first scaled out when defining the shape moduli

$$b^i = V^{-1/3}a^i \ . \tag{43}$$

Note that the $h^{1,1}$ moduli b^i represent only $h^{1,1}-1$ independent degrees of freedom as they satisfy the constraint

$$\mathcal{K}(b) \equiv d_{ijk}b^i b^j b^k = 6 \ . \tag{44}$$

The graviton and graviphoton of the gravity multiplet are given by

$$(g_{\alpha\beta}, \frac{2}{3}b_i \mathcal{A}^i_\alpha). \tag{45}$$

Therefore, in total, the five dimensional bulk theory contains a gravity multiplet, the universal hypermultiplet and $h^{1,1}-1$ vector multiplets. The inclusion of the $(2,1)$ sector of the Calabi–Yau space would lead to an additional $h^{2,1}$ set of hypermultiplets in the theory. Since they will not play a prominent rôle in our context they will not be explicitly included in the following.

On the boundary $M_4^{(1)}$ we have an E_6 gauge multiplet $(A_\mu^{(1)}, \chi^{(1)})$ and $h^{1,1}$ chiral multiplets (C^{ip}, η^{ip}) in the fundamental **27** representation of E_6. Here C^{ip} denote the complex scalars and η^{ip} the chiral fermions. The other boundary, $M_4^{(2)}$, carries an E_8 gauge multiplet $(A_\mu^{(2)}, \chi^{(2)})$ only. Inclusion of the $(2,1)$ sector would add $h^{2,1}$ chiral multiplets in the $\overline{\mathbf{27}}$ representation of E_6 to the field content of the boundary $M_4^{(1)}$. Any even bulk field will also survive on the boundary. Thus, in addition to the four–dimensional part of the metric, the scalars b^i together with $\mathcal{A}^i_{11}$, and V and σ survive on the boundaries. These pair into $h^{1,1}$ chiral muliplets.

We are now ready to derive the bosonic part of the five–dimensional effective action for the $(1,1)$ sector. Inserting the expressions for the various fields into the 11-dimensional supergravity action (1) and dropping higher derivative terms we find

$$S_5 = S_{\text{grav,vec}} + S_{\text{hyper}} + S_{\text{bound}} + S_{\text{matter}} \tag{46}$$

with

$$S_{\rm grav,vec} = -\frac{1}{2\kappa_5^2}\int_{M_5}\sqrt{-g}\,\Big[R + G_{ij}\partial_\alpha b^i\partial^\alpha b^j +$$

$$G_{ij}\mathcal{F}^i_{\alpha\beta}\mathcal{F}^{j\alpha\beta} + \frac{\sqrt{2}}{12}\epsilon^{\alpha\beta\gamma\delta\epsilon}d_{ijk}\mathcal{A}^i_\alpha\mathcal{F}^j_{\beta\gamma}\mathcal{F}^k_{\delta\epsilon}\Big] \tag{47a}$$

$$S_{\rm hyper} = -\frac{1}{2\kappa_5^2}\int_{M_5}\sqrt{-g}\,\Big[\frac{1}{2}V^{-2}\partial_\alpha V\partial^\alpha V$$

$$+\ 2V^{-1}X_\alpha\bar{X}^\alpha + \frac{1}{24}V^2 G_{\alpha\beta\gamma\delta}G^{\alpha\beta\gamma\delta}$$

$$+\ \frac{\sqrt{2}}{24}\epsilon^{\alpha\beta\gamma\delta\epsilon}G_{\alpha\beta\gamma\delta}\left(i(\xi\bar{X}_\epsilon - \bar{\xi}X_\epsilon) - 2\epsilon(x^{11})\alpha_i\mathcal{A}^i_\epsilon\right)$$

$$+\ \frac{1}{2}V^{-2}G^{ij}\alpha_i\alpha_j\Big] \tag{47b}$$

$$S_{\rm bound} = -\frac{\sqrt{2}}{\kappa_5^2}\int_{M_4^{(1)}}\sqrt{-g}V^{-1}\alpha_i b^i + \frac{\sqrt{2}}{\kappa_5^2}\int_{M_4^{(2)}}\sqrt{-g}\,V^{-1}\alpha_i b^i \tag{47c}$$

$$S_{\rm matter} = -\frac{1}{16\pi\alpha_{\rm GUT}}\sum_{n=1}^{2}\int_{M_4^{(n)}}\sqrt{-g}\,V{\rm tr}{F^{(n)}_{\mu\nu}}^2$$

$$-\ \frac{1}{2\pi\alpha_{\rm GUT}}\int_{M_4^{(1)}}\sqrt{-g}\,\Big[G_{ij}(D_\mu C)^i(D^\mu\bar{C})^j$$

$$+\,V^{-1}G^{ij}\frac{\partial W}{\partial C^{ip}}\frac{\partial\bar{W}}{\partial\bar{C}^j_p} + D^{(u)}D^{(u)}\Big]\ . \tag{47d}$$

All fields in this action that originate from the 11–dimensional antisymmetric tensor field are subject to a nontrivial Bianchi identity. Specifically, from eq. (4) we have

$$(dG)_{11\mu\nu\rho\sigma} = -\frac{2\sqrt{2}\pi\kappa_5^2}{\alpha_{\rm GUT}}\left\{J^{(1)}\delta(x^{11}) + J^{(2)}\delta(x^{11}-\pi\rho)\right\}_{\mu\nu\rho\sigma} \tag{48a}$$

$$(d\mathcal{F}^i)_{11\mu\nu} = -\frac{\kappa_5^2}{4\sqrt{2}\pi\alpha_{\rm GUT}}J^i_{\mu\nu}\delta(x^{11}) \tag{48b}$$

$$(dX)_{11\mu} = -\frac{\kappa_5^2}{4\sqrt{2}\pi\alpha_{\rm GUT}}J_\mu\delta(x^{11}) \tag{48c}$$

with the currents defined by

$$J^{(n)}_{\mu\nu\rho\sigma} = \frac{1}{16\pi^2}\left(\mathrm{tr}F^{(n)}\wedge F^{(n)} - \frac{1}{2}\mathrm{tr}R\wedge R\right)_{\mu\nu\rho\sigma} \tag{49a}$$

$$J^i_{\mu\nu} = -2iV^{-1}\Gamma^i_{jk}\left((D_\mu C)^{jp}(D_\nu \bar{C})^k_p - (D_\mu \bar{C})^k_p(D_\nu C)^{jp}\right) \tag{49b}$$

$$J_\mu = -\frac{i}{2}V^{-1}d_{ijk}f_{pqr}(D_\mu C)^{ip}C^{jq}C^{kr}\ . \tag{49c}$$

The five-dimensional Newton constant κ_5 and the Yang–Mills coupling α_{GUT} are expressed in terms of 11-dimensional quantities as

$$\kappa_5^2 = \frac{\kappa^2}{v}\ , \qquad \alpha_{\mathrm{GUT}} = \frac{\kappa^2}{2v}\left(\frac{4\pi}{\kappa}\right)^{2/3}\ . \tag{50}$$

We still need to define various quantities in the above action. The metric G_{ij} is given in terms of the Kähler potential $\mathcal{K}$ as

$$G_{ij} = -\frac{1}{2}\frac{\partial}{\partial b^i}\frac{\partial}{\partial b^j}\ln\mathcal{K}\ . \tag{51}$$

The corresponding connection Γ^i_{jk} is defined as

$$\Gamma^i_{jk} = \frac{1}{2}G^{il}\frac{\partial G_{jk}}{\partial b^l}\ . \tag{52}$$

We recall that

$$\mathcal{K} = d_{ijk}b^ib^jb^k\ , \tag{53}$$

where d_{ijk} are the Calabi–Yau intersection numbers. All indices $i, j, k, \cdots$ in the five-dimensional theory are raised and lowered with the metric G_{ij}. We recall that the fields b^i are subject to the constraint

$$\mathcal{K} = 6 \tag{54}$$

which should be taken into account when equations of motion are derived from the above action. Most conveniently, it can be implemented by adding a Lagrange multiplier term $\sqrt{-g}\lambda(\mathcal{K}(b) - 6)$ to the bulk action. Furthermore, we need to define the superpotential

$$W = \frac{1}{6}d_{ijk}f_{pqr}C^{ip}C^{jq}C^{kr} \tag{55}$$

and the D-term

$$D^{(u)} = G_{ij}\bar{C}^jT^{(u)}C^i \tag{56}$$

where $T^{(u)}$, $u = 1, \ldots, 78$ are the E_6 generators in the fundamental representation. The consistency of the above theory has been explicitly checked by a reduction of the 11-dimensional equations of motion.

The most notable features of this action, at first sight, are the bulk and boundary potentials for the $(1,1)$ moduli V and b^i that appear in $S_{\rm hyper}$ and $S_{\rm bound}$. Those potentials involve the five–brane charges α_i, defined by eq. (38), that characterize the nonzero mode. The bulk potential in the hypermultiplet part of the action arises directly from the kinetic term G^2 of the antisymmetric tensor field with the expression (37) for the nonzero mode inserted. It can therefore be interpreted as the energy contribution of the nonzero mode. The origin of the boundary potentials, on the other hand, can be directly seen from eq. (40) and the 10-dimensional boundary actions. Essentially, they arise because the standard embedding leads to nonvanishing internal boundary actions due to the crucial factor $1/2$ in front of the ${\rm tr}R^2$ terms. This is in complete analogy with the appearance of nonvanishing sources in the internal part of the Bianchi identity which led us to introduce the nonzero mode. The action presented in (46) and (47b) was first derived in Ref. 4.

3. Bulk Five-Branes and Non-Standard Embeddings

In this third lecture, we begin by discussing the simplest BPS three-brane solution of the generalized five-dimensional heterotic M-theory presented in Lecture 2. We then commence a major extension of heterotic M-theory. Until now, we have employed the standard embedding of the spin connection of the Calabi-Yau threefold into the gauge connection of the visible brane. However, unlike the case of the weakly coupled heterotic string, there is nothing compelling about the standard embedding in heterotic M-theory. Quite the contrary, it is more natural to consider "non-standard" embeddings. Here, we will only briefly discuss such embeddings, referring the reader to the TASI 2001 lectures by Daniel Waldram for details. In this lecture, we focus on one of the important phenomena assoaciated with non-standard embeddings, namely, the appearance of one or more bulk space three-branes (actually, $M5$-branes wrapped on holomorphic curves in the Calabi-Yau threefold). In the second part of this lecture, we will discuss the existence and properties of bulk space wrapped five-branes in detail.

We would now like to find the simplest BPS domain wall solutions of the generalized five–dimensional heterotic M-theory. ¿From the above results, it is clear that the proper Ansatz for the type of solutions we are looking for is given by

$$\begin{aligned} ds_5^2 &= a(y)^2 dx^\mu dx^\nu \eta_{\mu\nu} + b(y)^2 dy^2 \\ V &= V(y) \\ b^i &= b^i(y) \ , \end{aligned} \tag{57}$$

where we use $y = x^{11}$ from now on. A solution to the generalized equations of motion is somewhat hard to find, essentially due to the complication caused by the inclusion of all $(1,1)$ moduli and the associated Kähler structure. The trick is to express the solution in terms of certain functions $f^i = f^i(y)$ which are only implicitly defined rather than trying to find fully explicit formulae. It turns out that those functions are fixed by the equations

$$d_{ijk} f^j f^k = H_i \ , \quad H_i = -2\sqrt{2} k \alpha_i |y| + k_i \tag{58}$$

where k and k_i are arbitrary constants. Then the solution can be written as

$$\begin{aligned} V &= \left(\frac{1}{6} d_{ijk} f^i f^j f^k\right)^2 \\ a &= \tilde{k} V^{1/6} \\ b &= k V^{2/3} . \\ b^i &= V^{-1/6} f^i \end{aligned} \tag{59}$$

where $\tilde{k}$ is another arbitrary constant. We have checked that this solution is indeed a BPS state of the theory; that is, that it preserves four of the eight supercharges. Note that we have chosen the above solution to have no singularities other than those at the two boundaries. Specifically, the harmonic functions H_i in eq. (58) satisfy

$$H_i{}'' = 4\sqrt{2} k \alpha_i (\delta(y) - \delta(y - \pi\rho)) \ , \tag{60}$$

indicating sources at the orbifold planes $y = 0, \pi\rho$. Recall that we have restricted the range of y to $y \in [-\pi\rho, \pi\rho]$ with the endpoints identified. This explains the second delta–function at $y = \pi\rho$ in the above equation. We conclude that the solution (60) represents a multi–charged double domain wall (three–brane) solution with the two walls located at the orbifold planes. It preserves four–dimensional Poincaré invariance as well as four of the eight supercharges.

In Lecture 2, we have presented a related three–brane solution which was less general in that it involved the universal Calabi–Yau modulus V only. Clearly, we should be able to recover this solution from eq. (60) if we consider the specific case $h^{1,1} = 1$. Then we have $d_{111} = 6$ and it follows from eq. (58) that

$$f^1 = \left(\frac{\sqrt{2}}{3} k \alpha_1 |y| + k_1\right)^{1/2} . \tag{61}$$

Inserting this into eq. (60) provides us with the explicit solution in this case which is given by

$$\begin{aligned} a &= a_0 H^{1/2} \\ b &= b_0 H^2 \qquad\qquad H = -\frac{\sqrt{2}}{3}\alpha|y| + c_0 \ , \quad \alpha = \alpha^1 \\ V &= b_0 H^3 \ . \end{aligned} \tag{62}$$

The constant a_0, b_0 and c_0 are related to the integration constants in eq. (60) by

$$a_0 = \tilde{k}k^{1/2} \ , \qquad b_0 = k^3 \ , \qquad c_0 = \frac{k_1}{k} \ . \tag{63}$$

Eq. (62) is indeed exactly the solution that was found in Lecture 2. It still represents a double domain wall. However, in contrast to the general solution it couples to one charge $\alpha = \alpha^1$ only. Geometrically, it describes a variation of the five–dimensional metric and the Calabi–Yau volume across the orbifold.

At this point, we introduce an important generalization which greatly expands the scope, theoretical interest and phenomenological implications of heterotic M-theory. First, note that all of our previous results have assumed that the gauge field vacuum on the Calabi–Yau threefold is identical to the geometrical spin connection. That is, we have assumed the standard embedding defined in (6). Since any Calabi–Yau threefold has holonomy group $SU(3)$, it follows that the spin connection and, hence, the gauge connection has structure group $G = SU(3)$. The four-dimensional low energy theory then exhibits a gauge group H which is the commutant of G in E_8. Since $G = SU(3)$, it follows that $H = E_6$, as we discussed above. Although the choice of the standard embedding was natural within the context of weakly coupled heterotic superstring theory, there is no reason to single it out from other gauge vacua in M-theory. Indeed, the only constraint on the gauge vacua in heterotic M-theory is that they be compatible with $N = 1$ supersymmetry on the boundary planes. That is, the gauge connection on the Calabi–Yau threefold must satisfy the Hermitian Yang–Mills equation, but is otherwise arbitrary. Clearly, it would be of significant interest to demonstrate the existence of gauge vacua other than the standard embedding. For example, if one could construct a "non–standard" embedding gauge vacuum with structure group, say, $G = SU(5) \times \mathbf{Z}_2$, then the low energy gauge group in four-dimensions would be the standard model group $SU(3)_C \times SU(2)_L \times U(1)_Y$. Since the Calabi–Yau threefold has a

Euclidean signature and is compact, we will refer to any gauge configuration with structure group $G \subset E_8$ that satisfies the Hermitian Yang–Mills equation as a G-instanton. We will, therefore, expand the vacua of heterotic M-theory by compactifying Hořava-Witten theory on Calabi–Yau manifolds with G-instantons.

Initially, this seems to be a very difficult task, since not a single solution to the Hermitian Yang–Mills equations on a Calabi–Yau threefold is known, with the exception of the standard embedding. However, at this point, some important mathemetical results become relevant, which allow us to demonstrate the existence and compute the properties of very large classes of G-instantons. The fundamental results in this regard are two–fold. First, it was shown by Donaldson and Uhlenbeck and Yau that there is a one-to-one correspondence between any G-instanton solution of the Hermitian Yang–Mills equation and the existence of a stable holomorphic vector bundle with structure group G over the Calabi–Yau threefold. Given one the other is determined, at least in principle. Now, even though it appears to be very difficult to find solutions of the Hermitian Yang–Mills equations, it was demonstrated by Friedman, Morgan and Witten[33,35] and Donagi[34] that one can, rather straightforwardly, construct stable holomorphic vector bundles over Calabi–Yau threefolds. Using, and extending, the technology introduced in these papers, large classes of heterotic M-theory vacua with non-standard G-bundles have been constructed.[36,37] It was shown in these papers that heterotic M-theory vacua corresponding to grand unified theories, with gauge groups such as $SU(5)$ and $SO(10)$,[36] and the standard model with gauge group $SU(3)_C \times SU(2)_L \times U(1)_Y$[37] can, indeed, be constructed in this manner. I will not discuss these holomorphic bundle constructions in these lectures, referring the reader to the TASI lectures by Daniel Waldram. Here, instead, I will discuss an important implication of non-standard G-bundle vacua, namely, the necessary appearance of $M5$-branes, wrapped on holomorphic curves, in the bulk space.

Recall from above that anomaly cancellation requires that the Bianchi identity for the four-form field strength $G = dC$ be modified as in equation (4). It is useful to rewrite this expression as

$$(dG)_{11\bar{I}\bar{J}\bar{K}\bar{L}} = -4\sqrt{2}\pi \left(\frac{\kappa}{4\pi}\right)^{2/3} \left\{J^{(1)}\delta(x^{11}) + J^{(2)}\delta(x^{11} - \pi\rho)\right\}_{\bar{I}\bar{J}\bar{K}\bar{L}} \quad (64)$$

where sources are defined by

$$J^{(n)} = c_2(V^n) - \frac{1}{2}c_2(TX) \qquad n = 1, 2, \quad (65)$$

and

$$c_2(V^n) = -\frac{1}{16\pi^2} trF^n \wedge F^n, \qquad c_2(TX) = -\frac{1}{16\pi^2} trR \wedge R, \tag{66}$$

V^n is the stable holomorphic vector bundle on the n-th plane, F^n is the field strength associated with the gauge theory, and R is the Ricci tensor of the Calabi-Yau manifold. Note that $c_2(V^n)$ and $c_2(TX)$ are the second Chern class of the vector bundle on the n-th boundary plane and the second Chern class of the Calabi-Yau tangent bundle respectively. Integrating (64) over a five-cycle which spans the orbifold interval and is otherwise an arbitrary four-cycle in the Calabi-Yau three-fold, we find the topological condition that

$$c_2(V^1) + c_2(V^2) - c_2(TX) = 0. \tag{67}$$

When N bulk five-branes, located at coordinates x_i for $i = 1, \ldots, N$ in the 11-direction, are present in the vacuum, cancellation of their world-volume anomalies, as well as the gravitational and gauge anomalies on the orbifold fixed planes, requires that Bianchi identity be further modified to

$$\begin{aligned}(dG)_{11\bar{I}\bar{J}\bar{K}\bar{L}} = 4\sqrt{2}\pi \left(\frac{\kappa}{4\pi}\right)^{2/3} & (J^{(1)}\delta(x^{11}) + J^{(2)}\delta(x^{11} - \pi\rho) \\ & + \Sigma_{i=1}^{N} \hat{J}^{(i)}\delta(x^{11} - x_i))_{\bar{I}\bar{J}\bar{K}\bar{L}}.\end{aligned} \tag{68}$$

Each five-brane source $\hat{J}^{(i)}$ is defined to be the four-form which is Poincaré dual to the holomorphic curve in the Calabi-Yau threefold around which the i-th five-brane is wrapped. If we define the five-brane class

$$W = \Sigma_{i=1}^{N} \hat{J}^{(i)}, \tag{69}$$

then the topological condition (67) is modified to

$$c_2(V^1) + c_2(V^2) - c_2(TX) + W = 0. \tag{70}$$

The simplest example one can present is the standard embedding, where one fixes the Calabi-Yau three-fold and chooses the two holomorphic vector bundles so that $V^1 = TX$ and $V^2 = 0$. It follows that

$$c_2(V^1) = c_2(TX), \qquad c_2(V^2) = 0. \tag{71}$$

Note that these Chern classes satisfy the topological condition given in (70) with

$$W = 0. \tag{72}$$

That is, for the standard embedding there are no $M5$–branes in the bulk space, as we already know from the previous lectures. However, as was shown in Ref. 36, most non-standard G-bundles correspond to Chern classes that require a non-vanishing five-brane class W in order to be anomaly free. In particular, phenomenologically relevant heterotic M-theory vacua, such as those leading to the standard model gauge group with three families of quarks and leptons,[37] must have bulk five–branes. We will, therefore, spend the remainder of this lecture discussing the structure and physical properties of bulk space $M5$–branes wrapped on holomorphic curves.

The inclusion of five-branes in the bulk space not only generalizes the types of background one can consider, but also introduces new degrees of freedom into the theory, namely, the dynamical fields on the five-branes themselves. We will now consider what low-energy fields survive on one of the five-branes when it is wrapped around a two-cycle in the Calabi–Yau threefold.

In general, the fields on a single five-brane are as follows.[38,39] The simplest are the bosonic coordinates X^I describing the embedding of the brane into 11-dimensional spacetime. The additional bosonic field is a world-volume two-form potential B with field strength $H = dB$ satisfying a generalized self-duality condition. For small fluctuations, the duality condition simplifies to the conventional constraint $H = *H$. These degrees of freedom are paired with spacetime fermions θ, leading to a Green–Schwarz type action, with manifest spacetime supersymmetry and local kappa-symmetry.[40,41] (As usual, including the self-dual field in the action is difficult, but is made possible by either including an auxiliary field or abandoning a covariant formulation.) For a five-brane in flat space, one can choose a gauge such that the dynamical fields fall into a six-dimensional massless tensor multiplet with $(0,2)$ supersymmetry on the brane world-volume.[42,43] This multiplet has five scalars describing the motion in directions transverse to the five-brane, together with the self-dual tensor H.

For a five-brane embedded in $S^1/Z_2 \times X \times M_4$, to preserve Lorentz invariance in M_4, $3+1$ dimensions of the five-brane must be left uncompactified. The remaining two spatial dimensions are then wrapped on a two-cycle of the Calabi–Yau three-fold. To preserve supersymmetry, the two-cycle must be a holomorphic curve.[17,44,45] Thus, from the point of view of a five-dimensional effective theory on $S^1/Z_2 \times M_4$, since two of the five-brane directions are compactified, it appears as a flat three-brane (or equivalently a domain wall) located at some point $x^{11} = x$ on the orbifold. Thus, at low energy, the degrees of freedom on the brane must fall into four-dimensional supersymmetric multiplets.

An important question is how much supersymmetry is preserved in the low-energy theory. One way to address this problem is directly from the symmetries of the Green–Schwarz action, following the discussion for similar brane configurations in Ref. 44. Locally, the 11-dimensional spacetime $S^1/Z_2 \times X \times M_4$ admits eight independent Killing spinors η, so should be described by a theory with eight supercharges. (Globally, only half of the spinors survive the non-local orbifold quotienting condition $\Gamma_{11}\eta(-x^{11}) = \eta(x^{11})$, so that, for instance, the eleven-dimensional bulk fields lead to $N = 1$, not $N = 2$, supergravity in four dimensions.) The Green–Schwarz form of the five-brane action is then invariant under supertranslations generated by η, as well as local kappa-transformations. In general the fermion fields θ transform as (see for instance Ref. 43)

$$\delta\theta = \eta + P_+\kappa \tag{73}$$

where P_+ is a projection operator. If the brane configuration is purely bosonic then $\theta = 0$ and the variation of the bosonic fields is identically zero. Furthermore, if $H = 0$ then the projection operator takes the simple form

$$P_\pm = \frac{1}{2}\left(1 \pm \frac{1}{6!\sqrt{g}}\epsilon^{m_1\ldots m_6}\partial_{m_1}X^{I_1}\ldots\partial_{m_6}X^{I_6}\Gamma_{I_1\ldots I_6}\right) \tag{74}$$

where σ^m, $m = 0,\ldots,5$ label the coordinates on the five-brane and g is the determinant of the induced metric

$$g_{mn} = \partial_m X^I \partial_n X^J g_{IJ}\ . \tag{75}$$

If the brane configuration is invariant for some combination of supertranslation η and kappa-transformation, then we say it is supersymmetric. Now κ is a local parameter which can be chosen at will. Since the projection operators satisfy $P_+ + P_- = 1$, we see that for a solution of $\delta\theta = 0$, one is required to set $\kappa = -\eta$, together with imposing the condition

$$P_-\eta = 0 \tag{76}$$

For a brane wrapped on a two-cycle in the Calabi–Yau space, spanning M_4 and located at $x^{11} = x$ in the orbifold interval, we can choose the parameterization

$$X^\mu = \sigma^\mu, \qquad X^A = X^A(\sigma,\bar\sigma), \qquad X^{11} = x \tag{77}$$

where $\sigma = \sigma^4 + i\sigma^5$. The condition (76) then reads

$$-(i/\sqrt{g})\,\partial X^A\bar\partial X^B\Gamma^{(4)}\Gamma_{AB}\,\eta = \eta \tag{78}$$

where we have introduced the four-dimensional chirality operator $\Gamma^{(4)} = \Gamma_0\Gamma_1\Gamma_2\Gamma_3$. Recalling that on the Calabi–Yau three-fold the Killing spinor satisfies $\Gamma^{\bar{b}}\eta = 0$, it is easy to show that this condition can only be satisfied if the embedding is holomorphic, that is $X^a = X^a(\sigma)$, independent of $\bar{\sigma}$. The condition then further reduces to

$$\Gamma^{(4)}\eta = i\eta \tag{79}$$

which, given that the spinor has definite chirality in eleven dimensions as well as on the Calabi–Yau space, implies that $\Gamma^{11}\eta = \eta$, compatible with the global orbifold quotient condition. Thus, finally, we see that only half of the eight Killing spinors, namely those satisfying (79), lead to preserved supersymmetries on the five-brane. Consequently the low-energy four-dimensional theory describing the five-brane dynamics will have $N = 1$ supersymmetry.

The simplest excitations on the five-brane surviving in the low-energy four-dimensional effective theory are the moduli describing the position of the five-brane in eleven dimensions. There is a single modulus X^{11} giving the position of the brane in the orbifold interval. In addition, there is the moduli space of holomorphic curves $\mathcal{C}_2$ in X describing the position of the brane in the Calabi–Yau space. This moduli space is generally complicated, and we will not address its detailed structure here. (As an example, the moduli space of genus one curves in K3 is K3 itself.[45]) However, we note that these moduli are scalars in four dimensions, and we expect them to arrange themselves as a set of chiral multiplets, with a complex structure presumably inherited from that of the Calabi–Yau manifold.

Now let us consider the reduction of the self-dual three-form degrees of freedom. (Here we are essentially repeating a discussion given in Refs. 46, 47). The holomorphic curve is a Riemann surface and, so, is characterized by its genus g. One recalls that the number of independent harmonic one-forms on a Riemann surface is given by $2g$. In addition, there is the harmonic volume two-form Ω. Thus, if we decompose the five-brane worldvolume as $\mathcal{C}_2 \times M_4$, we can expand H in zero modes as

$$H = da \wedge \Omega + F^u \wedge \lambda_u + h \tag{80}$$

where λ_u are a basis $u = 1, \ldots, 2g$ of harmonic one-forms on $\mathcal{C}_2$, while the four-dimensional fields are a scalar a, $2g$ $U(1)$ vector fields $F^u = dA^u$ and a three-form field strength $h = db$. However, not all these fields are independent because of the self-duality condition $H = *H$. Rather, one easily concludes that

$$h = *da \tag{81}$$

and, hence, that the four-dimensional scalar a and two-form b describe the same degree of freedom. To analyze the vector fields, we introduce the matrix $T_u{}^v$ defined by

$$*\lambda_u = T_u{}^v \lambda_v \tag{82}$$

If we choose the basis λ_u such that the moduli space metric $\int_{\mathcal{C}_2} \lambda_u \wedge (*\lambda_v)$ is the unit matrix, T is antisymmetric and, of course, $T^2 = -1$. The self-duality constraint implies for the vector fields that

$$F^u = T_v{}^u * F^v \ . \tag{83}$$

If we choose a basis for F^u such that

$$T = \text{diag}\left(\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \ldots, \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}\right) \tag{84}$$

with g two by two blocks on the diagonal, one easily concludes that only g of the $2g$ vector fields are independent. In conclusion, for a genus g curve $\mathcal{C}_2$, we have found one scalar and g $U(1)$ vector fields from the two-form on the five-brane worldvolume. The scalar has to pair with another scalar to form a chiral $N = 1$ multiplet. The only other universal scalar available is the zero mode of the transverse coordinate X^{11} in the orbifold direction.

Thus, in general, the $N = 1$ low-energy theory of a single five-brane wrapped on a genus g holomorphic curve $\mathcal{C}_2$ has gauge group $U(1)^g$ with g $U(1)$ vector multiplets and a universal chiral multiplet with bosonic fields (a, X^{11}). Furthermore, there is some number of additional chiral multiplets describing the moduli space of the curve $\mathcal{C}_2$ in the Calabi–Yau three-fold.

It is well known that when two regions of the five-brane world-volume in M–theory come into close proximity, new massless states appear.[48,13] These are associated with membranes stretching between the two nearly overlapping five-brane surfaces. In general, this can lead to enhancement of the gauge symmetry. Let us now consider this possibility, heretofore ignored in our discussion. In general, one can consider two types of brane degeneracy where parts of the five-brane world-volumes are in close proximity. The first, and simplest, is to have N distinct but coincident five-branes, all wrapping the same cycle $\mathcal{C}_2$ in the Calabi–Yau space and all located at the same point in the orbifold interval. Here, the new massless states come from membranes stretching between the distinct five-brane world-volumes. The second, and more complicated, situation is where there is a degeneracy of the embedding of a single five-brane, such that parts of the curve C_2 become close together in the Calabi–Yau space. In this case, the new states come from membranes stretching between different parts of the same five-brane world-volume.[49,50] Let us consider these two possibilities separately.

The first case of distinct five-branes is analogous to the M–theory description of N overlapping type IIB D3-branes, which arise as N coincident five-branes wrapping the same cycle in a flat torus. In that case, the $U(1)$ gauge theory on each D3-brane is enhanced to a $U(N)$ theory describing the full collection of branes. Thus, by analogy, in our case we would expect a similar enhancement of each of the g $U(1)$ fields on each five-brane. That is, when wrapped on a holomorphic curve of genus g, the full gauge group for the low-energy theory describing N coincident five-branes becomes $U(N)^g$.

The second case is inherently more complicated. It can, however, be clearly elucidated and studied for Calabi-Yau threefolds which are elliptically fibered. These manifolds consist of a base two-fold, over any point of which is fibered an elliptic curve. At almost all points in the base, the elliptic curve is smooth. However, there is a locus of points, called the discriminant locus, over which the fibers degenerate. These degeneracies have specific characteristics and have been classified by Kodaira.[51] If the five-brane is wrapped over a smooth fiber, away from the discriminant locus, then there are no new massless states. However, as the fiber approaches the discriminant it degenerates to a specific Kodaira singularity. Accordingly, the five-brane wrapped on such a fiber begins to "approach itself" near the singularity, leading to new, massless states appearing in the theory. The general theory for computing these massless states was presented for fibers over both the smooth and singular parts of discriminant curves in Ref. 49 and 50 respectively. For example, consider an elliptically fibered Calabi-Yau threefold over an $\mathbf{F}_3$ Hirzebruch base and let the five-brane be wrapped on a fiber near a smooth part of the discriminant curve with Kodaira type I_2. Then, it was shown in Ref. 49 that, in addition to the usual states, the I_2 degeneracy of the elliptic fiber produces an $SU(2)$ doublet $\mathbf{2}$ of massless $N = 2$ hypermultiplets with unit electric charge. In general, one gets a complicated spectrum of new hypermultiplets and, for sufficiently intricate Kodaira singularities, new non-Abelian vector multiplets as well.

Summarizing the two cases, we see that for N five-branes wrapping the same curve C_2 of genus g, we expect that the symmetry is enhanced from N copies of $U(1)^g$ to $U(N)^g$. Alternatively, in the second case, even for a single brane we can get new massless states if the holomorphic curve degenerates. These states form hypermultiplets and entended non-Abelian gauge vector multiplets depending on the exact form of the curve degeneracy.

4. Beyond Hořava-Witten Theory

It is of interest to ask whether one can construct other orbifolds of M-theory beyond the $S^1/\mathbf{Z}_2$ example of Refs. 1, 2. A first step in this direction was taken by Dasgupta and Muhki[52] and Witten[53] who discussed both local and global anomaly cancellation within the context of $T^4/\mathbf{Z}_2$ orbifolds. A major generalization of these results was presented in Refs. 54, 55, 56, 57 and 58, 59 where all the M-theory orbifolds associated with the spacetime $\mathbf{R}^6 \times K3$ were constructed. In this fourth lecture, we will, for specificity, consider M-theory orbifolds on $S^1/\mathbf{Z}_2 \times T^4/\mathbf{Z}_2$. It will be demonstrated, in detail, how such orbifolds can be made anaomaly free, completely determining both the twisted and untwisted sector spectra in the process, even on odd dimensional orbifold planes where all anomalies vanish.

The spacetime has topology $\mathbf{R}^6 \times S^1 \times T^4$, where each of the five compact coordinates takes values on the interval $[-\pi, \pi]$ with the endpoints identified. Let x^μ parameterize the six non-compact dimensions, while x^i and x^{11} parameterize the T^4 and S^1 factors respectively. Then the $\mathbf{Z}_2$ action on S^1 is defined by

$$\alpha : (x^\mu, x^i, x^{11}) \longrightarrow (x^\mu, x^i, -x^{11}) \tag{85}$$

whereas the $\mathbf{Z}_2$ action on T^4 is

$$\beta : (x^\mu, x^i, x^{11}) \longrightarrow (x^\mu, -x^i, x^{11}) \,. \tag{86}$$

The element α leaves invariant the two ten-planes defined by $x^{11} = 0$ and $x^{11} = \pi$, while β leaves invariant the sixteen seven-planes defined when the four coordinates x^i individually assume the values 0 or π. Finally, $\alpha\beta$ leaves invariant the thirty-two six-planes defined when all five compact coordinates individually assume the values 0 or π. The $\alpha\beta$ six-planes coincide with the intersections of the α ten-planes with the β seven-planes.

A gravitational anomaly arises on each ten-plane due to the coupling of chiral projections of the bulk gravitino to currents localized on the fixed planes. Since the two ten-planes are indistinguishable aside from their position, this anomaly is identical on each of the two planes and can be computed by conventional means if proper care is used. The reason why extra care is needed is that each ten-plane anomaly arises from the coupling of eleven-dimensional fermions to ten-dimensional currents, whereas standard index theorem results only apply to ten-dimensional fermions coupled to ten-dimensional currents. If one notes that the index theorem can be applied to the small radius limit where the two ten-planes coincide, then the gravitational anomaly on each individual ten-plane can be computed;

it is simply one-half of the index theorem anomaly derived using the "untwisted" sector spectrum in ten-dimensions. By untwisted sector, we mean the $\mathbf{Z}_2$ projection of the eleven-dimensional bulk space supergravity multiplet onto each ten-dimension fixed plane. This untwisted spectrum forms the ten-dimensional $N = 1$ supergravity multiplet containing a graviton, a chiral gravitino, a two-form and a scalar dilaton. We denote by R the ten-dimensional Riemann tensor, regarded as an $SO(9,1)$-valued form.

As pointed out in Refs. 1, 2, in addition to the untwisted spectrum, one must allow for the possibility of "twisted" sector $N = 1$ supermultiplets that live on each ten-dimensional orbifold plane only. For the case at hand, the twisted sector spectrum must fall into $N = 1$ Yang-Mills supermultiplets consisting of gauge fields and chiral gauginos. These will give rise to an additional contribution to the gravitational anomaly on each ten-plane, as well as to mixed and pure-gauge anomalies. However, since the twisted sector fields are ten-dimensional, these anomalies can be computed directly from the standard formulas, without multiplying by one-half. The twisted sector gauge group, the dimension of the gauge group and the gauge field strength on the i-th ten-plane are denoted by $\mathcal{G}_i$, $n_i = \dim \mathcal{G}_i$ and F_i respectively, for $i = 1, 2$.

The quantum mechanical one-loop local chiral anomaly on the i-th ten-plane is characterized by the twelve-form

$$\begin{aligned} I_{12}(1-\text{loop})_i = {} & \frac{1}{4}\left(I^{(3/2)}_{GRAV}(R) - I^{(1/2)}_{GRAV}(R)\right) \\ & + \frac{1}{2}\left(n_i\, I^{(1/2)}_{GRAV}(R) + I^{(1/2)}_{MIXED}(R, F_i) + I^{(1/2)}_{GAUGE}(F_i)\right) \end{aligned} \tag{87}$$

from which the anomaly arises by descent. The constituent polynomials contributing to the pure gravitational anomaly due to the chiral spin 3/2 and chiral spin 1/2 fermions are

$$I^{(3/2)}_{GRAV}(R) = \frac{1}{(2\pi)^5 6!}\left(\frac{55}{56}\operatorname{tr} R^6 - \frac{75}{128}\operatorname{tr} R^4 \wedge \operatorname{tr} R^2 + \frac{35}{512}(\operatorname{tr} R^2)^3\right) \tag{88}$$

and

$$I^{(1/2)}_{GRAV}(R) = \frac{1}{(2\pi)^5 6!}\left(-\frac{1}{504}\operatorname{tr} R^6 - \frac{1}{384}\operatorname{tr} R^4 \wedge \operatorname{tr} R^2 - \frac{5}{4608}(\operatorname{tr} R^2)^3\right) \tag{89}$$

respectively, where tr is the trace of the $SO(9,1)$ indices. The polynomials contributing to the mixed and pure-gauge anomalies are due to chiral spin

1/2 fermions only and are given by

$$I^{(1/2)}_{MIXED}(R,F_i) = \frac{1}{(2\pi)^5 6!}\left(\frac{1}{16}\,\mathrm{tr}\,R^4\wedge \mathrm{Tr}\,F_i^2 + \frac{5}{64}\,(\,\mathrm{tr}\,R^2)^2\wedge \mathrm{Tr}\,F_i^2 \right.$$
$$\left. -\frac{5}{8}\,\mathrm{tr}\,R^2\wedge \mathrm{Tr}\,F_i^4\right) \tag{90}$$

and

$$I^{(1/2)}_{GAUGE}(F_i) = \frac{1}{(2\pi)^5 6!}\,\mathrm{Tr}\,F_i^6\,. \tag{91}$$

Here Tr is the trace over the adjoint representation of $\mathcal{G}_i$. All the anomaly polynomials are computed using standard index theorems. Each term in (87) has a factor of 1/2 because the relevant fermions are Majorana-Weyl with half the degrees of freedom of Weyl fermions. The first two terms in (87) arise from untwisted sector fermions, whereas the last three terms are contributed by the twisted sector. It follows from the above discussion that the first two terms must have an additional factor of 1/2, accounting for the overall coefficient of 1/4, whereas the remaining three terms are given exactly by the index theorems.

The quantum anomaly (87) would spoil the consistency of the theory were it not to cancel against some sort of classical inflow anomaly. Hence, it is imperative to discern the presence of appropriate local classical countert-erms to cancel against (87). One begins the analysis of anomaly cancellation by considering the pure $\mathrm{tr}\,R^6$ term in (87) which is irreducible and must therefore identically vanish. It follows from the above that this term is

$$-\frac{1}{2(2\pi)^5 6!}\,\frac{(n_i-248)}{494}\,\mathrm{tr}\,R^6\,. \tag{92}$$

Therefore, the $\mathrm{tr}\,R^6$ term will vanish if and only if each gauge group $\mathcal{G}_i$ satisfies the constraint

$$n_i = 248\,. \tag{93}$$

Without yet specifying which 248-dimensional gauge group is permitted, we substitute 248 for n_i in (87) obtaining

$$I_{12}(1-\mathrm{loop})_i = \frac{1}{2(2\pi)^5 6!}\left[-\frac{15}{16}\,\mathrm{tr}\,R^4\wedge \mathrm{tr}\,R^2 - \frac{15}{64}\,(\,\mathrm{tr}\,R^2)^3\right.$$
$$\left.+\frac{1}{16}\,\mathrm{tr}\,R^4\wedge \mathrm{Tr}\,F_i^2 + \frac{5}{64}\,(\,\mathrm{tr}\,R^2)^2\wedge \mathrm{Tr}\,F_i^2 - \frac{5}{8}\,\mathrm{tr}\,R^2\wedge \mathrm{Tr}\,F_i^4 + \mathrm{Tr}\,F_i^6\right] \tag{94}$$

Although non-vanishing, this part of the anomaly is reducible. It follows that it can be made to cancel as long as it can be factorized into the product of two terms, a four-form and an eight-form. A necessary requirement for this to be the case is that

$$\operatorname{Tr} F_i^6 = \frac{1}{24} \operatorname{Tr} F_i^4 \wedge \operatorname{Tr} F_i^2 - \frac{1}{3600} (\operatorname{Tr} F_i^2)^3 . \tag{95}$$

There are two Lie groups with dimension 248 that satisfy this condition, the non-Abelian group E_8 and the Abelian group $U(1)$.[248] Both groups represent allowed twisted matter gauge groups on each ten-plane. Hence, from anomaly considerations alone one can determine the twisted sector on each ten-plane, albeit with a small ambiguity in the allowed twisted sector gauge group. In this paper, we consider only the non-Abelian gauge group E_8. Using (95) and several E_8 trace relations, the anomaly polynomial (94) can be re-expressed as follows

$$I_{12}(1-\text{loop})_i = \frac{1}{3} \pi I_{4(i)}^3 + X_8 \wedge I_{4(i)} \tag{96}$$

where X_8 is the eight-form

$$X_8 = \frac{1}{(2\pi)^3 4!} \left(\frac{1}{8} \operatorname{tr} R^4 - \frac{1}{32} (\operatorname{tr} R^2)^2 \right) \tag{97}$$

and $I_{4\,(i)}$ is the four-form given by

$$I_{4\,(i)} = \frac{1}{16\pi^2} \left(\frac{1}{30} \operatorname{Tr} F_i^2 - \frac{1}{2} \operatorname{tr} R^2 \right) . \tag{98}$$

Once in this factorized form, the anomaly $I_{12}(1-\text{loop})_i$ can be cancelled as follows.

First, the Bianchi identity $dG = 0$, where G is the field strength of the three-form C in the eleven-dimensional supergravity multiplet, is modified to

$$dG = \sum_{i=1}^{2} I_{4(i)} \wedge \delta^{(1)}_{M_i^{10}} \tag{99}$$

where $I_{4(i)}$ is the four-form given in (98) and $\delta^{(1)}_{M_i^{10}}$ is a one-form brane current with support on the i-th ten-plane. Second, we note that the eleven-dimensional supergravity action contains the terms

$$S = \cdots - \frac{\pi}{3} \int C \wedge G \wedge G + \int G \wedge X_7 \tag{100}$$

where X_7 satisfies $dX_7 = X_8$. The CGG interaction is required by the minimally-coupled supergravity action, while the GX_7 term is an additional higher-derivative interaction necessitated by five-brane anomaly cancellation. Using the modified Bianchi identity (99), one can compute the variation of these two terms under Lorentz and gauge transformations. The result is that the CGG and GX_7 terms have classical anomalies which descend from the polynomials

$$I_{12}(CGG)_i = -\frac{\pi}{3} I^3_{4\,(i)} \tag{101}$$

and

$$I_{12}(GX_7)_i = -X_8 \wedge I_{4(i)} \,. \tag{102}$$

respectively. It follows that

$$I_{12}(1-\text{loop})_i + I_{12}(CGG)_i + I_{12}(GX_7)_i = 0 \tag{103}$$

and, hence, the total anomaly cancels exactly.

We conclude that the requirement of local anomaly cancellation on the each of the two $S^1/\mathbf{Z}_2$ orbifold ten-planes specifies the twisted spectrum of the theory. This specification is almost, but not quite, unique, allowing $N = 1$ vector supermultiplets with either gauge group E_8 or $U(1)^{248}$. An important ingredient in this analysis was the fact that the contribution to the anomaly on each ten-plane from the untwisted sector was a factor of $1/2$ smaller than the index theorem result. This followed from the fact that the index theorem had to be spread over two equivalent ten-planes. A direct consequence of this is that the non-Abelian gauge group on each ten-plane is E_8, not $E_8 \times E_8$, and that the gauge group $SO(32)$ is disallowed. Since $S^1/\mathbf{Z}_2$ is a subspace of $S^1/\mathbf{Z}_2 \times T^4/\mathbf{Z}_2$, the results of this section continue to hold on the larger orbifold. We now discuss the cancellation of local anomalies in the other factor space, $T^4/\mathbf{Z}_2$.

The quantum anomalies on each of the sixteen indistinguishable seven-planes of the $T^4/\mathbf{Z}_2$ orbifold are easy to analyze. In analogy with the ten-planes, an untwisted sector is induced on each seven-plane by the $\mathbf{Z}_2$ projection of the eleven-dimensional supergravity multiplet. This untwisted spectrum forms the seven-dimensional $N = 1$ supergravity multiplet consisting of a graviton, a gravitino, three vector fields, a two-form, a real scalar dilaton and a spin $1/2$ dilitino. However, unlike the case of a ten-plane, gravitational anomalies cannot be supported on a seven-plane. In fact, since there are no chiral fermions in seven-dimensions, no chiral anomaly of any kind, gravitational or gauge, can arise. Hence, with no local chiral anomalies to cancel, it would appear to be impossible to compute the twisted sector spectrum of any seven-plane. As long as we focus on the seven-planes exclusively, this conclusion is correct. However, as we will see below, the cancellation of the local anomalies on the thirty-two six-dimensional $\alpha\beta$ orbifold planes, formed from the intersection of the α ten-planes with the β

seven-planes, will require a non-vanishing twisted sector spectrum on each seven-plane and dictate its structure. With this in mind, we now turn to the analysis of anomalies localized on the intersection six-planes in the full $S^1/\mathbf{Z}_2 \times T^4/\mathbf{Z}_2$ orbifold.

As in the case for the ten-planes, a gravitational anomaly will arise on each six-plane due to the coupling of chiral projections of the bulk gravitino to currents localized on the thirty-two fixed planes. Since the thirty-two six-planes are indistinguishable, the anomaly is the same on each plane and can be computed by conventional means if proper care is taken. Noting that the standard index theorems can be applied to the small radius limit where the thirty-two six-planes coincide, it follows that the gravitational anomaly on each six-plane is simply one-thirty-second of the index theorem anomaly derived using the untwisted sector spectrum in six-dimensions. In this case, the untwisted sector spectrum is the $\mathbf{Z}_2 \times \mathbf{Z}_2$ projection of the eleven-dimensional bulk supergravity multiplet onto each six-dimensional fixed plane. This untwisted spectrum forms several $N = 1$ six-dimensional supermultiplets. Namely, the supergravity multiplet consisting of a graviton, a chiral gravitino and a self-dual two-form, four hypermultiplets each with four scalars and an anti-chiral hyperino, and one tensor multiplet with one anti-self-dual two-form, one scalar and an anti-chiral spin 1/2 fermion. A one-loop quantum gravitational anomaly then arises from one chiral spin 3/2 fermion, five anti-chiral spin 1/2 fermions and one each of self-dual and anti-self-dual tensors. However, the anomalies due to the tensors cancel each other. Noting that a chiral anomaly in six-dimensions is characterized by an eight-form, from which the anomaly arises by descent, we find, for the i-th six-plane, that

$$I_8(SG)_i = \frac{1}{32}\left(I^{(3/2)}_{GRAV}(R) - 5\, I^{(1/2)}_{GRAV}(R)\right) \tag{104}$$

where

$$I^{(3/2)}_{GRAV}(R) = \frac{1}{(2\pi)^3 4!}\left(-\frac{49}{48}\operatorname{tr} R^4 + \frac{43}{192}(\operatorname{tr} R^2)^2\right) \tag{105}$$

and

$$I^{(1/2)}_{GRAV}(R) = \frac{1}{(2\pi)^3 4!}\left(-\frac{1}{240}\operatorname{tr} R^4 - \frac{1}{192}(\operatorname{tr} R^2)^2\right), \tag{106}$$

where R is the six-dimensional Riemann tensor, regarded as an $SO(5,1)$-valued form. Note that the terms in brackets in (104) are the anomaly as computed by the index theorem. $I_8(SG)_i$ is obtained from that result by dividing by 32.

Noting that each six-plane is embedded in one of the two ten-dimensional planes, we see that there are additional "untwisted" sector fields on each six-plane. These arise from the β $\mathbf{Z}_2$ projection of the $N=1$ E_8 Yang-Mills supermultiplet on the associated ten-plane. Such fields are untwisted from the point of view of the six-dimensional plane, although they arise from fields that were in the twisted sector of the ten-plane. In this lecture, we will assume that the β action on the ten-dimensional vector multiplets does not break the E_8 gauge group. A discussion of the case where E_8 is broken to a subgroup by the action of β can be found in Refs. 55, 56. A ten-dimensional $N=1$ vector supermultiplet decomposes in six-dimensions into an $N=1$ vector multiplet and an $N=1$ hypermultiplet. However, the action of β projects out the hypermultiplet. Therefore, the ten-plane contribution to the untwisted sector of each six-plane is an $N=1$ E_8 vector supermultiplet, which consists of gauge fields and chiral gauginos. The gauginos contribute to the gravitational anomaly on each six-plane, as well as adding mixed and E_8 gauge anomalies. Noting that the standard index theorems can be applied to the small radius limit, where each ten-plane shrinks to zero size and, hence, the sixteen six-planes it contains coincide, it follows that the anomaly is simply one-sixteenth of the index theorem result. We find that the one-loop quantum contribution of this E_8 supermultiplet to the gravitational, mixed and E_8 gauge anomalies on the i-th six-plane is

$$I_8(E_8)_i = \frac{1}{16}\left(248\, I_{GRAV}^{(1/2)(R)} + I_{MIXED}^{(1/2)}(R,F_i) + I_{GAUGE}^{(1/2)}(F_i)\right) \tag{107}$$

where

$$I_{MIXED}^{(1/2)}(R,F_i) = \frac{1}{(2\pi)^3 4!}\left(\frac{1}{4}\operatorname{tr} R^2 \wedge \operatorname{Tr} F_i^2\right) \tag{108}$$

and

$$I_{GAUGE}^{(1/2)}(F_i) = \frac{1}{(2\pi)^3 4!}\left(-\operatorname{Tr} F_i^4\right). \tag{109}$$

Here Tr is over the adjoint **248** representation of E_8. Note that the terms in brackets in (107) are the index theorem anomaly. $I_8(E_8)_i$ is obtained from that result by dividing by 16.

Are there other sources of untwisted sector anomalies on a six-plane? The answer is, potentially yes. We note that, in addition to being embedded in one of the two ten-planes, each six-plane is also embedded in one of the sixteen seven-dimensional orbifold planes. In analogy with the discussion above, if there were to be a non-vanishing twisted sector spectrum on each seven-plane, then this could descend under the α $\mathbf{Z}_2$ projection as an addition to the untwisted spectrum on each six-plane. This additional untwisted spectrum could then contribute to the chiral anomalies on the six-plane. However, as noted above, a priori, there is no reason for one to believe that there is any twisted sector on a seven-dimensional orbifold plane. Therefore, for the time being, let us assume that there is no such contribution to the six-dimensional anomaly. We will see below that this assumption must be carefully revisited.

As for the ten-dimensional planes, one must allow for the possibility of twisted sector $N = 1$ supermultiplets on each of the thirty-two six-planes. The most general allowed spectrum on the i-th six-plane would be n_{Vi} vector multiplets transforming in the adjoint representation of some as yet unspecified gauge group $\mathcal{G}_i$, n_{Hi} hypermultiplets transforming under some representation (possibly reducible) $\mathcal{R}$ of $\mathcal{G}_i$, and n_{Ti} gauge-singlet tensor multiplets. We denote by $\mathcal{F}_i$ the gauge field strength. Since these fields are in the twisted sector, their contribution to the chiral anomalies can be determined directly from the index theorems without modification. We find that the one-loop quantum contribution of the twisted spectrum to the gravitational, mixed and $\mathcal{G}_i$ gauge anomalies on the i-th six-plane is

$$I_8(\mathcal{G}_i) = (n_V - n_H - n_T)_i \, I^{(1/2)}_{GRAV}(R) - n_{Ti} \, I^{(3\text{–form})}_{GRAV}(R) + I^{(1/2)}_{MIXED}(R, \mathcal{F}_i) + I^{(1/2)}_{GAUGE}(\mathcal{F}_i) \qquad (110)$$

where $I^{(1/2)}_{GRAV}(R)$ is given in (106) and

$$I^{(3\text{–form})}_{GRAV}(R) = \frac{1}{(2\pi)^3 4!}\left(-\frac{7}{60}\,\mathrm{tr}\,R^4 + \frac{1}{24}\,(\,\mathrm{tr}\,R^2\,)^2 \right). \qquad (111)$$

Furthermore, the mixed and pure-gauge anomaly polynomials are modified to

$$I^{(1/2)}_{MIXED}(R, \mathcal{F}_i) = \frac{1}{(2\pi)^3 4!}\left(\frac{1}{4}\,\mathrm{tr}\,R^2 \wedge \mathrm{trace}\,\mathcal{F}_i^2 \right) \qquad (112)$$

and

$$I^{(1/2)}_{GAUGE}(\mathcal{F}_i) = \frac{1}{(2\pi)^3 4!}\left(-\,\mathrm{trace}\,\mathcal{F}_i^4 \right), \qquad (113)$$

where

$$\mathrm{trace}\,\mathcal{F}_i^n = \mathrm{Tr}\,\mathcal{F}_i^n - \sum_\alpha h_\alpha \,\mathrm{tr}_\alpha \mathcal{F}_i^n\,. \qquad (114)$$

Here Tr is an adjoint trace, h_α is the number of hypermultiplets transforming in the $\mathcal{R}_\alpha$ representation and tr_α is a trace over the $\mathcal{R}_\alpha$ representation.

Note that the total number of vector multiplets is $n_{Vi} = \dim(\mathcal{G}_i)$, while the total number of hypermultiplets is $n_{Hi} = \sum_\alpha h_\alpha \times \dim(\mathcal{R}_\alpha)$. The relative minus sign in (114) reflects the anti-chirality of the hyperinos.

Combining the contributions from the two untwisted sector sources and the twisted sector, the total one-loop quantum anomaly on the i-th six-plane is the sum

$$I_8(1-\text{loop})_i = I_8(SG)_i + I_8(E_8)_i + I_8(\mathcal{G}_i) \tag{115}$$

where $I_8(SG)_i$, $I_8(E_8)_i$ and $I_8(\mathcal{G}_i)$ are given in (104), (107) and (110) respectively.

Unlike the case for the ten-dimensional planes, the classical anomaly associated with the GX_7 term in the eleven-dimensional action (100) can contribute to the irreducible curvature term which, in six-dimensions, is $\operatorname{tr} R^4$. Therefore, our next step is to further modify the Bianchi identity for $G = dC$ from expression (99) to

$$dG = \sum_{i=1}^{2} I_{4(i)} \wedge \delta^{(1)}_{M_i^{10}} + \sum_{i=1}^{32} g_i\, \delta^{(5)}_{M_i^6} \tag{116}$$

where $\delta^{(5)}_{M_i^6}$ has support on the six-planes M_i^6. As discussed in Refs. 54, 55, the magnetic charges g_i are required to take the values

$$g_i = -3/4, -1/4, +1/4, \ldots \tag{117}$$

Using the modified Bianchi identity (115), one can compute the variation of the GX_7 term under Lorentz and gauge transformations. The result is that this term gives rise to a classical anomaly that descends from the polynomial

$$I_8(GX_7)_i = -g_i\, X_8 \tag{118}$$

where X_8 is presented in expression (97). The relevant anomaly is then

$$I_8(1-\text{loop})_i + I_8(GX_7)_i \tag{119}$$

where $I_8(1-\text{loop})_i$ is given in (115). This anomaly spoils the consistency of the theory and, hence, must cancel. One begins the analysis of anomaly cancellation by considering the pure $\operatorname{tr} R^4$ term in (119) which is irreducible and must identically vanish. It follows from the above that this term is

$$-\frac{1}{(2\pi)^3 4!\, 240}\left(n_{Vi} - n_{Hi} - 29 n_{Ti} + 30\, g_i + 23\right) \operatorname{tr} R^4 \,. \tag{120}$$

Therefore, the $\operatorname{tr} R^4$ term will vanish if and only if on each orbifold plane the constraint

$$n_{Vi} - n_{Hi} - 29n_{Ti} + 30g_i + 23 = 0 \tag{121}$$

is satisfied. Herein lies a problem, and the main point of paper.[57] Noting from (117) that $g_i = c_i/4$ where $c_i = -3, -1, 1, 3, 5, \ldots$, we see that cancelling the $\operatorname{tr} R^4$ term requires that we satisfy

$$n_{Vi} - n_{Hi} - 29n_{Ti} = (-15c_i - 46)/2\,. \tag{122}$$

However, this is not possible since the left hand side of this expression is an integer and the right hand side always half integer. There is only one possible resolution of this problem, which is to carefully review the only assumption that was made above, that is, that there is no twisted sector on a seven-plane and, hence, no contribution of the seven-planes by α $\mathbf{Z}_2$ projection to the untwisted anomaly on a six-plane. As we now show, this assumption is false.

Let us now allow for the possibility that there is a twisted sector of $N = 1$ supermultiplets on each of the sixteen seven-planes. The most general allowed spectrum on the i-th seven-plane would be n_{7Vi} vector supermultiplets transforming in the adjoint representation of some as yet unspecified gauge group G_{7i}. Each seven-dimensional vector multiplet contains a gauge field, three scalars and a gaugino. With respect to six-dimensions, this vector multiplet decomposes into an $N = 1$ vector supermultiplet and a single hypermultiplet. Under the α $\mathbf{Z}_2$ projection to each of the two embedded six-planes, the gauge group G_{7i} can be preserved or broken to a subgroup. In either case, we denote the six-dimensional gauge group arising in this manner as $\tilde{\mathcal{G}}_i$, define $\tilde{n}_{Vi} = \dim \tilde{\mathcal{G}}_i$ and write the associated gauge field strength as $\tilde{\mathcal{F}}_i$. In this lecture, for simplicity, we will assume that the gauge group is unbroken by the orbifold projection, that is, $\tilde{\mathcal{G}}_i = G_{7i}$. The more general case where it is broken to a subgroup is discussed in Refs. 55, 56. Furthermore, the α action projects out either the six-dimensional vector supermultiplet, in which case the hypermultiplet descends to the six-dimensional untwisted sector, or the six-dimensional hypermultiplet, in which case the vector supermultiplet enters the six-dimensional untwisted sector. We denote by $\tilde{n}_{Hi}$ the number of hypermultiplets arising in the six-dimensional untwisted sector by projection from the seven-plane, and specify their (possibly reducible) representation under $\tilde{\mathcal{G}}_i$ as $\tilde{\mathcal{R}}$. Since these fields are in the untwisted sector associated with a single seven-plane, and since there are two six-planes embedded in each seven-plane, their contribution to the quantum anomaly on each six-plane can be determined by taking 1/2 of the index theorem result. We find that the one-loop quantum contribution of this part of the the untwisted spectrum to the gravitational, mixed and $\tilde{\mathcal{G}}_i$ gauge anomalies on the i-th six-plane is

$$I_8(\tilde{\mathcal{G}}_i) = \frac{1}{2}\left[(\tilde{n}_V - \tilde{n}_H)_i\, I^{(1/2)}_{GRAV}(R) + I^{(1/2)}_{MIXED}(R, \tilde{\mathcal{F}}_i) + I^{(1/2)}_{GAUGE}(\tilde{\mathcal{F}}_i)\right] \tag{123}$$

where $I^{(1/2)}_{GRAV}(R), I^{(1/2)}_{MIXED}(R, \tilde{\mathcal{F}}_i)$ and $I^{(1/2)}_{GAUGE}(\tilde{\mathcal{F}}_i)$ are given in (106),(112) and (113) respectively with the gauge and hypermultiplet quantities replaced by their " $\sim$ " equivalents.

The total quantum anomaly on the i-th six-plane is now modified to

$$I_8(1-\text{loop})_i + I_8(\tilde{\mathcal{G}}_i) \tag{124}$$

where $I_8(1-\text{loop})_i$ and $I_8(\tilde{\mathcal{G}}_i)$ are given in (115) and (123) respectively. It follows that the relevant anomaly contributing to, among other things, the irreducible $\operatorname{tr} R^4$ term is modified to

$$I_8(1-\text{loop})_i + I_8(\tilde{\mathcal{G}}_i) + I_8(GX_7)_i\,. \tag{125}$$

This anomaly spoils the quantum consistency of the theory and, hence, must cancel. We again begin by considering the pure $\operatorname{tr} R^4$ term in (125). This term is irreducible and must identically vanish. It follows from the above that this term is

$$\frac{-1}{(2\pi)^3 4!\, 240}\,(\, n_{Vi} - n_{Hi} + \tfrac{1}{2}\,\tilde{n}_{Vi} - \tfrac{1}{2}\,\tilde{n}_{Hi} - 29 n_{Ti} + 30 g_i + 23\,)\operatorname{tr} R^4. \tag{126}$$

Therefore, the $\operatorname{tr} R^4$ term will vanish if and only if on each orbifold plane the constraint

$$n_{Vi} - n_{Hi} + \tfrac{1}{2}\,\tilde{n}_{Vi} - \tfrac{1}{2}\,\tilde{n}_{Hi} - 29\, n_{Ti} + 30\, g_i + 23 = 0 \tag{127}$$

is satisfied. Again, noting that $g_i = c_i/4$ where $c_i = -3, -1, 1, 3, 5, \ldots$, we see that we must satisfy

$$n_{Vi} - n_{Hi} + \tfrac{1}{2}\,\tilde{n}_{Vi} - \tfrac{1}{2}\,\tilde{n}_{Hi} - 29\, n_{Ti} = \tfrac{1}{2}\,(\, -15 c_i - 46\,)\,. \tag{128}$$

As above, the right hand side is always a half integer. Now, however, because of the addition of the untwisted spectrum arising from the seven-plane, the left hand side can also be chosen to be half integer. Hence, the pure $\operatorname{tr} R^4$ term can be cancelled.

Having cancelled the irreducible $\operatorname{tr} R^4$ term, we now compute the remaining terms in the anomaly eight-form. In addition to the contributions from (125), we must also take into account the classical anomaly associated with the CGG term in the eleven-dimensional action (100). Using the modified Bianchi identity (115), one can compute the variation of the CGG term under Lorentz and gauge transformations. The result is that this term gives rise to a classical anomaly that descends from the polynomial

$$I_8(CGG)_i = -\pi\, g_i\, I^2_{4\,(i)} \tag{129}$$

where $I_{4\,(i)}$ is given in expression (98). Adding this anomaly to (125), and cancelling the $\operatorname{tr} R^4$ term by imposing constraint (127), we can now determine the remaining terms in the anomaly eight-form.

Recall that, in this lecture, we are assuming that the β action on the ten-dimensional vector supermultiplet does not break the E_8 gauge group. In this case, we can readily show that there can be no twisted sector vector multiplets on any six-plane. Rather than complicate the present discussion, we will simply assume here that gauge field strengths $\mathcal{F}_i$ do not appear. Furthermore, cancellation of the complete anomaly, in the case where E_8 is unbroken, requires that $\tilde{\mathcal{G}}_i$ be a product of $U(1)$ factors. Here, we will limit the discussion to the simplest case where

$$\tilde{\mathcal{G}}_i = U(1) \tag{130}$$

The β action on the seven-dimensional plane then either projects a single vector supermultiplet, or a single chargeless hypermultiplet, onto the untwisted sector of the six-plane. In either case, no $U(1)$ anomaly exists. Hence, the gauge field strengths $\tilde{\mathcal{F}}_i$ also do not appear. With this in mind, we now compute the remaining terms in the anomaly eight-form. They are

$$\frac{1}{(2\pi)^3 4!\, 16}\Big(\tfrac{3}{4}\,(\,1-4\,n_{Ti}\,)\,(\operatorname{tr} R^2)^2 + \tfrac{1}{20}\,(\,5+8\,g_i\,)\operatorname{tr} R^2 \wedge \operatorname{Tr} F_i^2$$
$$-\tfrac{1}{100}\,(\,1+\tfrac{4}{3}\,g_i\,)\,(\,\operatorname{Tr} F_i^2)^2\Big) \tag{131}$$

where we have used the E_8 trace relation $\operatorname{Tr} F^4 = \frac{1}{100}(\operatorname{Tr} F^2)^2$. Note that, since n_{Ti} is a non-negative integer and g_i must satisfy (117), the first two terms of this expression term can never vanish. Furthermore, it is straightforward to show that (131) will factor into an exact square, and, hence, be potentially cancelled by a six-plane Green-Schwarz mechanism, if and only if

$$4\,(\,4\,n_{Ti}-1\,)(\,3+4\,g_i\,) = (\,5+8\,g_i\,)^2 \tag{132}$$

Again, this equation has no solutions for the allowed values of n_{Ti} and g_i. It follows that anomaly (131), as it presently stands, cannot be be made to identically vanish or cancel. The resolution of this problem was first described in Ref. 54, and consists of the realization that the existence of seven-planes in the theory necessitates the introduction of additional Chern-Simons interactions in the action, one for each seven-plane. The required terms are

$$S = \cdots + \sum_{i=1}^{16} \int \delta^{(4)}_{M^7_i} \wedge G \wedge Y^0_{3(i)} \tag{133}$$

where $dY^0_{3(i)} = Y_{4(i)}$ is a gauge-invariant four-form polynomial. $Y_{4(i)}$ arises from the curvature R and also the field strength $\tilde{\mathcal{F}}_i$ associated with the additional adjoint super-gauge fields localized on the i-th seven-plane. It is given by

$$Y_{4(i)} = \frac{1}{4\pi}\Big(-\frac{1}{32}\,\eta\,\mathrm{tr}\,R^2 + \rho\,\mathrm{tr}\,\tilde{\mathcal{F}}_i \Big) \tag{134}$$

where η and ρ are rational coefficients. Using the modified Bianchi identity (116), one can compute the variation of the $\delta^7 G Y_3$ terms under Lorentz and gauge transformations. The result is that these give rise to a classical anomaly that descends from the polynomial

$$I_8(\delta^7 G Y_3)_i = -I_{4\,(i)} \wedge Y_{4(i)} \tag{135}$$

where $I_{4\,(i)}$ is the four-form given in (98).

The total anomaly on the i-th six-plane is now modified to

$$I_8(1-\mathrm{loop})_i + I_8(\tilde{\mathcal{G}}_i) + I_8(GX_7)_i + I_8(CGG)_i + I_8(\delta^7 G Y_3)_i \tag{136}$$

where $I_8(1-\mathrm{loop})_i, I_8(\tilde{\mathcal{G}}_i), I_8(GX_7)_i, I_8(CGG)_i$ and $I_8(\delta^7 G Y_3)_i$ are given in (115), (123), (118), (129) and (135) respectively. Note that for the fixed plane intersection presently under discussion, the field strength $\tilde{\mathcal{F}}_i$ does not enter the anomaly eight-form (131). Therefore, within this context, we must take

$$\rho = 0\,. \tag{137}$$

After cancelling the irreducible $\mathrm{tr}\,R^4$ term, the remaining anomaly now becomes

$$\frac{1}{(2\pi)^3 4!16}\Big(\frac{3}{4}\,(1 - 4n_{Ti} - \eta)\,(\mathrm{tr}\,R^2)^2 + \frac{1}{20}\,(5 + 8g_i + \eta)\,\mathrm{tr}\,R^2 \wedge \mathrm{Tr}\,F_i^2 - \frac{1}{100}\,(1 + \frac{4}{3}g_i)\,(\mathrm{Tr}\,F_i^2)^2 \Big) \tag{138}$$

Depending on the number of untwisted hypermultiplets, n_{Ti}, these terms can be made to cancel or to factor into the sum of exact squares. In this lecture, we consider the $n_{Ti} = 0, 1$ cases only. As discussed in Refs. 55, 56, the solutions where $n_{Ti} \geq 2$ are related to the $n_{Ti} = 0, 1$ solutions by the absorption of one or more five-branes from the bulk space onto the i-th six-plane.

We first consider the case where

$$n_{Ti} = 0\,. \tag{139}$$

In this case, no further Green-Schwarz type mechanism in six-dimensions is possible and the anomaly must vanish identically. We see from (138) that this is possible if and only if

$$g_i = -3/4, \qquad \eta = 1\,. \tag{140}$$

It is important to note that this solution only exists for a non-vanishing value of parameter η. Hence, the additional Chern-Simons interactions (133) are essential for the anomaly to vanish identically in the $n_{Ti} = 0$ case. Inserting these results into expression (127) for the vanishing of the irreducible $\operatorname{tr} R^4$ term, and recalling that $n_{Vi} = 0$, we find that

$$-2n_{Hi} + \tilde{n}_{Vi} - \tilde{n}_{Hi} = -1\,. \tag{141}$$

Equation (141) can be solved in several ways. Remembering that $\tilde{\mathcal{G}}_i = U(1)$, the first solution then consists of allowing the $U(1)$ hypermultiplet to descend to the six-plane while projecting out the $U(1)$ vector multiplet. Equation (141) is then solved by taking the number of twisted hypermultiplets to vanish. That is, take

$$\tilde{n}_{Hi} = 1, \qquad \tilde{n}_{Vi} = 0, \qquad n_{Hi} = 0\,. \tag{142}$$

The second solution follows by doing the reverse, that is, projecting out the $U(1)$ hypermultiplet and allowing the $U(1)$ vector multiplet to descend to the six-plane. In this case, equation (141) is solved by taking

$$\tilde{n}_{Hi} = 0, \qquad \tilde{n}_{Vi} = 1, \qquad n_{Hi} = 1\,. \tag{143}$$

Let us now consider the case where

$$n_{Ti} = 1\,. \tag{144}$$

In this case, the anomaly (138) can be removed by a six-dimensional Green-Schwarz mechanism as long as it factors into an exact square. It is straightforward to show that this will be the case if and only if

$$4\,(3+\eta)\,(3+4\,g_i) = (5+8\,g_i+\eta)^2\,. \tag{145}$$

This equation has two solutions

$$g_i = -3/4, \qquad \eta = 1 \tag{146}$$

and

$$g_i = 1/4, \qquad \eta = 1\,. \tag{147}$$

Again, note that these solutions require a non-vanishing value of the parameter η. Hence, the additional Chern-Simons interactions (133) are also essential for anomaly factorization in the $n_{Ti} = 1$ case. Inserting these into the expression for the vanishing of the irreducible $\mathrm{tr}\, R^4$ term, and recalling that $n_{Vi} = 0$, we find

$$-2n_{Hi} + \tilde{n}_{Vi} - \tilde{n}_{Hi} = 57 \tag{148}$$

and

$$-2n_{Hi} + \tilde{n}_{Vi} - \tilde{n}_{Hi} = -3\,. \tag{149}$$

The first equation (148) cannot be solved within the context of $\tilde{\mathcal{G}}_i = U(1)$, since $\tilde{n}_{Vi} \leq 1$. The second equation, however, has two solutions

$$\tilde{n}_{Hi} = 1, \qquad \tilde{n}_{Vi} = 0, \qquad n_{Hi} = 1 \tag{150}$$

and

$$\tilde{n}_{Hi} = 0, \qquad \tilde{n}_{Vi} = 1, \qquad n_{Hi} = 2\,. \tag{151}$$

In either case, the anomaly (138) factors into an exact square given by

$$-\frac{3}{(2\pi)^3 4!\, 16}\left(\mathrm{tr}\, R^2 - \frac{1}{15}\,\mathrm{Tr}\, F_i^2\right)^2 . \tag{152}$$

The anomaly can now be cancelled by a Green-Schwarz mechanism on the six-plane. First, one alters the Bianchi identity for the anti-self-dual tensor in the twisted sector tensor multiplet from $dH_{Ti} = 0$, where H_{Ti} is the tensor field strength three-form, to

$$dH_{Ti} = \frac{1}{16\pi^2}(\,\mathrm{tr}\, R^2 - \frac{1}{15}\mathrm{Tr}\, F_i^2\,)\,. \tag{153}$$

Second, additional Chern-Simons terms are added to the action, one for each six-plane. The required terms are

$$S = \cdots - \frac{1}{64\pi}\sum_{i=1}^{32}\int \delta^{(5)}_{M_i^6} \wedge B_{Ti} \wedge (\,\mathrm{tr}\, R^2 - \frac{1}{15}\mathrm{Tr}\, F_i^2\,)\,, \tag{154}$$

where B_{Ti} is the anti-self-dual tensor two-form on the i-th six-plane. Using Bianchi identity (153), one can compute the variation of each such term under Lorentz and gauge transformations. The result is a classical anomaly that descends from an eight-form that exactly cancels expression (152). The theory is now anomaly free.

Thus, we have demonstrated, within the context of an explicit orbifold fixed plane intersection where the β $\mathbf{Z}_2$ projection to the six-plane leaves E_8 unbroken, that all local anomalies can be cancelled. However, this cancellation requires that the intersecting seven-plane support a twisted sector consisting of a $U(1)$ $N = 1$ vector supermultiplet and an associated Chern-Simons term. This term is of the form (133) with $\eta = 1$ and $\rho = 0$. The fact that $\rho = 0$ in this context follows directly from the property that E_8 is unbroken by the β projection.

We conclude that, as has been discussed in detail in Refs. 54, 55, 56, 57 and Refs. 58, 59, anomaly free M-theory orbifolds associated with the spacetime $\mathbf{R}^6 \times K3$ can be constructed in detail, including the entire twisted and untwisted spectra. This work has now been extended to orbifolds of spacetime $\mathbf{R}^4 \times CY_3$, where CY_3 is a Calabi-Yau threefold, in Ref. 60. This last work opens the door to finding realistic standard model-like M-theory vacua within this context.

5. Discussion

Hořava-Witten theory and its compactification on Calabi-Yau threefolds to heterotic M-theory have stimulated a great deal of both formal M-theory research as well as discussions of the associated phenomenology. In addition to the papers referenced in the above lectures, further relevant literature can be found in Refs. 80 61.

Heterotic M-theory has also served as a consistent and phenomenologically relevant venue for studying M-theory cosmology. This research comes in two catagories. The first consists of work discussing subluminal expansion, Kasner-like solutions and inflation within the context of brane world scenarios associated with heterotic M-theory. These results can be found in Refs. 81, 82, 83, 84, 85, 86, 87. Very recently, a new theory of the early universe, called the Ekpyrotic Universe, has been constructed for generic brane world scenarios, including heterotic M-theory. In the Ekpyrotic scenario, all expansion is subluminal, with no period of inflation. A nearly scale-invariant spectrum of fluctuations in the microwave background is obtained, not as quantum fluctuations in deSitter space but, rather, as the fluctuations on a bulk brane or end-of-the-world brane as it moves through the fifth-dimension. The fundamental papers on this subject can be found in Refs. 88, 89, 90, 91, 92, 93.

Acknowledgements

Burt Ovrut is supported in part by the DOE under contract No. DE-AC02-76-ER-03071.

References

1. P. Hořava and E. Witten, *Nucl. Phys.* **B460** (1996) 506.
2. P. Hořava and E. Witten, *Nucl. Phys.* **B475** (1996) 94.
3. A. Lukas, B. A. Ovrut, K.S. Stelle and D. Waldram, *The Universe as a Domain Wall*, Phys.Rev. D59 (1999) 086001.
4. A. Lukas, B. A. Ovrut, K.S. Stelle and D. Waldram, *Heterotic M–theory in Five Dimensions*, Nucl.Phys. B552 (1999) 246-290.
5. E. Cremmer, in *Superspace and Supergravity*, ed. S. W. Hawking and M. Rocek, pp. 267–282, Cambridge University Press (1981).
6. M. Günaydin, G. Sierra and P.K. Townsend, *Nucl. Phys.* **B242** (1084) 244.
7. A. C. Cadavid, A. Ceresole, R. D'Auria and S. Ferrara, *Phys. Lett.* **B357** 1995 76; I. Antoniadis, S. Ferrara and T. R. Taylor, *Nucl. Phys.* **B460** (1996) 489.
8. M. B. Green, J. H. Schwarz and E. Witten, *Superstring Theory*, vol. 2, pp. 386, Cambridge University Press (1987).
9. I. V. Lavrinenko, H. Lü and C. N. Pope, *Nucl. Phys.* **B492** (1996) 278.
10. A. H. Chamseddine and H. Nicolai, *Phys. Lett.* **B96** (1980) 89.
11. M. Günaydin, G. Sierra and P.K. Townsend, *Nucl. Phys.* **B253** (1985) 573.
12. G. Sierra, *Phys. Lett.* **157B** (1985) 379.
13. A. Strominger, *Phys. Lett.* **B383** (1996) 44.
14. L. Andrianopoli *et al.*, *Nucl. Phys.* **B476** (1996) 397.
15. J. Polchinski and A. Strominger, *Phys. Lett.* **B388** (1996) 736.
16. J. Michelson, *Nucl. Phys.* **B495** (1997) 127.
17. E. Witten, *Nucl. Phys.* **B471** (1996) 135.
18. T. Banks and M. Dine, *Nucl. Phys.* **B479** (1996) 173.
19. P. Hořava, *Phys. Rev.* **D54** (1996) 7561.
20. A. Lukas, B. A. Ovrut and D. Waldram, *On the Four-Dimensional Effective Action of Strongly Coupled Heterotic String Theory*, Nucl.Phys. B532 (1998) 43-82.
21. K. Choi, *Phys. Rev.* **D56** (1997) 6588.
22. H. P. Nilles, M. Olechowski and M. Yamaguchi, *Phys. Lett.* **B415** (1997) 24.
23. I. Antoniadis and M. Quirós, *Phys. Lett.* **B416**(1998) 327.
24. K. Choi, H. B. Kim and C. Muñoz, *Four–Dimensional Effective Supergravity and Soft Terms in M–Theory*, KAIST-TH 97/20, FTUAM 97/16, hep-th/9711158.
25. A. Lukas, B. A. Ovrut and D. Waldram, *Gaugino Condensation in M–theory on S^1/Z_2*, Phys.Rev. D57 (1998) 7529-7538.
26. H. P. Nilles, M. Olechowski and M. Yamaguchi, *Supersymmetry Breakdown at a Hidden Wall*, BONN-TH-97-10, hep-th/9801030.
27. A. Lukas, B. A. Ovrut and D. Waldram, *The Ten-dimensional Effective Ac-*

tion of Strongly Coupled Heterotic String Theory, Nucl.Phys. B540 (1999) 230-246.
28. T. Li, *Compactification and Supersymmetry Breaking in M Theory*, hep-th/9801123.
29. M. Faux, *New Consistent Limits of M Theory*, HUB-EP-98-9, hep-th/9801204.
30. E. Bergshoeff, M. de Roo, M. B. Green, G. Papadopoulos and P. K. Townsend, *Nucl. Phys.* **B470** (1996) 113.
31. L. J. Romans, *Phys. Lett.* **B169** (1986) 374.
32. H. Lu, C. N. Pope, E. Sezgin and K. S. Stelle, *Phys. Lett.* **B371** (1996) 46; P. M. Cowdall, H. Lu, C. N. Pope, K. S. Stelle and P. K. Townsend, *Nucl. Phys.* **B486** (1997) 49.
33. R. Friedman, J. Morgan, E. Witten, Commun.Math.Phys. **187** (1997) 679-743.
34. R. Donagi, Asian Journal of Math., Vol 1, (1997)214, math.AG/ 9702002.
35. R. Friedman, J. Morgan, E. Witten, math.AG/9709029 (1997).
36. R. Donagi, A. Lukas, B.A. Ovrut and D. Waldram, JHEP 9905 (1999) 018, JHEP 9906 (1999) 034; R. Donagi, B.A. Ovrut and D. Waldram, JHEP 9911 (1999) 030; R. Donagi, J. Khoury, B.A. Ovrut, P.J. Steinhardt and N. Turok, hep-th/0105199 (2001).
37. R. Donagi, B.A. Ovrut, T. Pantev and D. Waldram, hep-th/9912208 (1999) to appear; JHEP 0108 (2001) 053; math.AG/0008010 (2000) to appear; math.AG/0008011 (2000) to appear.
38. G. W. Gibbons and P. K. Townsend, *Phys. Rev. Lett.* **71** (1993) 3754.
39. D. M. Kaplan and J. Michelson, *Phys. Rev.* **D53** (1996) 3474.
40. I. Bandos *et al.*, *Phys. Rev. Lett.* **78** (1997) 4332.
41. M. Aganagic *et al.*, *Nucl. Phys.* **B496** (1997) 191.
42. R. Kallosh, *Phys. Rev.* **D57** (1998) 3214.
43. P. Claus, R. Kallosh and A. Van Proeyen, *Nucl. Phys.* **B518** (1998) 117.
44. K. Becker, M. Becker and A. Strominger, *Nucl. Phys.* **B456** (1995) 130.
45. M. Bershadsky, V. Sadov and C. Vafa, *Nucl. Phys.* **B463** (1996) 420.
46. E. Witten, *Nucl. Phys.* **B500** (1997) 3.
47. A. Klemm *et al.*, *Nucl. Phys.* **B477** (1996) 746.
48. E. Witten, *J. Geom. Phys.* **22** (1997) 1.
49. A. Grassi, Z. Guralnik and B.A. Ovrut, JHEP 0101 (2001) 037.
50. A. Grassi, Z. Guralnik and B.A. Ovrut, hep-th/0110036 (2001).
51. K. Kodaira, Ann. of Math. **77** (1963) 563.
52. K.Dasgupta and S.Mukhi, Nucl.Phys. **B**465 (1996) 399-412.
53. E.Witten, Nucl.Phys. **B**463 (1996) 383-397.
54. M.Faux, D.Lüst and B.A.Ovrut, Nucl.Phys. **B**554 (1999) 437-483.
55. M.Faux, D.Lüst and B.A.Ovrut, Nucl.Phys. B589 (2000) 269-291.
56. M.Faux, D.Lüst and B.A.Ovrut, hep-th/0010087 (2000).
57. M.Faux, D.Lüst and B.A.Ovrut, hep-th/0011031 (2000).
58. A. Bilal, J.-P. Derendinger and R. Sauser, Nucl.Phys. B576 (2000) 347-374.
59. V. Kaplunovsky, J. Sonnenschein, S. Theisen and S. Yankielowicz, Nucl.Phys. B590 (2000) 123-160.

60. C. Doran, M. Faux and B.A. Ovrut, hep-th/0108078 (2001).
61. I. Antoniadis and M. Quirós, *Phys. Lett.* **B392** (1997) 61.
62. E. Caceres, V. S. Kaplunovsky and I. M. Mandelberg, *Nucl. Phys.* **B493** (1997) 73.
63. H.-P. Nilles and S. Stieberger, *Nucl. Phys.* **B499** (1997) 3.
64. T. Li, J. L. Lopez and D. V. Nanopoulos, *Mod. Phys. Lett.* **A12** (1997) 2647.
65. J. O. Conrad, TUM-HEP-289-97, hep-th/9708031.
66. E. Dudas and C. Grojean, *Nucl. Phys.* **B507** (1997) 553.
67. E. Dudas and J. Mourad, *Phys. Lett.* **B400** (1997) 71.
68. Z. Lalak, A. Lukas and B. A. Ovrut, Phys.Lett. B425 (1998) 59-70.
69. Z. Lalak, S. Pokorski and S. Thomas, CERN-TH/98-230, hep-ph/9807503.
70. Z. Lalak and S. Thomas, *Gaugino Condensation, Moduli Potential and Supersymmetry Breaking in M–theory Models*, QMW-PH-97-23, hep-th/9707223.
71. E. Dudas, *Supersymmetry breaking in M–theory and Quantization Rules*, CERN-TH/97-236, hep-th/9709043.
72. J. Ellis, A. E. Faraggi and D. V. Nanopoulos, *M–theory Model-Building and Proton Stability*, CERN-TH-97-237, hep-th/9709049.
73. D. V. Nanopoulos, *M Phenomenology*, CTP-TAMU-42-97, hep-th/9711080.
74. P. Majumdar and S. SenGupta, *Gaugino Mass in the Heterotic String with Scherk–Schwarz Compactification*, SINP-98-05, hep-th/9802111.
75. N. Wyllard, *The Heterotic Prepotential from Eleven Dimensions*, GOTEBORG-ITP-98-01, hep-th/9802148.
76. D. Bailin, G.V. Kraniotis and A. Love, *Sparticle Spectrum and Dark Matter in M Theory*, SUSX-TH-98-003, hep-ph/9803274.
77. E. Sharpe, *Boundary Superpotentials*, PUPT-1660, hep-th/9611196.
78. E. A. Mirabelli and M. E. Peskin, *Transmission of Supersymmetry Breaking from a Four–Dimensional Boundary*, SLAC-PUB-7681, hep-th/9712214.
79. J. Ellis, Z. Lalak, S. Pokorski and W. Pokorski, *Five–dimensional Aspects of M–theory and Supersymmetry Breaking*, CERN-TH-98-118, hep-ph/9805377.
80. S. P. de Alwis, *Phys. Lett.* **B388** (1996) 291; *Phys. Lett.* **B392** (1997) 332.
81. A. Lukas and B. A. Ovrut, UPR-770-T, Phys.Lett. B437 (1998) 291-302.
82. K. Benakli, hep-th/9804096.
83. A. Lukas, B. A. Ovrut and D. Waldram, Phys.Rev. D60 (1999) 086001.
84. A. Lukas, B. A. Ovrut and D. Waldram, hep-th/9812052, Lectures presented at the Advanced School on Cosmology and Particle Physics, Peniscola, Spain, June 1998.
85. A. Lukas, B. A. Ovrut and D. Waldram, Phys.Rev. D61 (2000) 023506.
86. G. Huey, P.J. Steinhardt, B.A. Ovrut and Daniel Waldram, Phys.Lett. B476 (2000) 379-386.
87. M. Braendle, A. Lukas, B.A. Ovrut, Phys.Rev. D63 (2001) 026003.
88. J. Khoury, B.A. Ovrut, P.J. Steinhardt and N. Turok, hep-th/0103239 (2001); hep-th/0105212 (2001).
89. R. Donagi, J. Khoury, B.A. Ovrut, P.J. Steinhardt and N. Turok, hep-th/0105199 (2001).
90. J. Khoury, B.A. Ovrut, N. Seiberg, P.J. Steinhardt and N. Turok, hep-th/0108187 (2001).

91. J. Khoury, B.A. Ovrut, P.J. Steinhardt and N. Turok, hep-th/0109050 (2001).
92. P.J. Steinhardt and N. Turok, hep-th/0111030 (2001).
93. P.J. Steinhardt and N. Turok, hep-th/0111098 (2001).

KRISHNA RAJAGOPAL

This page intentionally left blank

COLOR SUPERCONDUCTIVITY

KRISHNA RAJAGOPAL
Center for Theoretical Physics Massachusetts Institute of Technology, Cambridge, MA
E-mail: krishna@ctp.mit.edu

1. Introduction

In the decades that have passed since its inception, QCD has become firmly established as the theory underlying all of strong-interaction physics—a pillar of the standard model. Perturbative QCD has been verified in deep inelastic scattering, and the spectrum and structural properties of the hadrons are gradually being calculated by the nonperturbative lattice formulation of QCD.

Even so, there remain tantalizing questions. As well as predicting the behavior of small numbers of particles, QCD should also be able to tell us about the thermodynamics of matter in the realm of extraordinarily high temperatures and densities at which it comes to dominate the physics. These regions are of more than academic interest: neutron stars are believed to consist of matter squeezed beyond nuclear density by gravitational forces, and the whole universe was hotter than 100 MeV for the first crucial microseconds of its history. However, only in the last few years have these regions begun to be probed experimentally in heavy ion collisions and astrophysical observations of neutron stars, and our theoretical understanding of them remains elementary.

High densities have proven particularly difficult to study, in part because lattice gauge theory has been blocked by the complexity of the fermion determinant. We are still trying to establish the symmetries of the ground state, and find effective theories for its lowest excitations. These questions are of direct physical relevance: an understanding of the symmetry properties of dense matter can be expected to inform our understanding of neutron star astrophysics and conceivably also heavy ion collisions which somehow achieve high baryon densities without reaching high temperatures.

In these lectures, we explore the progress that has been made in the last few years in understanding the possible phases of QCD at low temperatures and high densities, and go on to discuss the possible consequences for in compact stars phenomenology. Other reviews, with different emphases, will also prove useful to the reader.[1]

2. The Fermi Surface and Cooper Instability

One of the most striking features of QCD is asymptotic freedom,[2] the fact that the force between quarks becomes arbitrarily weak as the characteristic momentum scale of their interaction grows larger. This immediately suggests that at sufficiently high densities and low temperatures, QCD will become tractable.[3] High density brings in a large energy scale, the chemical potential μ, and one might hope that the relevant coupling to describe the dynamics is $g(\mu)$. Matter will then consist of a Fermi sea of essentially free quarks, whose behavior is dominated by the freest of them all: the high-momentum quarks that live at the Fermi surface.

This naive expectation does not stand up to critical scrutiny. As we shall discuss at length below, perturbation theory around the naive ground state (free quark Fermi spheres) encounters infrared divergences. Furthermore, the naive perturbative ground state is unstable. Fortunately, related difficulties have been met and overcome previously, in the theory of superconductivity. There we learn that arbitrarily weak attractive interactions can change the ground state qualitatively. In the true ground state there is an effective mass for photons — the Meissner effect — and energy gaps for charged excitations. These phenomena remove potential infrared divergences, and render the perturbation theory around the true ground state regular (nondegenerate).

This can be seen intuitively in the following way. Consider a system of free particles. The Helmholtz free energy is $F = E - \mu N$, where E is the total energy of the system, μ is the chemical potential, and N is the number of particles. The Fermi surface is defined by a Fermi energy $E_F = \mu$, at which the free energy is minimized, so adding or subtracting a single particle costs zero free energy. Now, suppose a weak attractive interaction is introduced. Bardeen, Cooper, and Schrieffer (BCS)[4] showed that this favors a complete rearrangement of the states near the Fermi surface, because it costs no free energy to make a pair of particles (or holes), and the attractive interaction makes it favorable to do so. Many such pairs will therefore be created, in all the modes near the Fermi surface, and these pairs, being bosonic, will form a condensate. The ground state will be a superposition of states with all numbers of pairs, breaking the fermion number symmetry. An arbitrarily weak interaction has lead to spontaneous symmetry breaking.

Since pairs of quarks cannot be color singlets, the resulting condensate will break the local color symmetry $SU(3)_{\mathrm{color}}$. This is the definition of color superconductivity.[5,6,7,8,9] Note that the quark pairs play the same role here as the Higgs particle does in the standard model: the color-superconducting phase can be thought of as the Higgsed (as opposed to confined) phase of QCD.

It is important to remember that the breaking of a gauge symmetry cannot be characterized by a gauge-invariant local order parameter which vanishes on one side of a phase boundary.[10] The superconducting phase can be characterized rigorously only by its global symmetries. In electromagnetism there is a non-local order parameter, the mass of the magnetic photons, that corresponds physically to the Meissner effect and distinguishes the free phase from the superconducting one. In nonabelian theories like QCD there is no free phase: even without pairing the gluons are not states in the spectrum. No order parameter distinguishes the Higgsed phase from a confined phase or a plasma, so we have to look at the global symmetries. This is just what happens in the standard electroweak model, whose Higgs phase can also be understood as a confined phase.[11] The absence of massless gauge bosons and of long-range forces is the essence of the Meissner-Anderson-Higgs effect, and it is also the essence of confinement.

For detailed weak-coupling calculations it is more convenient to work in a fixed gauge, where *after* gauge-fixing the gauge potentials make only small fluctuations around zero. Of course at the end of any calculation we must restore the gauge symmetry, by averaging over the gauge fixing parameters, and only gauge-invariant quantities will survive. However, in the intermediate steps, within a fixed gauge, one can capture important correlations that characterize the ground state by specifying the existence of nonzero condensates relative to the ambient gauge. This is the procedure used in the standard model, where intermediate steps in the calculations involve a nonzero vacuum expectation value for a Higgs doublet field $\langle\phi^a\rangle = v\delta_1^a$, which is not gauge invariant. In superconductivity, the essence of the physics is the correlation in the fermionic wave function which describes the Cooper pairs, and the resulting modification of the dispersion relations which describe the excitation spectrum. In particular, the gap in the spectrum of fermionic excitations at the Fermi surface is a gauge invariant quantity. Describing this physics within a fixed gauge as a condensate which "breaks" the gauge symmetry is a convenient fiction. By forging a connection with superconductivity and condensate formation, it brings the universality class of confinement down to earth, and makes it accessible to weak coupling methods. These condensates need not break any true (i.e. global) symmetries. If a global symmetry *is* broken, some combination of the condensates themselves is a gauge invariant physical observable, and not just a convenient fiction.

Compared to ordinary superconductivity, color superconductivity, though it appears superficially to be more complex mathematically, is in a profound sense simpler and more directly related to fundamentals. Ordinary superconductivity takes place in solids and the accurate effective interactions are determined by band structure and other complicated effects. Furthermore, ordinary superconductivity in a metal involves electron pairing, and the fundamental interaction between electrons (the screened Coulomb interaction) is repulsive. The effective attraction near the Fermi surface that leads to superconductivity arises in classic superconductors only as a subtle consequence of retarded phonon interactions, and in the cuprate superconductors through some mechanism yet unknown. In color superconductivity, by contrast, the attractive interaction can arise already from the primary, strong, interactions. This has two consequences. First, the accurate form of these interactions can be calculated from first principles, using asymptotic freedom. This makes calculations at high enough density robust. Second, at accessible densities, where the strong interactions are much stronger than the electromagnetic interactions, we expect the color superconductors themselves to be robust in the sense that the ratio of their gaps and critical temperatures to the Fermi energy should be quite large.

In QCD with three colors and three flavors, we find an improved ground state at high density, based on color superconductivity, around which weak-coupling perturbation theory is valid. In particular, all the colored degrees of freedom acquire gaps. Thus, the improved ground state differs qualitatively from the naive one.

The resulting predictions regarding the low-energy spectrum and dynamics are striking. Color symmetry and chiral symmetry are spontaneously broken. The spectrum of elementary excitations is quite different from that found in naive perturbation theory. Nominally massless quarks and gluons become massive, new massless collective modes appear, and various quantum numbers get modified. All the elementary excitations carry integral electric charges.[12] Altogether, one finds an uncanny resemblance between the properties one computes at asymptotic densities, directly from the microscopic Lagrangian, and the properties one expects to hold at low density, based on the known phenomenology of hadrons. In particular, the traditional "mysteries" of confinement and chiral symmetry breaking are fully embodied in a controlled, fully microscopic, weak-coupling (but non-perturbative!) calculation, that accurately describes a directly physical, intrinsically interesting regime.[12,13]

3. The Gap Equation

We shall return to the main line of these lectures, namely a description of the physical properties of color superconducting quark matter, in Section 4. In this section, we give an incomplete introduction to the methods used to calculate the gap, which is the fundamental energy scale characterizing a (color) superconductor. The gap Δ is the free energy of a fermionic excitation about the ground state. Also, the critical temperature T_c below which quark matter is a color superconductor and above which the condensate melts, yielding a quark-gluon plasma, is given by 0.57Δ, as in any BCS superconductor.

It would be ideal if the calculation of the gap were within the scope of lattice gauge theory as is, for example, the calculation of the critical temperature for the QCD phase transition at zero baryon density. Unfortunately, lattice methods relying on importance sampling have to this point been rendered exponentially impractical at nonzero baryon density by the complex action at nonzero μ. Various lattice methods *can* be applied for $\mu \neq 0$ as long as T/μ is large enough;[14] so far, though, none have proved applicable at temperatures which are low enough that color superconductivity occurs. Lattice simulations are possible in two-color QCD and in QCD at large isospin density. Finally, sophisticated algorithms have recently allowed theories which are simpler than QCD but which have as severe a fermion sign problem as that in QCD at nonzero chemical potential to be simulated.[15] All of this bodes well for the future.

To date, in the absence of suitable lattice methods, quantitative analyses of color superconductivity have followed two distinct strategies. The first approach is utilitarian and semi-phenomenological, emphasizing the use of simplified models. The overarching theme here is to define models which incorporate the salient physical effects, yet are tractable using known mathematical techniques of quantum many-body theory. Free parameters within a model of choice are chosen to give reasonable vacuum physics. Examples include analyses in which the interaction between quarks is replaced simply by four-fermion interactions with the quantum numbers of the instanton interaction[8,9,16] or of one-gluon exchange,[12,17] random matrix models,[18] and more sophisticated analyses done using the instanton liquid model.[19,20] Renormalization group methods have also been used to explore the space of all possible effective four-fermion interactions.[21,22] These methods yield results which are in qualitative agreement: the gaps range between several tens of MeV up to of order 100 MeV; the associated critical temperatures (above which the diquark condensates vanish) can therefore be as large as about $T_c \sim 50$ MeV. This agreement between different models reflects the fact that what matters most is simply the strength of the attraction between quarks in the color $\bar{\mathbf{3}}$ channel, and by fixing the parameters of the model interaction to fit, say, the magnitude of the vacuum chiral condensate, one ends up with attractions of similar strengths in different models.

The second, more ambitious approach is fully microscopic. Such an approach is feasible, for high-density QCD, due to asymptotic freedom. Several important results have been obtained from the microscopic approach, perhaps most notably the asymptotic form of the gap.[23] Very significant challenges remain, however. It is not really known, for example, how to calculate corrections to the leading term in a systematic way.

These approaches have complementary virtues – simplicity versus rigor, openness to phenomenological input versus quantitative predictive power at asymptotically high density. Fortunately, they broadly agree in their main conclusions as to the patterns of symmetry breaking and the magnitude of the gap at accessible densities of order $\mu = 400 - 500$ MeV.

In a field-theoretic approach to fermion pairing, the relevant quantity is the fermion self energy, i.e. the one-particle irreducible (1PI) Green function of two quark fields. Its poles give gaps, namely the gauge-invariant masses of the quasiquarks, the lowest energy fermionic excitations around the quark Fermi surface. To see if quark condensation occurs in some channel, one writes down a self-consistency equation, the gap equation, for a self energy with that structure, and solves it to find the actual self energy (the gap). If it is zero, there is no condensation in that channel. If not, there can be condensation, but it may just be a local minimum of the free energy. There may be other solutions to the gap equation, and the one with the lowest free energy is the true ground state.

There are several possible choices for the interaction to be used in the gap equation. At asymptotically high densities QCD is weakly coupled, so one gluon exchange is appropriate. Such calculations[23–36] are extremely important, since they demonstrate from first principles that color superconductivity occurs in QCD. However, the density regime of physical interest for neutron stars is several to ten times nuclear density ($\mu \sim 400 - 500$ MeV) and weak coupling calculations are unlikely to be trustworthy in that regime. In fact, current weak-coupling calculations cannot be extrapolated below about 10^8 MeV because of gauge dependence arising from the neglect of vertex corrections.[34]

The alternative is to use some phenomenological interaction that can be argued to capture the essential physics of QCD in the regime of interest. The interaction can be normalized to reproduce known low-density physics such as the vacuum chiral condensate, and then extrapolated to the desired chemical potential. In two-flavor theories, the instanton vertex is a natural choice,[8,9,16,19] since it is a four-fermion interaction. With more flavors, the one gluon exchange vertex without a gluon propagator[7,12,37] is more convenient. It has been found that these both give the same results, to within a factor of 2 or less. This is well within the inherent uncertainties of such phenomenological approaches. One caveat to bear in mind is that the single-gluon-exchange interaction is symmetric under $U(1)_A$, and so it sees no distinction between condensates of the form $\langle qCq\rangle$ and $\langle qC\gamma_5 q\rangle$ (C is the Dirac charge conjugation matrix). However, once instantons are included the Lorentz scalar $\langle qC\gamma_5 q\rangle$ is favored,[8,9] so in single-gluon-exchange calculations the parity-violating condensate is usually ignored.

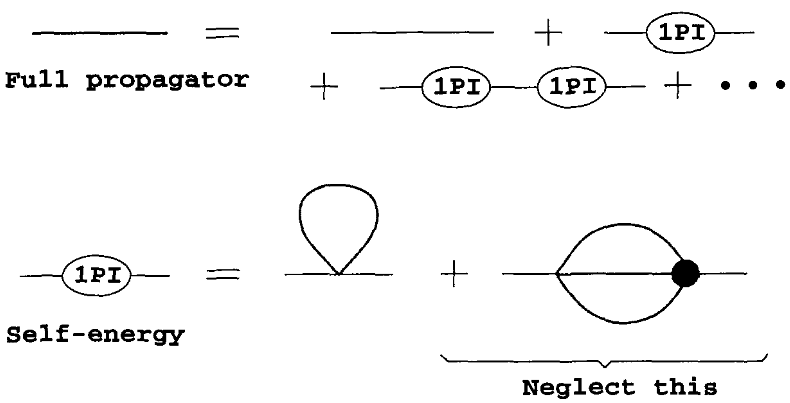


Figure 1. Mean-field Schwinger-Dyson (gap) equations

The mean-field approximation to the Schwinger-Dyson equations is shown diagramatically in Fig. 1, relating the full propagator to the self-energy. In the mean-field approximation, only daisy-type diagrams are included in the resummation, vertex corrections are excluded. Algebraically, the equation takes the form

$$\Sigma(k) = -\frac{1}{(2\pi)^4}\int d^4q\, S(q)D(k-q), \tag{1}$$

where $\Sigma(k)$ is the self-energy, S is the full fermion propagator, and $D(k-q)$ is the vertex, which in NJL models will be momentum-independent, but in a weak-coupling QCD calculation will include the gluon propagator and couplings. Since we want to study quark-quark condensation, we have to write propagators in a form that allows for this possibility, just as to study chiral symmetry breaking it is necessary to use 4-component Dirac spinors rather than 2-component Weyl spinors, even if there is no mass term in the action. We therefore use Nambu-Gorkov 8-component spinors, $\Psi = (\psi, \bar{\psi}^T)$, so the self-energy Σ can include a quark-quark pairing term Δ. The fermion inverse propagator S^{-1} then takes the form

$$S^{-1}(q) = S^{-1}_{\text{free}}(q) + \Sigma = \begin{pmatrix} \not{q} + \mu\gamma_0 & \gamma_0\Delta^\dagger\gamma_0 \\ \Delta & (\not{q} - \mu\gamma_0)^T \end{pmatrix}. \tag{2}$$

Equations (1) and (2) can be combined to give a self-consistency condition for Δ, the gap equation. If the interaction is a point-like four-fermion NJL interaction then the gap parameter Δ will be a color-flavor-spin matrix, independent of momentum. If the gluon propagator is included, Δ will be momentum-dependent, complicating the analysis considerably.

In NJL models, the simplicity of the model has allowed renormalization group analyses[21,22] that include a large class of four-fermion interactions, and follow their running couplings as modes are integrated out. This confirms that in QCD with two and three massless quarks the most attractive channels for condensation are those corresponding to the two-flavor superconducting (2SC) and color-flavor locked (CFL) phases studied below.

Following through the analysis outlined above, one typically finds gap equations of the form

$$1 = G\int_0^\Lambda k^2 dk\, \frac{1}{\sqrt{(k-\mu)^2+\Delta^2}}, \tag{3}$$

Here, the NJL model is specified by the four-fermion coupling G and a cutoff Λ, which may equally well be replaced by a smooth form-factor. The physics is reasonably insensitive to Λ in the following sense: if Λ is changed and G is changed simultaneously in such a way as to hold one physical observable (for example the vacuum chiral condensate calculated in the same model) fixed, then other physical observables (for example the gap Δ) change little. In the limit of small gap, the integral can be evaluated, giving

$$\Delta \sim \Lambda \exp\Bigl(-\frac{\text{const}}{G\mu^2}\Bigr). \tag{4}$$

This shows the non-analytic dependence of the gap on the coupling G. Condensation is a nonperturbative effect that cannot be seen to any order in perturbation theory. The reason it can be seen in the diagrammatic Schwinger-Dyson approach is that there is an additional ingredient: an ansatz for the form of the self energy. This corresponds to guessing the form of the ground state wavefunction in a many-body variational approach. All solutions to gap equations therefore represent possible stable ground states, but to find the favored ground state their free energies must be compared, and even then one can never be sure that the true ground state has been found, since there is always the possibility that there is another vacuum that solves its own gap equation and has an even lower free energy. When G is chosen to give a reasonable value of the vacuum chiral condensate, the gap Δ in quark matter at densities of interest (say $\mu \sim 400 - 500$ MeV) turns out to be of order 100 MeV.[8,9]

In weak-coupling QCD calculations, where the full single-gluon-exchange vertex complete with gluon propagator is used, the solution to the gap equation takes the form[5,6,23,25]

$$\Delta \sim \mu \frac{1}{g^5} \exp\Bigl(-\frac{3\pi^2}{\sqrt{2}}\frac{1}{g}\Bigr)\ , \tag{5}$$

or, making the weak-coupling expansion in the QCD gauge coupling g more explicit,

$$\ln\Bigl(\frac{\Delta}{\mu}\Bigr) = -\frac{3\pi^2}{\sqrt{2}}\frac{1}{g} - 5\ln g + \text{const} + \mathcal{O}(g). \tag{6}$$

This gap equation has two interesting features. Firstly, it does not correspond to what you would naively expect from the NJL model of single gluon exchange, in which the gluon propagator is discarded and $G \propto g^2$, yielding $\Delta \sim \exp(-1/g^2)$. The reason[6,23] is that at high density the gluon propagator has an infrared divergence at very small angle scattering, since magnetic gluons are only Landau damped, not screened. This divergence is regulated by the gap itself, weakening its dependence on the coupling.

Secondly, in (5) we have left unspecified the energy scale at which the coupling g is to be evaluated. Natural guesses range between μ and Δ. If we use $g(\mu)$ (this is pessimistic; using a lower energy scale results in significantly larger gaps) and assume $g(\mu)$ runs according to the one-loop formula $1/g^2 \sim \ln \mu$ then the exponential factor in (5) gives very weak suppression, and is in fact overwhelmed by the initial factor μ, so that the gap rises without limit at asymptotically high density, although Δ/μ shrinks to zero so that weak-coupling methods are still self-consistent. This means that color superconductivity will inevitably dominate the physics at high enough densities.

Although the value of Δ is under control asymptotically, it seems fair to say that applying asymptotic results at $\mu = 400-500$ MeV is at least as uncertain a proposition as applying estimates made using phenomenologically normalized models with point-like interactions. Nevertheless, if we take the estimates for the prefactor b provided by Schäfer and Wilczek's numerical results[28] and apply them at $\mu \sim 400$ MeV, they predict gaps of order 100 MeV. The consequent critical temperatures are related to the zero temperature gap Δ by the standard weak-coupling BCS result $T_c = 0.57\Delta$,[25,29] and are therefore of order 50 MeV. Some known corrections (like the fact that g must be evaluated at some scale lower than μ) push this estimate up while others (like the corrections described in Refs. 29, 36) push it down. And, regardless, the asymptotic calculation is of quantitative value only for $\mu \gg 10^8$ MeV.[34] It is nevertheless satisfying that two very different approaches, one using zero density phenomenology to normalize models, the other using weak-coupling methods valid at asymptotically high density, yield predictions for the gaps and critical temperatures at accessible densities which are in qualitative agreement. Neither can be trusted quantitatively for quark number chemical potentials $\mu \sim 400-500$ MeV, as appropriate for the quark matter which may occur in compact stars. Still, both methods agree that the gaps at the Fermi surface are of order tens to 100 MeV, with critical temperatures about half as large.

In subsequent sections, we shall treat the value of Δ as a parameter whose order of magnitude is understood but whose precise value must be determined some day either by a lattice calculation or by astrophysical observation. Instead of describing calculational methods in further detail, we focus on questions like who pairs with whom, what symmetries are broken, what are the quantum numbers of the excitations of the resulting phases, and what are the physical properties of these condensed matter phases of QCD.

Before leaving Δ, though, one comment on its magnitude: the ratio of the critical temperature to the Fermi energy in a color superconductor is $T_c/E_F = T_c/\mu \sim 0.1$, which is three orders of magnitude larger than in a traditional metallic BCS superconductor. This reflects the fact that color superconductivity is induced by an attraction due to the primary, strong, interaction in the theory, rather than having to rely on much weaker secondary interactions, as in phonon mediated superconductivity in metals. Quark matter is a high-T_c superconductor by any reasonable definition. It is unfortunate that its T_c is nevertheless low enough that it is unlikely the phenomenon can be realized in heavy ion collisions.

4. Color-Flavor Locking

In this Section we shall analyze the high-density, zero-temperature behavior of a slight idealization of QCD, in which the masses of the u, d and s quarks are set to zero, and those of the c, b and t quarks to infinity. This idealization gives rise to an especially clear and beautiful form of the theory. Also, as we shall discuss below, the analysis applies with only minor modifications to real-world QCD at high enough density. We concentrate on the physical properties of dense quark matter in the idealized three-flavor world. For a more quantitative discussion of the calculation of the magnitude of the gap than that provided in the previous section, the reader should consult the literature cited in that section. The attractive interaction between quarks, even if arbitrarily weak, renders the Fermi surface unstable to the formation of a condensate of quark Cooper pairs. We expect those quark quasiparticles which interact with the condensate to acquire an energy gap, and we expect a Meissner effect to occur for all gauge bosons except those which see the condensate as neutral.

4.1. *Form of the Condensate*

The most attractive color channel for quark pairing is the $\bar{\mathbf{3}}$. This is true at weak coupling, where single gluon exchange dominates, at intermediate coupling where instanton interaction becomes important, and at strong coupling, where the color $\bar{\mathbf{3}}$ combines two flux strings into one, lowering the energy of the pair. The relevant gap equation has been studied for pointlike 4-fermion interactions with the index structure of single gluon exchange[12,13,20] as well as a weakly coupled gluon propagator.[31,32,33] They agree that the true ground state contains nonzero condensates approximately of the form[12]

$$\langle \psi^{a\alpha}_{iL}(\vec{p})\psi^{b\beta}_{jL}(-\vec{p})\epsilon_{ab}\rangle \;=\; -\,\langle \psi(\vec{p})^{a\alpha}_{iR}\psi^{b\beta}_{jR}(-\vec{p})\epsilon_{ab}\rangle \;=\; \Delta(p^2)\epsilon^{\alpha\beta A}\epsilon_{ijA}\,. \qquad (7)$$

We have explicitly displayed color (α, β), flavor (i, j), and spinor (a, b) indices. The A-index is summed and therefore links color and flavor. We have used a two-component spinor notation; note that properly the right-helicity fields should involve dotted spinors. The important information conveyed by the spinors is that the condensation does not violate rotational invariance. The relative minus sign between left-helicity and right-helicity condensates signifies that the ground state is a scalar, rather than a pseudoscalar, so that parity is unbroken.

In reality, condensation in the color $\bar{\mathbf{3}}$ channel (7) induces a small but nonzero condensate in the color $\mathbf{6}$ channel even if the interaction is repulsive in this channel,[12] because this additional condensation breaks no further symmetries.[17] This means that the right hand side of (7) is slightly more complicated and should, in fact, be written in terms of two gap parameters κ_1 and κ_2,

$$\begin{aligned}\langle \psi^{a\alpha}_{iL}(\vec{p})\psi^{b\beta}_{jL}(-\vec{p})\varepsilon_{ab}\rangle &= -\langle \psi(\vec{p})^{a\alpha}_{iR}\psi^{b\beta}_{jR}(-\vec{p})\epsilon_{ab}\rangle \\ &= \kappa_1(p^2)\delta^{\alpha}_{i}\delta^{\beta}_{j} + \kappa_2(p^2)\delta^{\alpha}_{j}\delta^{\beta}_{i}\end{aligned} \qquad (8)$$

The pure color $\bar{\mathbf{3}}$ condensate displayed in (7) has $\kappa_2 = -\kappa_1$. Using (7) is a good approximation because the induced color $\mathbf{6}$ condensate is much smaller than the dominant color $\bar{\mathbf{3}}$ condensate mandated by the attraction in this channel.[12,31,32]

We now explain the term "color-flavor locking". The Kronecker delta functions in (8) link color and flavor indices. This means that the condensates transform nontrivially under separate color and flavor transformations, but remain invariant if we simultaneously rotate both color and flavor. Thus these symmetries are locked together. (Color-flavor locking is analogous to the B-phase of superfluid helium 3, where orbital and spin rotations, which are separate symmetries in a nonrelativistic system, are locked together.)

4.2. *Symmetry Breaking*

The color-flavor locked phase (7) features two condensates, one involving left-handed quarks alone and one involving right-handed quarks alone. Each is invariant only under equal and opposite color and flavor rotations, but since color is a vector symmetry, the combined effect is to break the color and flavor symmetries down to a global vector "diagonal" symmetry, that makes equal transformations in all three sectors – color, left-handed flavor, and right-handed flavor. Baryon number symmetry is broken down to a discrete Z_2 symmetry under which all quark fields are multiplied by -1. The symmetry breaking pattern is[12]

$$[SU(3)_{\text{color}}] \times \underbrace{SU(3)_L \times SU(3)_R}_{\supset [U(1)_Q]} \times U(1)_B \longrightarrow \underbrace{SU(3)_{C+L+R}}_{\supset [U(1)_{\tilde{Q}}]} \times Z_2 \tag{9}$$

where gauge symmetries are in square brackets, $SU(3)_L$ and $SU(3)_R$ are the global chiral flavor symmetries, and $U(1)_B$ is baryon number. (See also Ref. 38 in which a similar pattern was considered at zero density.)

We see that chiral symmetry has been broken, but not by the usual low-density $\langle\bar{\psi}\psi\rangle$ condensate which pairs left-handed quarks with right-handed antiquarks. Even though the condensates in (7,8) only pair left-handed quarks with left-handed quarks, and right-handed quarks with right-handed quarks, and do not *appear* to lock $SU(3)_L$ to $SU(3)_R$, they manage to do so by locking both to $SU(3)_{\text{color}}$. Color-flavor locking, therefore, provides a mechanism by which chiral symmetry can be broken.

Even without doing any detailed calculations, we can see that the symmetry breaking pattern (9) will have profound effects on the physics. We will summarize them quickly here, and investigate them in more detail below.

(1) The color gauge group is completely broken. All eight gluons become massive. This ensures that there are no infrared divergences associated with gluon propagators. Since the quark modes are also gapped (see below), we conclude that weak coupling perturbation theory *around the correct, condensed ground state* is free of the difficulties that appeared around the naive ground state.
(2) All the quark modes are gapped. This removes the other potential source of infrared divergences, from integration over low-energy excitations around the Fermi surface. The nine quasiquarks (three colors times three flavors) fall into an $\mathbf{8} \oplus \mathbf{1}$ of the unbroken global $SU(3)$, so there are two gap parameters. The singlet has a larger gap than the octet.
(3) Electromagnetism is no longer a separate symmetry, but corresponds to gauging one of the flavor generators. A rotated electromagnetism ("$\tilde{Q}$"), a combination of the original photon and one of the gluons, survives unbroken.

(4) Two global symmetries are broken, the chiral symmetry and baryon number, so there are two gauge-invariant order parameters that distinguish the CFL phase from the QGP. There are corresponding Goldstone bosons which are long-wavelength symmetry rotations of the order parameter, whose energy cost is arbitrarily low. The chiral Goldstone bosons form a pseudoscalar octet, like the zero-density $SU(3)_{\text{flavor}}$ pion octet. The breaking of the baryon number symmetry leads to a singlet scalar Goldstone boson which makes the CFL phase a superfluid, and remains massless even when quark masses are introduced.

(5) Quark-hadron continuity. It is striking that the symmetries of the 3-flavor CFL phase are the same as those one might expect for 3-flavor hypernuclear matter.[13] In hypernuclear matter one would expect the hyperons to pair in an $SU(3)_{\text{flavor}}$ singlet $(\langle\Lambda\Lambda\rangle, \langle\Sigma\Sigma\rangle, \langle N\Xi\rangle)$, breaking baryon number but leaving flavor and electromagnetism unbroken. Chiral symmetry would be broken by the chiral condensate. This means that one might be able to follow the spectrum continuously from hypernuclear matter to the CFL phase of quark matter. The pions would evolve into the pseudoscalar octet mentioned above. The vector mesons would evolve into the massive gauge bosons. This will be discussed in more detail below, but we note already now that nuclear matter (as opposed to the hypernuclear matter in a world with three degenerate quarks) *cannot* be continuously connected to the CFL phase.

4.3. *Global Symmetries and Order Parameters*

Color-flavor locking, unlike the Higgs mechanism in the electroweak sector of the standard model, breaks global symmetries as well as gauge symmetries. Physically, this implies that there are sharp differences between the color-flavor locked phase and the quark-gluon plasma phase (in which all symmetries of the QCD lagrangian are unbroken), so that any passage between them must be marked by one or more phase transitions. It is a simple matter to construct the corresponding gauge invariant order parameters, which have a strict meaning valid at any coupling, from our primary, gauge variant condensate at weak coupling. For instance, to form a gauge invariant order parameter capturing chiral symmetry breaking we may take the product of the left-handed version of (7) with the right-handed version and saturate the color indices, to obtain

$$\langle \psi^{\alpha}_{Li} \psi^{\beta}_{Lj} \bar{\psi}^{k}_{R\alpha} \bar{\psi}^{l}_{R\beta} \rangle \sim \langle \psi^{\alpha}_{Li} \psi^{\beta}_{Lj} \rangle \langle \bar{\psi}^{k}_{R\alpha} \bar{\psi}^{l}_{R\beta} \rangle \sim \Delta^2 \epsilon_{ijm} \epsilon^{klm} \ . \tag{10}$$

Likewise we can take a product of three copies of the condensate and saturate the color indices, to obtain a gauge invariant order parameter for the baryon-number violating superfluid order parameter. These secondary order parameters will survive gauge unfixing unscathed. Unlike the primary condensate, from which they were derived, they are more than convenient fictions.

The spontaneous violation of baryon number does not mean that compact stars can change the baryon number of the universe, any more than electron pairing in ordinary superconductors can change the lepton number of the universe. Actually, it is already generally believed that ordinary nuclear matter is a superfluid in which nucleon-nucleon pairing violates baryon number symmetry. The essential point in all these cases is that in a finite sized sample there is no true violation of the conservation laws, since a Gaussian surface can be constructed enclosing the sample. The correct interpretation of the formal "violation" of these symmetries is that there can be large fluctuations and easy transport of the corresponding quantum numbers within the sample. These are precisely the phenomena of superconductivity and superfluidity.

If we turn on a common mass for all the quarks, the chiral $SU(3)_L \times SU(3)_R$ flavor symmetry of the Lagrangian will be reduced to the diagonal $SU(3)_{L+R}$. If we turn on unequal masses, the symmetry will be even less. In any case, however, the $U(1)$ of baryon number a good microscopic symmetry, and the corresponding six-quark order parameter remains a strict signature of the color-flavor locked phase, distinguishing it from the quark-gluon plasma phase.

As it stands the order parameter (10) is not quite the usual one, but roughly speaking its square. It leaves invariant an additional Z_2, under which only the left-handed quark fields change sign. Actually this Z_2 is not a legitimate symmetry of the full theory, but suffers from an anomaly. So we might expect that the usual chiral order parameter is induced by the anomalous interactions that violate the axial baryon number symmetry of the classical Lagrangian. To put this another way, because axial baryon number is not a symmetry of QCD, once chiral symmetry is broken by color-flavor locking there is no symmetry argument precluding the existence of an ordinary chiral condensate. Indeed, instanton effects do induce a nonzero $\langle \bar{\psi}_R \psi_L \rangle$ because the instanton-induced interaction is a six-fermion operator which can be written as a product of $\bar{\psi}_R\psi_L$ and the operator in (10) which already has a nonzero expectation value,[12] but this turns out to be a small effect.[20,31]

At weak coupling, we can be more specific about these matters. The most important interactions near the Fermi surface, quantitatively, arise from gluon exchange. These are responsible for the primary condensation. The instanton interaction is much less important asymptotically because the gauge fields which make up the instantons themselves are screened, the effects of instantons are intrinsically smaller for more energetic quarks, and because the instanton-induced interaction involves six fermion fields, and hence (one can show) becomes irrelevant upon renormalization toward the Fermi surface. The instanton interaction is qualitatively important, however, because it represents the leading contribution to axial baryon number violation. It is only such $U(1)_A$ violating interactions that remove the degeneracy among states with different relative phases between the left- and right-handed condensates in (7). In the absence of intrinsic $U(1)_A$ breaking, the spontaneous violation of this symmetry in the color-flavor locked phase would be accompanied by the existence of a pseudoscalar $SU(3)_{\mathrm{color}+L+R}$ singlet Nambu-Goldstone bosons. Since the intrinsic violation of this symmetry is parametrically small, the corresponding boson will not be strictly massless, but only very light. Quantum fluctuations in this light η'-field, among other things, will keep the conventional chiral symmetry breaking order parameter small compared to (10) at high density.

4.4. *Elementary Excitations*

The physics of the excitations in the CFL phase has been the focus of much recent work.[12,13,17,20,31–33,39–58] There are three sorts of elementary excitations. They are the modes produced directly by the fundamental quark and gluon fields, and the collective modes associated with spontaneous symmetry breaking. These modes can be classified under the unbroken $SU(3) \times Z_2$ symmetry, and the unbroken rotation and parity symmetries.

The quark fields of course produce spin 1/2 fermions. Some of these are true long-lived quasiparticles, since there are no lighter states of half-integer spin that they might decay into. With the conventions we have been using, as expressed in (7), the quark fields are triplets and antitriplets under color and flavor, respectively. Thus they decompose into an octet and a singlet under the diagonal $SU(3)_{\text{color}+L+R}$. There is an energy gap for production of pairs above the ground state. More precisely, there are two gaps: a smaller one for the octet, and a larger one for the singlet.[12] The dispersion relations describing these fermionic quasiparticle excitations in the CFL phase have been described in some detail.[17,41,55]

The gluon fields produce an $SU(3)_{\text{color}+L+R}$ octet of spin 1 bosons. As previously mentioned, they acquire a common mass by the Meissner-Anderson-Higgs mechanism. The quantitative expressions for the masses of these vector mesons which have been computed at weak coupling[42,49,55,59] and in an instanton-liquid model.[60]

The fermionic excitations have a gap; the vector mesons have mass; but, the Nambu-Goldstone bosons are massless. These bosons form a pseudoscalar octet associated with chiral symmetry breaking, and a scalar singlet associated with baryon number superfluidity. The octet, but not the singlet, will be lifted from zero mass if the quarks are massive. Finally there is the parametrically light, but never strictly massless, pseudoscalar singlet associated with $U(1)_A$ breaking.

The Nambu-Goldstone bosons arising from chiral symmetry breaking in the CFL phase are Fermi surface excitations in which the orientation of the left-handed and right-handed diquark condensates oscillate out of phase in flavor space. The effective field theory describing these oscillations has been constructed.[40,42,47] Up to two derivatives, it is given by

$$\mathcal{L}_{\text{eff}} = \frac{f_\pi^2}{4} \mathrm{Tr}\left(\partial_0\Sigma\partial_0\Sigma^\dagger + v_\pi^2 \partial_i\Sigma\partial^i\Sigma^\dagger\right) - c\left(\det M \,\mathrm{Tr}(M^{-1}\Sigma) + \text{h.c.}\right) \; . \tag{11}$$

The Nambu-Goldstone boson field matrix Σ is a color singlet and transforms under $SU(3)_L \times SU(3)_R$ as $\Sigma \to U_L \Sigma U_R^\dagger$ as usual. $M = \text{diag}(m_u, m_d, m_s)$ is the quark mass matrix. The construction of Σ from rotations of the CFL condensates can be found in Ref. 40, 42: one first finds the 17 putative Nambu-Goldstone bosons expected when the full symmetry group is broken to $SU(3)_{C+L+R}$; one then identifies 8 of these which become the longitudinal parts of massive vector bosons; the remaining 9 are the octet described by (11), and the superfluid mode. In addition, although the singlet η' is not a true Goldstone boson it is very light, as discussed above. See Ref. 40, 42 for the singlet terms in the effective Lagrangian. The higher derivative terms in the effective Lagrangian have also been analyzed.[47]

The masses of the pseudoscalar mesons which are the pseudo-Nambu-Goldstone bosons associated with chiral symmetry breaking can be obtained from $\mathcal{L}_{\text{eff}}$ of (11).[42] For example,

$$m^2_{\pi^\pm} = \frac{2c}{f^2_\pi} m_s(m_u + m_d) \ , \qquad m^2_{K^\pm} = \frac{2c}{f^2_\pi} m_d(m_u + m_s) \ . \tag{12}$$

Thus, the kaon is lighter than the pion, by a factor of $m_d/(m_u + m_d)$.[42] Note that the effective Lagrangian is quadratic in M. This arises because L_{eff} respects the Z_2 symmetry under which only the left-handed quarks change sign.[12] As we discussed in the previous section, this is almost a symmetry of the CFL phase: it would be a symmetry if instanton effects could be neglected.[12] However, instanton effects generate a nonzero, but small, ordinary $\langle\bar{\psi}_R\psi_L\rangle$ condensate, which breaks the Z_2,[12,20,31] and results in a contribution to the meson m^2 which is linear in M and which may be numerically significant.[45,58]

If we were describing pions in vacuum, or pions in nuclear matter, the only way to obtain the coefficients in the effective theory would be to measure them in an experiment or, if possible, to calculate them on the lattice. Indeed in any theory with strong interactions, the purpose of writing an effective theory for the low energy degrees of freedom is to express the predictions for many low energy processes in terms of a few parameters, which must be obtained from experiment. In the color-flavor locked phase, however, the full theory is weakly coupled at asymptotically high densities. In this regime, therefore, the coefficients f^2_π, v^2_π and c are calculable from first principles using weak coupling methods! Up to possible logarithmic corrections, the result is[42–46,48,54,55,58]

$$f^2_\pi = \frac{21 - 8\log 2}{36\pi^2}\mu^2 \ , \quad v^2_\pi = \frac{1}{3} \ , \quad c = \frac{3\Delta^2}{2\pi^2} \ . \tag{13}$$

The electromagnetic[50,54] and nonzero temperature[54] corrections to these quantities have also been calculated. Recently, the instanton contributions to the CFL meson masses have been estimated; they may result in a significant increase.[61]

Quantitatively, (see Section 3 for a discussion of estimates of Δ) we estimate that the lightest pseudoscalar meson, the kaon, has a mass in the range of 5 to 20 MeV at $\mu = 400$ MeV, and becomes lighter still at higher densities. There are two reasons why the Nambu-Goldstone bosons are so much lighter in the CFL phase than in the vacuum. First, their mass^2 is proportional to m^2_{quark} rather than to m_{quark}, as at zero density. In addition, there is a further suppression by a factor of Δ/μ, which arises because the Nambu-Goldstone bosons are collective excitations of the condensates formed from particle-particle and hole-hole pairs near the Fermi surface, whereas the quark mass term connects particles with antiparticles, far from the Fermi surface.[44]

In QCD with unequal quark masses, at very high densities the CFL phase is much as we have described it, except that the gaps associated with $\langle us \rangle$, $\langle ds \rangle$ and $\langle ud \rangle$ pairing will differ slightly[17,39] and the CFL condensate may rotate in flavor space yielding a kaon condensate.[56] The formation of a kaon condensate in the CFL phase is a less dramatic effect than the formation of a kaon condensate in nuclear matter made of neutrons and protons only:[62] there, kaon condensation breaks $U(1)_S$. This symmetry is already broken in the CFL phase. Kaon condensation in the CFL phase is more akin to kaon condensation in hypernuclear matter made up of equal measures of all the octet baryons, in which $U(1)_S$ is already broken by hyperon-hyperon pairing. Remarkably, the Fermi momenta for the different quarks in the CFL phase remain degenerate even in the presence of a strange quark mass because separating them costs more pairing energy than it yields. This means that the CFL phase is electrically neutral in the absence of electrons.[63]

At high enough baryon density and low temperature, the ground state of QCD with three flavors of quarks is the color-flavor locked (CFL) phase. In this phase, quarks of all three colors and all three flavors form Cooper pairs, meaning that all fermionic quasiparticles are gapped. The gap Δ is likely of order tens to 100 MeV at astrophysically accessible densities, with quark chemical potential $\mu \sim (350-500)$ MeV. As we shall see momentarily, the condensate is charged with respect to eight of the nine massless gauge bosons of the ordinary vacuum, meaning that eight gauge bosons get a mass. Chiral symmetry is spontaneously broken, and so is baryon number (i.e., the material is a superfluid.) The CFL phase persists for finite masses and even for unequal masses, so long as the differences are not too large.[17,39] It is very likely the ground state for real QCD, assumed to be in equilibrium with respect to the weak interactions, over a substantial range of densities. Throughout the range of parameters over which the CFL phase exists as a bulk (and therefore electrically neutral) phase, it consists of equal numbers of u, d and s quarks and is therefore electrically neutral in the absence of any electrons.[63] The equality of the three quark number densities is enforced in the CFL phase by the fact that this equality maximizes the pairing energy associated with the formation of ud, us, and ds Cooper pairs. This equality is enforced even though the strange quark, with mass m_s, is heavier than the light quarks.

4.5. *The Modification of Electromagnetism*

It is physically significant, and proves extremely instructive, to consider the effect of color-flavor locking on the electromagnetic properties of high-density hadronic matter. To do this, we consider coupling in the appropriate additional $U(1)_{\mathrm{EM}}$ gauge field A_μ, representing the photon. This couples to u, d, s quarks with strength $\frac{2}{3}e, -\frac{1}{3}e, -\frac{1}{3}e$, respectively. Evidently this $U(1)_{\mathrm{EM}}$ symmetry is broken by the condensate (7), through the terms pairing differently-charged quarks. Were this the complete story, the color-flavor locked phase would be an electromagnetic superconductor. The truth is far different, however.

The situation is analogous to what occurs in the electroweak sector of the standard model. There, the Higgs field condensate breaks both the original weak $SU(2)$ and the hypercharge $U(1)$. However, one linear combination of these symmetries leaves the condensate invariant, and remains a valid gauge symmetry of the ground state. Indeed, this is how we identify electromagnetism within the standard model.

Here we must similarly consider the possibility of mixing color $SU(3)$ and electromagnetic $U(1)_{\mathrm{EM}}$ generators to find a valid residual symmetry. Indeed, we should expect this to occur, by the following argument. In QCD with three flavors, $U(1)_{\mathrm{EM}}$ is a subgroup of $SU(3)_{L+R}$. When we neglected electromagnetism, we found that in the color-flavor locked phase $SU(3)_{L+R}$ is broken but $SU(3)_{\mathrm{color+L+R}}$ is an unbroken global symmetry. We therefore expect that gauging a $U(1)$ subgroup of $SU(3)_{L+R}$ must correspond, in the CFL phase, to gauging a $U(1)$ subgroup of the unbroken $SU(3)_{\mathrm{color+L+R}}$.

Once we are alerted to this possibility, it is not difficult to identify the appropriate combination of the photon and gluons which remains unbroken.[12,64] In the CFL phase, there is an unbroken $U(1)_{\tilde{Q}}$ gauge symmetry and a corresponding massless photon given by a linear combination of the ordinary photon and one of the gluons.[12,64] $U(1)_{\tilde{Q}}$ is generated by

$$\tilde{Q} = Q + \eta T_8$$

with $\eta = 1/\sqrt{3}$ and where Q is the conventional electromagnetic charge generator and the color hypercharge generator T_8 is normalized such that, in the representation of the quarks, $T_8/\sqrt{3} = \mathrm{diag}(-\frac{2}{3}, \frac{1}{3}, \frac{1}{3})$ in color space. The CFL condensate is $\tilde{Q}$-neutral, the $U(1)$ symmetry generated by $\tilde{Q}$ is therefore unbroken, the associated $\tilde{Q}$-photon remains massless, and within the CFL phase the $\tilde{Q}$-electric and $\tilde{Q}$-magnetic fields satisfy Maxwell's equations. The massless combination of the photon and the eighth gluon, $A_\mu^{\tilde{Q}}$, and the orthogonal massive combination which experiences the Meissner effect, A_μ^X, are given by

$$A_\mu^{\tilde{Q}} = \cos\theta A_\mu + \sin\theta G_\mu^8 \,, \tag{14}$$
$$A_\mu^X = -\sin\theta A_\mu + \cos\theta G_\mu^8 \,. \tag{15}$$

The mixing angle θ (called α_0 in Ref. 64) which specifies the unbroken $U(1)$ is given by

$$\cos\theta = \frac{g}{\sqrt{g^2+\eta^2 e^2}} \,. \tag{16}$$

θ is the analogue of the Weinberg angle in electroweak theory. At accessible densities, the gluons are strongly coupled ($g^2/4\pi \sim 1$) and the photons are weakly coupled ($e^2/4\pi \approx 1/137$), so θ is small, perhaps of order 1/20. The "rotated photon" consists mostly of the usual photon, with only a small admixture of the G^8 gluon.

Let us now consider the charges of all the elementary excitations which we enumerated previously. For reference, the electron couples to $A_\mu^{\tilde{Q}}$ with charge

$$\tilde{e} = \frac{eg}{\sqrt{\eta^2 e^2 + g^2}} \,. \tag{17}$$

which is less than e because the electron couples only to the A_μ component of $A_\mu^{\tilde{Q}}$. Now in computing the $\tilde{Q}$-charge of the quark with color and flavor indices α, a we must take the appropriate combination from

$$\frac{e(-\frac{2}{3}g, \frac{1}{3}g, \frac{1}{3}g) + g(\frac{2}{3}e, -\frac{1}{3}e, -\frac{1}{3})}{\sqrt{\eta^2 e^2 + g^2}} \,.$$

One readily perceives that the possible values are $0, \pm\tilde{e}$. Thus, in units of the electron charge, the quarks carry integer $\tilde{Q}$-charge! Quite remarkably, high-density QCD realizes a mathematically consistent gauge theory version of the old vision of Han and Nambu: the physical quark excitations have integer electric charges that depend on an internal color quantum number!

Similarly, the gluons all have $\tilde{Q}$-charges $0, \pm\tilde{e}$. Indeed, they have the $\tilde{Q}$-charges one would expect for an octet of massive vector bosons. The Nambu-Goldstone bosons arising from the breaking of chiral symmetry, of course, have the same charge assignments as the familiar π, K and η octet of pseudoscalars. The baryon superfluid mode is $\tilde{Q}$-neutral. In the color-flavor locked phase, we conclude, all the elementary excitations are integrally charged.[a] This is a classic aspect of confinement, here embodied in a controlled, weak-coupling framework.

All elementary excitations in the CFL phase are either $\tilde{Q}$-neutral or couple to $A_\mu^{\tilde{Q}}$ with charges which are integer multiples of the $\tilde{Q}$-charge of the electron $\tilde{e} = e\cos\theta$, which is less than e because the electron couples only to the A_μ component of $A_\mu^{\tilde{Q}}$. The only massless excitation (the superfluid mode) is $\tilde{Q}$-neutral. Because all charged excitations have nonzero mass and there are no electrons present,[63] the CFL phase in bulk is a transparent insulator at low temperatures: $\tilde{Q}$-magnetic and $\tilde{Q}$-electric fields within it evolve simply according to Maxwell's equations, and low frequency $\tilde{Q}$-light traverses it without scattering.

It is fun to consider how a chunk of our color-flavor locked material would look. Imagine shining light on a chunk of dense quark matter in the CFL phase. If CFL matter occurs only within the cores of neutron stars, cloaked under kilometers of hadronic matter,[65] the thought experiment we describe here in which light waves travelling in vacuum strike CFL matter can never arise in nature. If, however, the fact that quark matter features many more strange quarks than ordinary nuclear matter renders it stable even at zero pressure, then one may imagine quark stars in nature.[66] Such a quark star may be made of CFL quark matter throughout, or may have an outer layer in which a less symmetric pattern of pairing occurs. For example, quarks of only two colors and flavors may pair, yielding the 2SC phase which we shall discuss in the next section. (Some of the remaining quarks with differing Fermi momenta may also form a crystalline color superconductor, which we shall discuss below.) As in the CFL phase, the 2SC condensate leaves a (slightly different) $\tilde{Q}$-photon massless. However, the 2SC phase is a good $\tilde{Q}$-conductor because of the presence of unpaired quarks and electrons. Thus, 2SC matter is opaque and metallic rather than transparent and insulating. Illuminating it would result in absorption and reflection, but no refraction. We shall assume that the quark matter we illuminate is in the transparent CFL phase all the way out to its surface.

Consider, then, an enormously dense, but transparent, illuminated quark star. Some light falling on its surface will reflect, and some will refract into the star in the form of $\tilde{Q}$-light. The reflection and refraction

---

[a] We shall see below that in two-flavor QCD, in which color-flavor locking does not occur, the color superconducting condensate which forms also leaves a $\tilde{Q}$-photon massless. The only difference relative to the CFL phase is that $\eta = -1/2\sqrt{3}$. (However, the $\tilde{Q}$-charges of the excitations are not all integral in this theory.)

angles and the intensity of the reflected light and refracted $\tilde{Q}$-light have all been calculated.[67] The partial Meissner effect induced by a static magnetic field had been analyzed previously.[64] In Ref. 67, we analyze a time-varying electromagnetic field. As a bonus, our analysis allows us to use well understood properties of dense quark matter in the CFL phase to learn about the (less well understood) QCD vacuum.

We assume that the light has ω and k both much less than the energy needed to create a charged excitation in the CFL phase. This means $\omega, k \ll \Delta$, where Δ is the fermionic gap, to avoid the breaking of pairs and the creation of quasiparticles. It also means $\omega, k \ll m_{\pi^\pm}, m_{K^\pm}$, where $\pi^\pm$ and $K^\pm$ are the charged pions and kaons of the CFL phase.

In vacuum the electromagnetic fields obey the free Maxwell's equations

$$\nabla \cdot \mathbf{D} = 0 \, , \qquad \nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \, , \tag{18}$$

$$\nabla \cdot \mathbf{B} = 0 \, , \qquad \nabla \times \mathbf{H} = \frac{\partial \mathbf{D}}{\partial t} \, , \tag{19}$$

where $\mathbf{D} = \epsilon_0 \mathbf{E}$ and $\mathbf{B} = \mu_0 \mathbf{H}$, and ϵ_0 and μ_0 are the vacuum dielectric constant and magnetic permeability, respectively, such that the velocity of light $c = 1/\sqrt{\mu_0 \epsilon_0}$. Deep in the CFL phase, the rotated fields $\tilde{\mathbf{E}}$ and $\tilde{\mathbf{B}}$ obey the same field equations, but with dielectric constant[68]

$$\tilde{\epsilon} = \epsilon_0 \left(1 + \frac{8\alpha}{9\pi} \cos^2\theta \frac{\mu^2}{\Delta^2} \right) \, , \tag{20}$$

where α is the electromagnetic fine structure constant and μ is the chemical potential. This expression for $\tilde{\epsilon}$ is valid to leading order in α, and for $\omega, k \ll \Delta$. The dependence of $\tilde{\epsilon}$ on ω arises only in corrections to (20) which are suppressed by ω^2/Δ^2, and we therefore neglect dispersion in this letter. The magnetic permeability in the CFL phase remains unchanged to leading order, $\tilde{\mu} = \mu_0$. The index of refraction of CFL quark matter thus reduces to $\tilde{n} = \sqrt{\tilde{\mu}\tilde{\epsilon}/\mu_0\epsilon_0} = \sqrt{\tilde{\epsilon}/\epsilon_0}$. If we apply (20) for $\mu/\Delta \sim (4 - 10)$, we obtain $\tilde{n} \sim (1.02 - 1.1)$.

We take the surface of the CFL matter to be planar, with the CFL phase at $z > 0$ and vacuum at $z < 0$. (That is, we assume any curvature of the surface is on length scales long compared to the wavelength of the light.) For an ordinary dielectric, the analogous problem is solved in Ref. 69. The complication here is that we must match the ordinary electric and magnetic fields in vacuum onto $\tilde{Q}$-electric and $\tilde{Q}$-magnetic fields within the CFL phase. The properties of the reflected and refracted waves will therefore depend upon both the dielectric constant $\tilde{\varepsilon}$ and the mixing angle θ.

We are only interested in the reflected and refracted waves, and not in the detailed field configurations very close to the interface. This means that we can follow the strategy of Ref. 64 and encapsulate the physics of the interface into boundary conditions relating $\mathbf{E}$ and $\mathbf{B}$ on the vacuum side of the interface to $\tilde{\mathbf{E}}$ and $\tilde{\mathbf{B}}$ on the CFL side. On the CFL side, the massive X fields can be neglected as long as z is greater than some λ^{CFL}, while on the vacuum side, the confined gluon fields can be neglected as long as $|z|$ is greater than some λ^{QCD}. λ^{QCD} is a length scale characteristic of confinement. For the non-static fields of interest, and in the weak coupling regime, λ^{CFL} is of order $1/\Delta$, longer than the inverse Meissner mass $\sim g\mu$. In order to describe light whose wavelength is long compared to λ^{QCD} and λ^{CFL}, we need boundary conditions relating $\mathbf{E}$ and $\mathbf{B}$ at $z = -\lambda^{\mathrm{QCD}}$ to $\tilde{\mathbf{E}}$ and $\tilde{\mathbf{B}}$ at $z = +\lambda^{\mathrm{CFL}}$.

X-magnetic fields experience a Meissner effect in the CFL phase, meaning that supercurrents in the CFL matter within λ^{CFL} of the interface screen the X-component of any ordinary magnetic field parallel to the interface on the vacuum side, yielding the boundary condition

$$\tilde{\mathbf{H}}_{\parallel}(t,x,y,\lambda^{\mathrm{CFL}}) = \cos\theta\, \mathbf{H}_{\parallel}(t,x,y,-\lambda^{\mathrm{QCD}}) \ . \tag{21}$$

The CFL condensate is charged with respect to the X gauge boson, meaning that if there is an ordinary electric field perpendicular to the interface on the vacuum side, the X component of the electric flux will terminate in the CFL phase within λ^{CFL} of the interface, yielding

$$\tilde{\mathbf{D}}_{\perp}(t,x,y,\lambda^{\mathrm{CFL}}) = \cos\theta\, \mathbf{D}_{\perp}(t,x,y,-\lambda^{\mathrm{QCD}}) \ . \tag{22}$$

We expect that the confined QCD vacuum should behave as if it is a condensate of color-magnetic monopoles.[70] That is, in the vacuum color magnetic field lines end: if there is a $\tilde{Q}$-magnetic field perpendicular to the interface on the CFL side, the vacuum will ensure that only the ordinary magnetic field is admitted. Thus,

$$\mathbf{B}_{\perp}(t,x,y,-\lambda^{\mathrm{QCD}}) = \cos\theta\, \tilde{\mathbf{B}}_{\perp}(t,x,y,\lambda^{\mathrm{CFL}}) \ . \tag{23}$$

Finally, color magnetic currents on the vacuum side of the interface should exclude the color component of any $\tilde{Q}$-electric field parallel to the interface on the CFL side, ensuring that

$$\mathbf{E}_{\parallel}(t,x,y,-\lambda^{\mathrm{QCD}}) = \cos\theta\, \tilde{\mathbf{E}}_{\parallel}(t,x,y,\lambda^{\mathrm{CFL}}) \ . \tag{24}$$

At sufficiently high density, the property of CFL matter from which (21) and (22) follow, namely the Meissner effect for X-bosons, is a weak-coupling phenomenon which can be understood analytically. The properties of the QCD vacuum used to deduce (23) and (24) follow from a reasonable and familiar description of confinement as a dual Meissner effect, but confinement is not yet understood analytically. It is therefore of interest that our analysis below provides a *derivation* of (23) and (24) from (21).

Consider an incident wave, with wave vector $\mathbf{k} = \frac{\omega}{c}(\sin i, 0, \cos i)$, so that

$$\mathbf{E} = \mathcal{E}e^{i(\mathbf{k}\cdot\mathbf{r}-\omega t)} \ , \qquad \mathbf{B} = \sqrt{\mu_0\epsilon_0}\,\frac{\mathbf{k}}{k} \times \mathbf{E} \ . \tag{25}$$

There are two orthogonal linear polarizations, shown in Fig. 2, which we will treat separately. In the first, the vector $\mathcal{E}$ is parallel to the interface while in the second, $\mathcal{E}$ lies in the plane of incidence. The reflected wave is

$$\mathbf{E}' = \mathcal{E}'e^{i(\mathbf{k}'\cdot\mathbf{r}-\omega t)} \ , \qquad \mathbf{B}' = \sqrt{\mu_0\epsilon_0}\,\frac{\mathbf{k}'}{k'} \times \mathbf{E}' \ , \tag{26}$$

with wave vector $\mathbf{k}' = \frac{\omega}{c}(\sin i', 0, -\cos i')$, while the refracted wave is

$$\tilde{\mathbf{E}}_r = \tilde{\mathcal{E}}_r e^{i(\mathbf{k_r}\cdot\mathbf{r}-\omega t)} \ , \qquad \tilde{\mathbf{B}}_r = \sqrt{\tilde{\mu}\tilde{\epsilon}}\,\frac{\mathbf{k_r}}{k_r} \times \tilde{\mathbf{E}}_r \ , \tag{27}$$

with wave vector $\mathbf{k_r} = \omega\,\sqrt{\tilde{\mu}\tilde{\epsilon}}(\sin r, 0, \cos r)$.

The boundary conditions must be obeyed at all times, which immediately implies that the frequency of all the waves is the same, as above. The boundary conditions must be obeyed at all points on the planar interface. For $1/k \gg \lambda^{\mathrm{CFL}}, \lambda^{\mathrm{QCD}}$ this implies that $\mathbf{k}\cdot\mathbf{r} = \mathbf{k}'\cdot\mathbf{r} = \mathbf{k_r}\cdot\mathbf{r}$ at $z = 0$, independent of details of the boundary conditions. To satisfy this kinematic constraint, all three wave vectors must lie in a plane and $k\sin i = k'\sin i' = k_r \sin r$. Since $k = k'$, we must have $i = i'$: that is, the angle of incidence is the same as the angle of reflection. Since $k_r = \tilde{n}k$, we also reproduce Snell's law

$$\sin i = \tilde{n}\sin r \ . \tag{28}$$

The kinematics of the reflection and refraction of light on CFL quark matter are unaffected by the mixing angle θ.

We now use the boundary conditions to find the intensities of the reflected and refracted radiation. For the first polarization in Fig. 2, (21) and

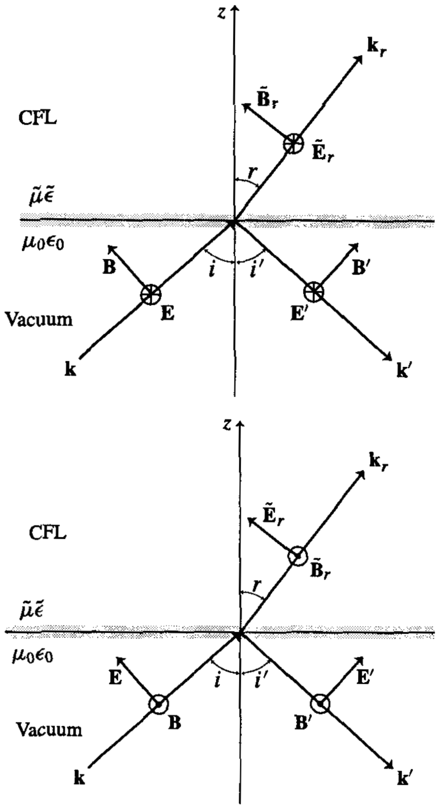


Figure 2. Incident wave $\mathbf{k}$ strikes a planar interface between vacuum and CFL quark matter, giving a reflected wave $\mathbf{k}'$ and a refracted wave $\mathbf{k_r}$. We use boundary conditions to relate electromagnetic fields just below the interface to $\tilde{Q}$-electromagnetic fields just above, assuming that the wavelength of the light is long compared to the screening length in the CFL phase and the confinement length in vacuum, symbolized by grey shading. Left panel: polarization perpendicular to the plane of incidence and thus parallel to the interface. Right panel: polarization parallel to the plane of incidence.

(24) yield

$$\begin{aligned} \cos\theta\,(\mathcal{E}-\mathcal{E}')\sqrt{\frac{\epsilon_0}{\mu_0}}\cos i &= \tilde{\mathcal{E}}_r\sqrt{\frac{\tilde{\epsilon}}{\tilde{\mu}}}\cos r\ , \\ \mathcal{E}+\mathcal{E}' &= \cos\theta\,\tilde{\mathcal{E}}_r\ , \end{aligned} \tag{29}$$

and, using Snell's law, (23) is equivalent to (24) in this case. Solving, we find

$$\frac{\tilde{\mathcal{E}}_r}{\mathcal{E}} = \frac{2\cos\theta\,\cos i}{\cos^2\theta\,\cos i + \frac{\mu_0}{\tilde{\mu}}\,\tilde{n}\cos r}\,,$$
$$\frac{\mathcal{E}'}{\mathcal{E}} = \frac{\cos^2\theta\,\cos i - \frac{\mu_0}{\tilde{\mu}}\,\tilde{n}\cos r}{\cos^2\theta\,\cos i + \frac{\mu_0}{\tilde{\mu}}\,\tilde{n}\cos r}\,, \tag{30}$$

where r is easily eliminated using Snell's law in the form $\tilde{n}\cos r = \sqrt{\tilde{n}^2 - \sin^2 i}$. To the order we are working, $\tilde{\mu} = \mu_0$. For the second polarization of Fig. 2, (24) and either (21) or (22) yield

$$(\mathcal{E} - \mathcal{E}')\cos i = \cos\theta\,\tilde{\mathcal{E}}_r\cos r\,,$$
$$\cos\theta\sqrt{\frac{\epsilon_0}{\mu_0}}\,(\mathcal{E} + \mathcal{E}') = \sqrt{\frac{\tilde{\epsilon}}{\tilde{\mu}}}\tilde{\mathcal{E}}_r\,, \tag{31}$$

and hence

$$\frac{\tilde{\mathcal{E}}_r}{\mathcal{E}} = \frac{2\tilde{n}\cos\theta\cos i}{\frac{\mu_0}{\tilde{\mu}}\,\tilde{n}^2\cos i + \tilde{n}\cos r\cos^2\theta}\,,$$
$$\frac{\mathcal{E}'}{\mathcal{E}} = \frac{\frac{\mu_0}{\tilde{\mu}}\,\tilde{n}^2\cos i - \tilde{n}\cos r\cos^2\theta}{\frac{\mu_0}{\tilde{\mu}}\,\tilde{n}^2\cos i + \tilde{n}\cos r\cos^2\theta}\,. \tag{32}$$

Upon setting $\cos\theta = 1$, the relations (30) and (32) reproduce results for reflection and refraction off standard dielectric media (see Ref. 69). Decreasing $\cos\theta$ decreases the A_μ component of $A_\mu^{\tilde{Q}}$, and thus favors reflection over refraction. For θ as small as in nature, the changes introduced by $\theta \neq 0$ are small. In the (unphysical) limit in which $\cos\theta = 0$, $A_\mu^{\tilde{Q}}$ would be orthogonal to A_μ making the CFL phase a superconductor with respect to ordinary electromagnetism. In this limit, we expect and find zero refraction and perfect reflection for both polarizations.

Are the solutions (30) and (32) consistent with energy conservation? The Poynting vector $\mathbf{S} = \frac{1}{2}\,(\mathbf{E}\times\mathbf{H})$ measures the energy flow per unit area and time. Continuity of the z-component of the Poynting vector requires

$$\mathcal{E}^2\sqrt{\frac{\epsilon_0}{\mu_0}}\cos i = (\mathcal{E}')^2\sqrt{\frac{\epsilon_0}{\mu_0}}\cos i + (\tilde{\mathcal{E}}_r)^2\sqrt{\frac{\tilde{\epsilon}}{\tilde{\mu}}}\cos r\,, \tag{33}$$

a relation which is indeed satisfied by both (30) and (32).

Notice that in our analysis of each of the two polarizations, one boundary condition was irrelevant and Snell's law could be used to eliminate a second. If we had used energy conservation in the form (33) in our derivation instead of just as a check, we could have derived all our results from the single boundary condition (21). That is, given only the boundary condition (21) which is easily derived, Snell's law (28) which is kinematic, and energy conservation (33), we can derive our solutions describing the reflection and refraction of light of both polarizations and, from these electromagnetic fields, we can then derive the remaining boundary conditions (22), (23) and (24). This means that we have *derived* the boundary conditions motivated above by the idea that the QCD vacuum behaves like a dual superconductor filled with a condensate of color-magnetic monopoles.[70] Having analyzed the illumination of dense quark matter, we find that in addition we have illuminated our understanding of the QCD vacuum.

It is perhaps of more practical value to analyze the response of a chunk of CFL matter to a static magnetic field, as in the core of a neutron star. The effect of a chunk of color superconducting quark matter (whether in the CFL phase or in the less symmetric phase in which only up and down quarks pair) on a static magnetic field has been described in Ref. 64, and is essentially the static limit of the refraction calculation just presented. Some fraction of an externally applied ordinary magnetic field penetrates the superconductor in the form of a $\tilde{Q}$-magnetic field, while some fraction of the ordinary magnetic field is expelled by the Meissner effect. The fraction of the field which is expelled depends both on θ and on the shape of the chunk color superconducting quark matter, but it is small when θ is small, as in nature. Most of the flux is admitted, as $\tilde{Q}$-flux. This $\tilde{Q}$-magnetic field satisfies Maxwell's equations and is not restricted to flux tubes.

4.6. *Quark-Hadron Continuity*

The universal features of the color-flavor locked state: confinement, chiral symmetry breaking leaving a vector $SU(3)$ unbroken, and baryon number superfluidity, are exactly what one expects to find in nuclear matter in three-flavor QCD.[13] Perhaps this is not immediately obvious in the case of baryon number superfluidity, but let us recall that pairing phenomena, which would go over into neutron superfluidity and proton superconductivity in nuclear matter, are very well established in ordinary nuclei. In three-flavor QCD, there are good reasons[71] to think that the pairing interaction in the flavor singlet dibaryon channel (the so-called H-dibaryon channel) would be quite attractive in three-flavor QCD, and support a robust baryon number superfluidity. Thus, the symmetries of the color-flavor locked phase are precisely those of nuclear matter in three-flavor QCD, perhaps better referred to as hypernuclear matter.[13]

Furthermore, there is an uncanny resemblance between the low-lying spectrum computed from first principles for QCD at asymptotically high density, and what one expects to find in hypernuclear matter, in a world with three degenerate quark flavors. It is hard to resist the inference that in this theory, there need be no phase transistion between nuclear density and high density.[13] There need be no sharp boundary between hypernuclear matter, where microscopic caculations are difficult but the convenient degrees of freedom are "obviously" hadrons, and the asymptotic high-density phase, where weak-coupling (but non-perturbative) calculations are possible, and the right degrees of freedom are elementary quarks and gluons, together with the collective Nambu-Goldstone modes. We call this quark-hadron continuity.[13] Perhaps the least surprising aspect of this, by now, is the continuity between the pseudoscalar mesons at nuclear density and those at asymptotically high densities, since in both regimes these are present as a result of the breaking of the same symmetry. It might seem more shocking that a quark can go over continuously into, or "be", a baryon, since baryons are supposed to contain three quarks, but remember that in the color-flavor locked phase the quarks are immersed in a diquark condensate, and so a one-quark excitation can pick two quarks up from (or lose to quarks to) the condensate at will. The difference between one and three is negotiable. What about the gluons? Within the color-flavor locked phase, similarly, they are quite literally the physical vector mesons. They are massive, as we have discussed, and have the right quantum numbers. Thus the original vision of Yang and Mills – who proposed non-abelian gauge theory as a model of ρ mesons – is here embodied.

Note that the hypothesis of continuity between hypernuclear and dense quark matter certainly does not preclude quantitative change. Far from it. The dispersion relation for a fermion — whether a quark in the CFL phase or a baryon in the hypernuclear phase — is characterized by a gap at the Fermi surface and by a gap at zero momentum, i.e. a mass. As a function of increasing density, gaps at the hyperon Fermi surfaces due to hyperon-hyperon pairing evolve continuously to become the gaps at the quark Fermi surfaces which characterize the color-flavor locked phase.[17] During this evolution, the gaps are thought to increase significantly. In contrast, the masses ("gaps at zero momentum") decrease dramatically with increasing density as they evolve from being of order the hyperon masses in hypernuclear matter to being of order the current quark masses at asymptotically high densities.

Note that in order for quark-hadron continuity to be realized, $U(1)_{EM}$ must not be broken by hyperon-hyperon pairing.[17] At every point during the evolution of the theory as a function of increasing density, there will be an unbroken $U(1)$ and a massless $\tilde{Q}$-photon and the excitations will be integer charged. As the density is increased, however, the definition of the $\tilde{Q}$-photon in terms of the vacuum photon and gluon fields rotates. The ordinary photon rotates to become the $\tilde{Q}$-photon of the CFL phase. Turning to the massive vector bosons, in hypernuclear matter there is both an octet and a singlet. The singlet must become much heavier than the octet as a function of increasing density, since in the low energy description of the color-flavor locked phase one finds the octet alone. We see that if quark-hadron continuity is realized in QCD with three degenerate quarks, it requires various quantitative (but continuous) changes. What is remarkable is that it is even possible to imagine watching all the physical excitations of the theory evolving continuously as one dials the density up and goes from a strongly coupled hadronic world to a weak-coupling world of quarks and gluons.

If the quarks are massless, the Nambu-Goldstone bosons are massless in both hypernuclear and CFL quark matter, and in between. Once nondegenerate quark masses are introduced, however, the evolution of the Nambu-Goldstone masses as a function of increasing density becomes more intricate, as the kaon must go from being heavier than the pion to being lighter.

Note, finally, that the whole story becomes further complicated once the strange quark is made as heavy as in nature.[17,39] Although the color-flavor locked phase is certainly obtained at asymptotically densities, where quark masses are neglectable, the nuclear matter phase, made of neutrons and protons only, is not continuously connectable with the color-flavor locked phase. If quark-hadron continuity is to be realized in the phase diagram of nature, what must happen is that, as a function of increasing density, one first goes from nuclear matter to hypernuclear matter, with sufficiently high density that all the hyperons have similar Fermi surfaces. This first stage must involve phase transitions, as the symmetries of hypernuclear matter differ from those of ordinary nuclear matter. Then, as the density is increased further, the hypernuclear matter may evolve continuosly to become CFL quark matter, with pairing among hyperons becoming CFL pairing among quarks.

We now have a description of the properties of the CFL phase and its excitations, in which much can be described quantitatively if the value of the gap Δ is known. In the next two sections, we describe the less symmetric forms of color superconductivity which arise in QCD with $N_f \neq 3$. Already, however, in our idealized world (in which we either have three degenerate quarks or such high densities that the quark mass differences can be neglected) let us pause to marvel at our theoretical good fortune. The color-flavor locked phase is a concrete realization of the idea of complementarity: the same phase of a gauge theory can be described simultaneously as one in which the gauge symmetry is spontaneously broken and as one in which color is confined.[11] This means that it provides us with a weak-coupling laboratory within which we can study a confined phase from first principles at weak coupling. It is furthermore a phase of QCD wherein the physics of chiral symmetry breaking — indeed all the parameters of the chiral effective Lagrangian and all known or conjectured phenomena of the pseudoscalar meson sector, including kaon condensation — are amenable to controlled, weak-coupling calculation.

5. 2SC and Other Variants

5.1. *Two Flavors*

In the previous section, we have described quark matter in QCD with three degenerate flavors of light quarks. In nature the strange quark is heavier than the other two, so we now go to the opposite extreme of an infinitely heavy strange quark and describe the color superconducting phase in QCD with two flavors of light quarks. The $\bar{\mathbf{3}}$ is still the most attractive color channel for quark pairing,[5,7,8,9] and the resulting condensate

$$\langle \psi_i^\alpha \psi_j^\beta \rangle \propto \epsilon_{\alpha\beta 3} \epsilon^{ij} \tag{34}$$

picks a color direction. In this case, quarks of the first two colors (red and green) participate in pairing, while the third color (blue) does not.

The ground state is invariant under an $SU(2)$ subgroup of the color rotations that mixes red and green, but the blue quarks are singled out as different. The pattern of symmetry breaking is therefore (with gauge symmetries in square brackets)

$$\begin{aligned} &[SU(3)_{\text{color}}] \times [U(1)_Q] \times SU(2)_L \times SU(2)_R \\ \longrightarrow\ &[SU(2)_{\text{color}}] \times [U(1)_{\tilde{Q}}] \times SU(2)_L \times SU(2)_R \end{aligned} \tag{35}$$

The features of this pattern of condensation are

(1) The color gauge group is broken down to $SU(2)$, so five of the gluons will become massive, with masses of order the gap (since the

coupling is of order 1). Their masses have been computed in the weak-coupling theory valid at asymptotically high densities[59] and in an instanton liquid model.[60] The remaining three gluons are associated with an unbroken $SU(2)$ red-green gauge symmetry, whose confinement distance rises exponentially with density.[72] This aspect of the infrared physics of the 2SC phase is not under perturbative control, so unlike the CFL phase we cannot claim that in the 2SC phase any physical quantity can be obtained from a controlled weak-coupling calculation at sufficiently high density.

(2) The red and green quark modes acquire a gap Δ, which is the mass of the physical excitations around the Fermi surface (quasiquarks). There is no gap for the blue quarks in this ansatz, and it is an interesting question whether they find some other channel in which to pair. The available attractive channels arising from the instanton interaction break rotational invariance, and are weak so the gap will be much smaller, perhaps in the keV range.[8]

(3) As in the CFL phase a linear combination of the photon and one gluon gives a massless gauge boson that couples to a new unbroken rotated electromagnetism $\tilde{Q}$. The two blue quarks have $\tilde{Q}$-charges 0 and 1.

(4) No global symmetries are broken so there are no light scalars—the 2SC phase is not a superfluid. It has the same symmetries as the quark-gluon plasma (QGP), so there need not be any phase transition between them, but there will be a chiral phase transition between the hadronic and 2SC phases. This phase transition is first order[8,16,19,73] since it involves a competition between chiral condensation and diquark condensation.[16,19] Although the quark pair condensate appears to break baryon number, it does not. In the two flavor case baryon number is a linear combination of electric charge and isospin, $B = 2Q - 2I_3$, so baryon number is already included in the symmetry groups of (35). Just as an admixture of gluon and photon survives unbroken as a rotated electromagnetism, so an admixture of B and T_8 survives unbroken as a rotated baryon number.

(5) Axial color is not a symmetry of the QCD action, but at asymptotically high densities where the QCD coupling g is weak, explicit axial color breaking is also weak. As a result, the pseudoscalar excitations of the condensate which would be Goldstone bosons arising from axial-$SU(3)_{\text{color}}$ to axial-$SU(2)_{\text{color}}$ breaking if g were zero may be rather light.[74]

It is interesting that both the 2SC and CFL phases satisfy anomaly matching constraints, even though it is not yet completely clear whether this must be the case when Lorentz invariance is broken by a nonzero density.[75] It is not yet clear how high density QCD with larger numbers of flavors,[31] which we discuss below, satisfies anomaly matching constraints. Also, anomaly matching in the 2SC phase requires that the up and down quarks of the third color remain ungapped; this requirement must, therefore, be modified once these quarks pair to form a $J = 1$ condensate, breaking rotational invariance.[8]

5.2. *Two+One Flavors*

Nature chooses two light quarks and one middle-weight strange quark. If we imagine beginning with the CFL phase and increasing m_s, how do we get to the 2SC phase? This question has been answered in Ref. 17, 39. A nonzero m_s weakens those condensates which involve pairing between light and strange quarks. The CFL phase requires nonzero $\langle us\rangle$ and $\langle ds\rangle$ condensates; because these condensates pair quarks with differing Fermi momenta they can only exist if the resulting gaps (call them Δ_{us} and Δ_{ds}) are larger than of order $m_s^2/2\mu$, the difference between the u and s Fermi momenta in the absence of pairing. This means that as a function of increasing m_s at fixed μ (or decreasing μ at fixed m_s) there must be a first order unlocking phase transition.[17,39] The argument can be phrased thus: the 2SC and CFL phases must be separated by a phase transition, because chiral symmetry is broken in the CFL phase but not in the 2SC phase; suppose this transition were second order; this would require Δ_{us} and Δ_{ds} to be infinitesimally small but nonzero just on the CFL side of the transition; however, these gaps must be greater than of order $m_s^2/2\mu$; a second order phase transition is therefore a logical impossibility, either in mean field theory or beyond; the transition must therefore be first order. Note that the m_s that appears in these estimates is a density dependent effective strange quark mass, somewhat greater than the current quark mass.[76]

Putting in reasonable numbers for quark matter which may arise in compact stars, for $m_s = 200 - 300$ MeV and $\mu = 400 - 500$ MeV we find that the CFL phase is obtained if the interactions are strong enough to generate a gap Δ which is larger than about $40 - 110$ MeV, while the 2SC phase is obtained if Δ is smaller. $\Delta \sim 40 - 110$ MeV is within the range of current estimates and present calculational methods are therefore not precise enough to determine whether quark matter with these parameters is in the CFL or 2SC phases. At asymptotically high densities, however, the CFL phase is necessarily favored.

Note that the 2SC phase in QCD with massive strange quarks *is* a superfluid: no linear combination of baryon number and a gauge symmetry remains unbroken. Note also that in this phase, five quarks (blue up; blue down; strange quarks of all colors) are left unpaired. As we shall discuss in the next section, they may in fact pair to form a crystalline color superconductor.

5.3. *Four or More Flavors*

We end this section with brief mention of four variants which are unphysical, but nevertheless instructive: QCD with more than three light flavors, QCD with two colors, QCD with many colors, and QCD with large isospin density and zero baryon density.

Dense quark matter in QCD with more than three flavors was studied in Ref. 31. The main result is that the color-flavor locking phenomenon persists: condensates form which lock color rotations to flavor rotations, and the $SU(N_f)_L \times SU(N_f)_R$ group is broken down to a vector subgroup. Unlike with $N_f = 3$, however, the unbroken group is not the full $SU(N_f)_{L+R}$ which is unbroken in the vacuum. In the case of $N_f = 4$, for example, one finds $SU(4)_L \times SU(4)_R \rightarrow O(4)_{L+R}$ while in the case of $N_f = 5$, $SU(5)_L \times SU(5)_R \rightarrow SU(2)_{L+R}$.[31] For $N_f = 4, 5$ as for $N_f = 3$, chiral symmetry is broken in dense quark matter. However, because the unbroken vector groups are smaller than $SU(N_f)_V$, there must be a phase transition between hadronic matter and dense quark matter in these theories.[31]

If N_f is a multiple of three, the order parameter takes the form of multiple copies of the $N_f = 3$ order parameter, each locking a block of three flavors to color.[31] All quarks are gapped in this phase, as in the $N_f = 3$ CFL phase. For $N_f = 6$, the resulting symmetry breaking pattern is $SU(6)_L \times SU(6)_R \rightarrow SU(3)_{L+R} \times U(1)_{L+R} \times U(1)_{L-R}$.[31] The unbroken $SU(3)_{L+R}$ is a simultaneous rotation of both three flavor blocks for L and R and a global color rotation. Note that the unbroken $U(1)$'s are subgroups of the original $SU(6)$ groups: they correspond to vector and axial flavor rotations which rotate one three flavor block relative to the other. Note that for $N_f = 6$, unlike for $N_f = 3, 4, 5$, chiral symmetry is not completely broken at high density: an axial $U(1)$ subgroup remains unbroken. As the primary condensate we have just described leaves no quarks ungapped, there is no reason to expect the formation of any subdominant condensate which could break the unbroken chiral symmetry. Both because of this unbroken chiral symmetry and because the unbroken vector symmetry differs from that of the vacuum, there must be a phase transition between hadronic matter and dense quark matter in QCD with $N_f = 6$.[31]

5.4. *Two Colors*

The simplest case of all to analyze is QCD with two colors and two flavors. The condensate is antisymmetric in color and flavor, and is therefore a singlet in both color and flavor. Because it is a singlet in color, dense quark matter in this theory is not a color superconductor. Although the condensate is a singlet under the ordinary $SU(2)_L \times SU(2)_R$ flavor group, it nevertheless does break symmetries because the symmetry of the vacuum in QCD with $N_f = N_c = 2$ is enhanced to $SU(4)$. One reason why $N_c = 2$ QCD is interesting to study at nonzero density is that it provides an example where quark pairing can be studied on the lattice.[77] The $N_c = 2$ case has also been studied analytically in Ref. 9, 78; pairing in this theory is simpler to analyze because quark Cooper pairs are color singlets. We refer the reader to these references for details.

5.5. *Many Colors*

The $N_c \to \infty$ limit of QCD is often one in which hard problems become tractable. However, the ground state of $N_c = \infty$ QCD is a chiral density wave, not a color superconductor.[79] At asymptotically high densities, color superconductivity persists up to N_c's of order thousands[80,81] before being supplanted by the phase described in Ref. 79. At any finite N_c, color superconductivity occurs at arbitrarily weak coupling whereas the chiral density wave does not. For $N_c = 3$, color superconductivity is still favored over the chiral density wave (although not by much) even if the interaction is so strong that the color superconductivity gap is $\sim \mu/2$.[82]

5.6. *QCD at Large Isospin Density*

The phase of $N_c = 3$ QCD with nonzero isospin density ($\mu_I \neq 0$) and zero baryon density ($\mu = 0$) *can* be simulated on the lattice.[83] The sign problems that plague simulations at $\mu \neq 0$ do not arise for $\mu_I \neq 0$. Although not physically realizable, physics with $\mu_I \neq 0$ and $\mu = 0$ is very interesting to consider because phenomena arise which are similar to those occurring at large μ and, in this context, these phenomena are accessible to numerical "experiments". Such lattice simulations can be used to test calculational methods which have also been applied at large μ, where lattice simulation is unavailable. At low isospin density, this theory describes a dilute gas of Bose-condensed pions. Large μ_I physics features large Fermi surfaces for down quarks and anti-up quarks, Cooper pairing of down and anti-up quarks, and a gap whose g-dependence is as in (5), albeit with a different coefficient of $1/g$ in the exponent.[83] This condensate has the same quantum numbers as the pion condensate expected at much lower μ_I, which means that a hypothesis of continuity between hadronic — in this case pionic — and quark matter as a function of μ_I. Both the dilute pion gas limit and the asymptotically large μ_I limit can be treated analytically; the possibility of continuity between these two limits can be tested on the lattice.[83] The transition from a weak coupling superconductor with condensed Cooper pairs to a gas of tightly bound bosons which form a Bose condensate can be studied in a completely controlled fashion.

6. Crystalline Color Superconductivity

At asymptotic densities, the ground state of QCD with quarks of three flavors (u, d and s) with equal masses is expected to be the color-flavor locked (CFL) phase. This phase features a condensate of Cooper pairs of quarks which includes ud, us, and ds pairs. Quarks of all colors and all flavors participate in the pairing, and all excitations with quark quantum numbers are gapped. As in any BCS state, the Cooper pairing in the CFL state pairs quarks whose momenta are equal in magnitude and opposite in direction, and pairing is strongest between pairs of quarks whose momenta are both near their respective Fermi surfaces. Pairing persists even in the face of a stress (such as a chemical potential difference or a mass difference) that seeks to push the Fermi surfaces apart, although a stress that is too strong will ultimately disrupt Cooper pairing.[17,39] Thus, the CFL phase persists for unequal quark masses, so long as the differences are not too large.[17,39] This means that the CFL phase is the ground state for real QCD, assumed to be in equilibrium with respect to the weak interactions, as long as the density is high enough.

Imagine decreasing the quark number chemical potential μ from asymptotically large values. The quark matter at first remains color-flavor locked, although the CFL condensate may rotate in flavor space as terms of order m_s^4 in the free energy become important.[56] Color-flavor locking is maintained until a transition to a state in which some quarks become ungapped. This "unlocking transition", which must be first order,[17,39] occurs when[17,39,63,65]

$$\mu \approx m_s^2/4\Delta_0 \ . \tag{36}$$

In this expression and throughout this section, we write the gap in the BCS state as Δ_0. As we have seen, estimates in both models and asymptotic analyses suggest that it is of order tens to 100 MeV. m_s is the strange quark mass parameter, which includes the contribution from any $\langle \bar{s}s \rangle$ condensate induced by the nonzero current strange quark mass, making it a density-dependent effective mass, decreasing as density increases and equaling the current strange quark mass only at asymptotically high densities. At densities which may arise at the center of compact stars, corresponding to $\mu \sim 400 - 500$ MeV, m_s is certainly significantly larger than the current quark mass, and its value is not well-known. In fact, m_s decreases discontinuously at the unlocking transition.[76] Thus, the criterion (36) can only be used as a rough guide to the location of the unlocking transition in nature.[76] Given this quantitative uncertainty, there remain two logical possibilities for what happens as a function of decreasing μ. One possibility is a first order phase transition directly from color-flavor locked quark matter to hadronic matter, as explored in Ref. 65. The second possibility is an unlocking transition[17,39] to quark matter in which not all quarks participate in the dominant pairing, followed only at a lower μ by a transition to hadronic matter. We assume the second possibility here, and explore its consequences.

Once CFL is disrupted, leaving some species of quarks with differing Fermi momenta and therefore unable to participate in BCS pairing, it is natural to ask whether there is some generalization of the ansatz in which pairing between two species of quarks persists even once their Fermi momenta differ. Crystalline color superconductivity is the answer to this question. The idea is that it may be favorable for quarks with differing Fermi momenta to form pairs whose momenta are *not* equal in magnitude and opposite in sign.[84,85] This generalization of the pairing ansatz (beyond BCS ansätze in which only quarks with momenta which add to zero pair) is favored because it gives rise to a region of phase space where *both* of the quarks in a pair are close to their respective Fermi surfaces, and such pairs can be created at low cost in free energy. Condensates of this sort spontaneously break translational and rotational invariance, leading to gaps which vary in a crystalline pattern. As a function of increasing depth in a compact star, μ increases, m_s decreases, and Δ_0 changes also. If in some shell within the quark matter core of a neutron star (or within a strange quark star) the quark number densities are such that crystalline color superconductivity arises, rotational vortices may be pinned in this shell, making it a locus for glitch phenomena.[85]

An analysis of these ideas in the context of the disruption of CFL pairing is complicated by the fact that in quark matter in which CFL pairing does not occur, up and down quarks may nevertheless continue to pair in the usual BCS fashion. In this 2SC phase, the attractive channel involves the formation of Cooper pairs which are antisymmetric in both color and flavor, yielding a condensate with color (Greek indices) and flavor (Latin indices) structure $\langle q_a^\alpha q_b^\beta \rangle \sim \epsilon_{ab}\epsilon^{\alpha\beta 3}$. This condensate leaves five quarks unpaired: up and down quarks of the third color, and strange quarks of all three colors. Because the BCS pairing scheme leaves ungapped quarks with differing Fermi momenta, crystalline color superconductivity may result.

Most analyses of crystalline color superconductivity have been done in the simplified model context with pairing between two quark species whose Fermi momenta are pushed apart by turning on a chemical potential difference,[85,86,87,88] rather than considering CFL pairing in the presence of quark mass differences. In Ref. 89, we investigate the ways in which the response of the system to mass differences is similar to or different from the response to chemical potential differences. We can address this question within the two-flavor model by generalizing it to describe pairing between massless up quarks and strange quarks with mass m_s. For completeness, we introduce

$$\begin{aligned} \mu_u &= \mu - \delta\mu \\ \mu_s &= \mu + \delta\mu \ , \end{aligned} \tag{37}$$

allowing us to consider the effects of m_s and $\delta\mu$ simultaneously. We shall use this two-flavor toy model throughout, deferring an analysis of crystalline color superconductivity induced by the effects of m_s on three-flavor quark matter to future work.

6.1. *Consequences of $\delta\mu \neq 0$, with m_s=0*

Before describing the consequences of $m_s \neq 0$, let us review the salient facts known about the consequences of $\delta\mu \neq 0$, upon taking $m_s = 0$. If $|\delta\mu|$ is nonzero but less than some $\delta\mu_1$, the ground state in the two-flavor toy-model is precisely that obtained for $\delta\mu = 0$.[90,91,85,b] In this 2SC state, red and green up and strange quarks pair, yielding four quasiparticles with superconducting gap Δ_0. Furthermore, the number density of red and green up quarks is the same as that of red and green strange quarks. As long as $|\delta\mu|$ is not too large, this BCS state remains unchanged (and favored) because maintaining equal number densities, and thus coincident Fermi surfaces, maximizes the pairing and hence the interaction energy. As $|\delta\mu|$ is increased, the BCS state remains the ground state of the system only as long as its negative interaction energy offsets the large positive free energy cost associated with forcing the Fermi seas to remain coincident. In the weak coupling limit, in which $\Delta_0/\mu \ll 1$, the BCS state persists for $|\delta\mu| < \delta\mu_1 = \Delta_0/\sqrt{2}$.[90,85] For larger Δ_0, the $1/\sqrt{2}$ coefficient changes in value. These conclusions are the same whether the interaction between quarks is modeled as a point-like four-fermion interaction or is approximated by single-gluon exchange. The loss of BCS pairing at $|\delta\mu| = \delta\mu_1$ is the analogue in this toy model of the unlocking transition.

If $|\delta\mu| > \delta\mu_1$, BCS pairing between u and s is not possible. However, in a range $\delta\mu_1 < |\delta\mu| < \delta\mu_2$ near the unpairing transition, it is favorable to form a crystalline color superconducting state in which the Cooper pairs have nonzero momentum. This phenomenon was first analyzed by Larkin and Ovchinnikov and Fulde and Ferrell[84] (LOFF) in the context of pairing between electrons in which spin-up and spin-down Fermi momenta differ. It has proven difficult to find a condensed matter physics system which is well described simply as BCS pairing in the presence of a Zeeman effect: any magnetic perturbation that may induce a Zeeman effect tends to have much larger effects on the motion of the electrons, as in the Meissner effect. The QCD context of interest to us, in which the Fermi momenta being split are those of different flavors rather than of different spins, therefore turns out to be the natural arena for the phenomenon first analyzed by LOFF.

The crystalline color superconducting phase (also called the LOFF phase) has been described in Ref. 85 (following Ref. 84) upon making the simplifying assumption that quarks interact via a four-fermion interaction with the quantum numbers of single gluon exchange. In the LOFF state, each Cooper pair carries momentum $2\mathbf{q}$ with $|\mathbf{q}| \approx 1.2\delta\mu$. The condensate and gap parameter vary in space with wavelength $\pi/|\mathbf{q}|$. Although the

[b]In this two-flavor toy-model the diquark condensate is a flavor singlet. As the condensate breaks no flavor symmetries, there is no analogue of the rotations of the condensate in flavor space which occur within the CFL phase with nonzero $\delta\mu$.[56]

magnitude $|\mathbf{q}|$ is determined energetically, the direction $\hat{\mathbf{q}}$ is chosen spontaneously. The LOFF state is characterized by a gap parameter Δ and a diquark condensate, but not by an energy gap in the dispersion relation: the quasiparticle dispersion relations vary with the direction of the momentum, yielding gaps that vary from zero up to a maximum of Δ. The condensate is dominated by those regions in momentum space in which a quark pair with total momentum $2\mathbf{q}$ has both members of the pair within $\sim \Delta$ of their respective Fermi surfaces. These regions form circular bands on the two Fermi surfaces. Making the ansatz that all Cooper pairs make the same choice of direction $\hat{\mathbf{q}}$ corresponds to choosing a single circular band on each Fermi surface. In position space, it corresponds to a condensate which varies in space like

$$\langle\psi(\mathbf{x})\psi(\mathbf{x})\rangle \propto \Delta e^{2i\mathbf{q}\cdot\mathbf{x}} \ . \tag{38}$$

This ansatz is certainly *not* the best choice. If a single plane wave is favored, why not two? That is, if one choice of $\hat{\mathbf{q}}$ is favored, why not add a second $\mathbf{q}$, with the same $|\mathbf{q}|$ but a different $\hat{\mathbf{q}}$? If two are favored, why not three? This question, namely, the determination of the favored crystal structure of the crystalline color superconductor phase, is unresolved but is under investigation. Note, however, that if we find a region $\delta\mu_1 < |\delta\mu| < \delta\mu_2$ in which the simple LOFF ansatz with a single $\hat{\mathbf{q}}$ is favored over the BCS state and over no pairing, then the LOFF state with whatever crystal structure turns out to be optimal must be favored in *at least* this region. Note also that even the single $\hat{\mathbf{q}}$ ansatz, which we use henceforth, breaks translational and rotational invariance spontaneously. The resulting phonon has been analyzed in Ref. 87.

Crystalline color superconductivity is favored within a window $\delta\mu_1 < |\delta\mu| < \delta\mu_2$. As $|\delta\mu|$ increases from 0, one finds a first order phase transition from the ordinary BCS phase to the crystalline color superconducting phase at $|\delta\mu| = \delta\mu_1$ and then a second order phase transition at $|\delta\mu| = \delta\mu_2$ at which Δ decreases to zero. Because the condensation energy in the LOFF phase is much smaller than that of the BCS condensate at $\delta\mu = 0$, the value of $\delta\mu_1$ is almost identical to that at which the naive unpairing transition from the BCS state to the state with no pairing would occur if one ignored the possibility of a LOFF phase, namely $\delta\mu_1 = \Delta_0/\sqrt{2}$. For all practical purposes, therefore, the LOFF gap equation is not required in order to determine $\delta\mu_1$. The LOFF gap equation is used to determine $\delta\mu_2$ and the properties of the crystalline color superconducting phase.[85] In the limit of a weak four-fermion interaction, the crystalline color superconductivity window is bounded by $\delta\mu_1 = \Delta_0/\sqrt{2}$ and $\delta\mu_2 = 0.754\Delta_0$, as first demonstrated in Ref. 84. These results have been extended beyond the weak four-fermion interaction limit in Ref. 85.

We now know that the use of the simplified point-like interaction significantly underestimates the width of the LOFF window: assuming instead that quarks interact by exchanging medium-modified gluons yields a much larger value of $\delta\mu_2$.[88] This can be understood upon noting that quark-quark interaction by gluon exchange is dominated by forward scattering. In most scatterings, the angular positions on their respective Fermi surfaces do not change much. In the LOFF state, small-angle scattering is advantageous because it cannot scatter a pair of quarks out of the region of momentum space in which both members of the pair are in their respective circular bands, where pairing is favored. This means that it is natural that a forward-scattering dominated interaction like single-gluon exchange is much more favorable for crystalline color superconductivity that a point-like interaction, which yields s-wave scattering. Thus, although for the present we use the point-like interaction in our analysis of m_s-induced crystalline color superconductivity, it is worth remembering that this is very conservative.

6.2. *Consequences of $m_s \neq 0$*

In the absence of any interaction, and thus in the absence of pairing, the effect of a strange quark mass is to shift the Fermi momenta to

$$\begin{aligned} p_F^u &= \mu - \delta\mu \\ p_F^s &= \sqrt{(\mu + \delta\mu)^2 - m_s^2}\ . \end{aligned} \tag{39}$$

Assuming both $|\delta\mu|/\mu$ and m_s/μ are small, the separation between the two Fermi momenta is $\approx |2\delta\mu - m_s^2/2\mu|$. This suggests the conjecture that even when $m_s \neq 0$ the description given in the previous subsection continues to be valid upon replacing $|\delta\mu|$ by $|\delta\mu - m_s^2/4\mu|$. We show in Ref. 89 that this conjecture is *incorrect* in one key respect: whereas if $m_s = 0$ a $|\delta\mu|$ which is nonzero but smaller than $\delta\mu_1$ has no effect on the BCS state, the BCS gap Δ_0 decreases with increasing m_s^2. We show that for small m_s^2, $\Delta_0(m_s)/\Delta_0(0)$ decreases linearly with m_s^2. Because Δ_0 occurs in the free energy in a term of order $\Delta_0^2\mu^2$, the m_s-dependence of Δ_0 corrects the free energy by of order $\Delta_0(0)^2 m_s^2$. As $\delta\mu$ has no analogous effect, we conclude that $m_s^2/4\mu$ and $\delta\mu$ have qualitatively different effects on the paired state.

At another level, however, the story *is* quite similar to that for $m_s = 0$: if $|\delta\mu - m_s^2/4\mu|$ is small enough, we find the BCS state; if $|\delta\mu - m_s^2/4\mu|$ lies within an intermediate window, we find LOFF pairing; if $|\delta\mu - m_s^2/4\mu|$ is large enough, no pairing is possible. The boundaries between the phases, however, are related to a $\Delta_0(m_s)$, rather than simply to a constant Δ_0. That is, the definitions of "small enough" and "large enough" are m_s-dependent. We map the $(m_s, \delta\mu)$ plane in Ref. 89. We find that the appropriate variable to use to describe the width of the crystalline color superconductivity window is $\delta\mu/\Delta_0(m_s)$, as opposed to $\delta\mu/\Delta_0(0)$. When described with this variable, m_s-induced and $\delta\mu$-induced crystalline color superconductivity are approximately equally robust. At all but the weakest of couplings, the width of the crystalline color superconductivity window increases with m_s, meaning that crystalline color superconductivity is somewhat more robust if it is m_s-induced than if it is $\delta\mu$-induced. Indeed, we find that at the moderate coupling corresponding to $\Delta_0(0) = 100$ MeV, m_s-induced crystalline color superconductivity occurs whereas $\delta\mu$-induced crystalline color superconductivity does not.

6.3. *Opening the Crystalline Color Superconductivity Window*

In Ref. 88, we analyze the crystalline color superconducting phase upon assuming that quarks interact by the exchange of a medium-modified gluon, as is quantitatively valid at asymptotically high densities. We obtain $\delta\mu_2$, the upper boundary of the crystalline color superconductivity window. This analysis is controlled at asymptotically high densities where the coupling g is weak.

At weak coupling, quark-quark scattering by single-gluon exchange is dominated by forward scattering. In most scatterings, the angular positions of the quarks on their respective Fermi surfaces do not change much. As a consequence, the weaker the coupling the more the physics can be thought of as a sum of many $(1+1)$-dimensional theories, with only rare large-angle scatterings able to connect one direction in momentum space with others.[26] In the LOFF state, small-angle scattering is advantageous because it cannot scatter a pair of quarks out of the region of momentum space in which both members of the pair are in their respective rings, where pairing is favored. This means that it is natural to expect that a forward-scattering-dominated interaction like single-gluon exchange is much more favorable for crystalline color superconductivity than a point-like interaction, which yields s-wave scattering.

Suppose for a moment that we were analyzing a truly $(1+1)$-dimensional theory. The momentum-space geometry of the LOFF state in one spatial dimension is qualitatively different from that in three. Instead of Fermi surfaces, we would have only "Fermi points" at $\pm\mu_u$ and $\pm\mu_d$. The only choice of $|\mathbf{q}|$ which allows pairing between u and d quarks at their respective Fermi points is $|\mathbf{q}| = \delta\mu$. In $(3+1)$ dimensions, in contrast, $|\mathbf{q}| > \delta\mu$ is favored because it allows LOFF pairing in ring-shaped regions of the Fermi surface, rather than just at antipodal points.[84,85] Also, in striking contrast to the $(3+1)$-dimensional case, it has long been known that in a true $(1+1)$-dimensional theory with a point-like interaction between fermions, $\delta\mu_2/\Delta_0 \to \infty$ in the weak-interaction limit.[92]

We expect that in $(3+1)$-dimensional QCD with the interaction given by single-gluon exchange, as $\mu \to \infty$ and $g(\mu) \to 0$ the $(1+1)$-dimensional results should be approached: the energetically favored value of $|\mathbf{q}|$ should become closer and closer to $\delta\mu$, and $\delta\mu_2/\Delta_0$ should diverge. We derive both these effects in Ref. 88 and furthermore show that both are clearly in evidence already at the rather large coupling $g = 3.43$, corresponding to $\mu = 400$ MeV using the conventions of Refs. 28, 34. At this coupling, $\delta\mu_2/\Delta_0 \approx 1.2$, meaning that $(\delta\mu_2 - \delta\mu_1) \approx (1.2 - 1/\sqrt{2})\Delta_0$, which is much larger than $(0.754 - 1/\sqrt{2})\Delta_0$. If we go to much higher densities, where the calculation is under quantitative control, we find an even more striking enhancement: when $g = 0.79$ we find $\delta\mu_2/\Delta_0 > 1000$! We see that (relative to expectations based on experience with point-like interactions) the crystalline color superconductivity window is wider by more than four orders of magnitude at this weak coupling, and is about one order of magnitude wider at accessible densities if weak-coupling results are applied there.[c]

We have found that $\delta\mu_2/\Delta_0$ diverges in QCD as the weak-coupling, high-density limit is taken. Applying results valid at asymptotically high densities to those of interest in compact stars, namely $\mu \sim 400$ MeV, we find that even here the crystalline color superconductivity window is an order of magnitude wider than that obtained previously upon approximating the interaction between quarks as point-like. The crystalline color super-

[c]LOFF condensates have also recently been considered in two other contexts. In QCD with $\mu_u < 0$, $\mu_d > 0$ and $\mu_u = -\mu_d$, one has equal Fermi momenta for $\bar{u}$ antiquarks and d quarks, BCS pairing occurs, and consequently a $\langle\bar{u}d\rangle$ condensate forms.[83,93] If $-\mu_u$ and μ_d differ, and if the difference lies in the appropriate range, a LOFF phase with a spatially varying $\langle\bar{u}d\rangle$ condensate results.[83,93] The result of Ref. 88 that the LOFF window is much wider than previously thought applies in this context also. Suitably isospin asymmetric nuclear matter may also admit LOFF pairing, as discussed recently in Ref. 94. Here, the interaction is not forward-scattering dominated.

conductivity window in parameter space may therefore be much wider than previously thought, making this phase a *generic* feature of the phase diagram for cold dense quark matter. The reason for this qualitative increase in $\delta\mu_2$ can be traced back to the fact that gluon exchange at weaker and weaker coupling is more and more dominated by forward-scattering, while point-like interactions describe s-wave scattering. What is perhaps surprising is that even at quite *large* values of g, gluon exchange yields an order of magnitude increase in $\delta\mu_2 - \delta\mu_1$.

This discovery has significant implications for the QCD phase diagram and may have significant implications for compact stars. At high enough baryon density the CFL phase in which all quarks pair to form a spatially uniform BCS condensate is favored. Suppose that as the density is lowered the nonzero strange quark mass induces the formation of some less symmetrically paired quark matter before the density is lowered so much that baryonic matter is obtained. In this less symmetric quark matter, some quarks may yet form a BCS condensate. Those which do not, however, will have differing Fermi momenta. These will form a crystalline color superconducting phase if the differences between their Fermi momenta lie within the appropriate window. In QCD, the interaction between quarks is forward-scattering dominated and the crystalline color superconductivity window is consequently wide open. This phase is therefore generic, occurring almost anywhere there are some quarks which cannot form BCS pairs. Evaluating the critical temperature T_c above which the crystalline condensate melts requires solving the nonzero temperature gap equation obtained as in Ref. 86 for the case of a point-like interaction. As in that case, we expect that all compact stars which are minutes old or older are much colder than T_c. This suggests that wherever quark matter which is not in the CFL phase occurs within a compact star, crystalline color superconductivity is to be found. As we discuss in the next section, wherever crystalline color superconductivity is found rotational vortices may be pinned resulting in the generation of glitches as the star spins down.

7. Color Superconductivity in Compact Stars

Our current understanding of the color superconducting state of quark matter leads us to believe that it may occur naturally in compact stars. These are the only places in the universe where we expect very high densities and low temperatures. They typically have masses close to 1.4 solar masses, and are believed to have radii of order 10 km. Their density ranges from around nuclear density near the surface to higher values further in, although uncertainty about the equation of state leaves us unsure of the value in the core.

Much of the work on the consequences of quark matter within a compact star has focussed on the effects of quark matter on the equation of state, and hence on the radius of the star. As a Fermi surface phenomenon, color superconductivity has little effect on the equation of state: the pressure is an integral over the whole Fermi volume. Color superconductivity modifies the equation of state at the $\sim (\Delta/\mu)^2$ level, typically by a few percent.[8] Such small effects can be neglected in present calculations, and for this reason we will not attempt to survey the many ways in which observations of neutron stars are being used to constrain the equation of state.[95] Color superconductivity gives mass to excitations around the ground state: it opens up a gap at the quark Fermi surface, and makes the gluons massive. One would therefore expect its main consequences to relate to transport properties, such as mean free paths, conductivities and viscosities.

The critical temperature T_c below which quark matter is a color superconductor is high enough that any quark matter which occurs within neutron stars that are more than a few seconds old is in a color superconducting state. In the absence of lattice simulations, present theoretical methods are not accurate enough to determine whether neutron star cores are made of hadronic matter or quark matter. They also cannot determine whether any quark matter which arises will be in the CFL or 2SC phase, and if the latter whether the quarks that do not participate in 2SC pairing form a crystalline color superconductor. Just as the higher temperature regions of the QCD phase diagram are being mapped out in heavy ion collisions, we need to learn how to use neutron star phenomena to determine whether they feature cores made of 2SC (possibly crystalline) quark matter, CFL quark matter or hadronic matter, thus teaching us about the high density region of the QCD phase diagram. It is therefore important to look for astrophysical consequences of color superconductivity.

7.1. *The Transition Region*

There are two possibilities for the transition from nuclear matter to quark matter in a neutron star: a mixed phase, or a sharp interface. The surface tension of the interface determines which is favored.

To be concrete, we will consider the case where the strange quark is light enough so that quark pairing is always of the CFL type. Figure 3 shows

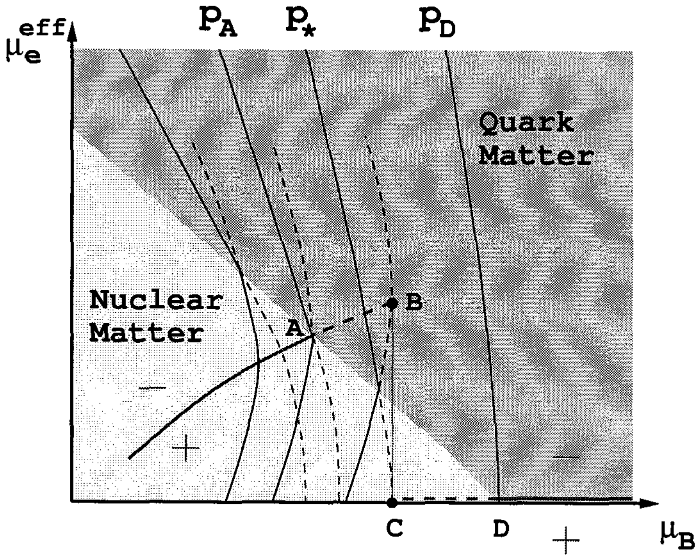


Figure 3. A schematic form of the μ_B-μ_e phase diagram for nuclear matter and CFL quark matter, ignoring electromagnetism. For an explanation see the text.

the μ_B-μ_e^{eff} phase diagram, ignoring electromagnetism. The lightly shaded region is where nuclear matter (NM) has higher pressure. The darker region is where quark matter (QM) has higher pressure. Where they meet is the coexistence line. The medium solid lines labelled by values of the pressure are isobars. Below the coexistence line they are given by the NM equation of state, above it by the QM equation of state.

The thick (red) lines are the neutrality lines. Each phase is negatively charged above its neutrality line and positively charged below it. Dotted lines show extensions onto the unfavored sheet (NM above the coexistence line, QM below it).

The electric charge density is

$$Q = -\left.\frac{\partial p}{\partial \mu_e}\right|_{\mu_B} \tag{40}$$

so the neutrality line goes through the right-most extremum of each isobar, since there the derivative of pressure with respect to μ_e is zero. For the CFL phase, the neutrality line is $\mu_e = 0$.[63]

Two possible paths from nuclear to CFL matter as a function of increasing μ are shown. In the absence of electromagnetism and surface tension, the favored option is to progress along the coexistence line from A to D, giving an overall neutral phase made of appropriate relative volumes of negatively charged CFL matter and positively charged nuclear matter.

If, on the other hand, Coulomb and surface energies are large, then the system remains on the nuclear neutrality line up to B, where there is a single interface between nuclear matter at B and CFL matter at C. The effective chemical potential μ_e^{eff} changes across the interface, meaning that there must be an electric field at the interface.[65] As a result, charged boundary layers develop with negative charge leaking into the CFL phase in the form of electrons and perhaps CFL mesons, leaving a net positive charge (in the form of both protons and a depletion of electrons) on the nuclear side of the interface. These charged boundary layers, the analogues of inversion layers in semiconductor physics, are of order tens of fermi thick.[65] This minimal interface, with its attendant charged boundary layers occurs between phases with the same μ_e, $\mu = \mu_B = \mu_C$, and pressure P_*. The effective chemical potential μ_e^{eff} changes across the interface, though, as a result of the presence of the electric field.

As yet, not much work has been done on signatures related to these features. The single interface creates a dramatic density discontinuity in the star: CFL quark matter at about four times nuclear density floats on nuclear matter at about twice nuclear density. This may affect the mass vs. radius relationship for neutron stars with quark matter cores. It may also have qualitative effects on the gravitational wave profile emitted during the inspiral and merger of two compact stars of this type.

If the surface tension σ_{QCD} and the electrostatic forces are ignored, then a mixed phase is favored over a sharp interface.[96,97,98] If we treat σ_{QCD} as an independent parameter, we can estimate the surface and Coulomb energy cost of the mixed phase.[98,65] We find that if $\sigma_{\text{QCD}} > 40\ \text{MeV/fm}^2$, as seems likely, the single sharp interface with its attendant charged boundary layers is free-energetically favored over the mixed phase.[65]

7.2. *Cooling by Neutrino Emission*

We turn now to neutron star phenomena which *are* affected by Fermi surface physics. For the first 10^{5-6} years of its life, the cooling of a neutron star is governed by the balance between heat capacity and the loss of heat by neutrino emission. How are these quantities affected by the presence of a quark matter core? This has been addressed recently in Ref. 99, 100, following earlier work in Ref. 101. Both the specific heat C_V and the neutrino emission rate L_ν are dominated by physics within T of the Fermi surface. If, as in the CFL phase, all quarks have a gap $\Delta \gg T$ then the contribution of quark quasiparticles to C_V and L_ν is suppressed by $\sim \exp(-\Delta/T)$. There may be other contributions to L_ν,[99] but these are also very small. In the CFL phase, the specific heat is dominated by that of the superfluid mode — i.e. the Goldstone boson associated with the spontaneous breaking of $U(1)_B$ — and there may also be small contributions from the light but not massless pseudo-Goldstone bosons associated with chiral symmetry breaking. Although further work is required, it is already clear that both C_V and L_ν are much smaller than in the nuclear matter outside the quark matter core. This means that the total heat capacity and the total neutrino emission rate (and hence the cooling rate) of a neutron star with a CFL core will be determined completely by the nuclear matter outside the core. The quark matter core is "inert": with its small heat capacity and emission rate it has little influence on the temperature of the star as a whole. As the rest of the star emits neutrinos and cools, the core cools by conduction, because the electrons keep it in good thermal contact with the rest of the star. These qualitative expectations are nicely borne out in the calculations presented in Ref. 100.

The analysis of the cooling history of a neutron star with a quark matter core in the 2SC phase is more complicated. The red and green up and down quarks pair with a gap many orders of magnitude larger than the temperature, which is of order 10 keV, and are therefore inert as described above. The remaining quarks may form a crystalline color superconductor. In addition, strange quarks may form an $\langle ss \rangle$ condensate with angular momentum $J = 1$ which locks to color in such a way that rotational invariance is not broken.[102] The resulting gap has been estimated to be of order hundreds of keV,[102] although applying results of Ref. 8 suggests a somewhat smaller gap, around 10 keV. The critical temperature T_c above which no condensate forms is of order the zero-temperature gap Δ. ($T_c = 0.57\Delta$ for $J = 0$ condensates.[25]) Therefore, if there are quarks for which $\Delta \sim T$ or smaller, these quarks do not pair at temperature T. Such quark quasiparticles will radiate neutrinos rapidly (via direct URCA reactions like $d \to u + e + \bar{\nu}$, $u \to d + e^+ + \nu$, etc.) and the quark matter core will cool rapidly and determine the cooling history of the star as a whole.[101,100] The star will cool rapidly until its interior temperature is $T < T_c \sim \Delta$, at which time the quark matter core will become inert and the further cooling history will be dominated by neutrino emission from the nuclear matter fraction of the star. If future data were to show that neutron stars first cool rapidly (direct URCA) and then cool more slowly, such data would allow an estimate of the smallest quark matter gap. We are unlikely to be so lucky. The simple observation of rapid cooling would *not* be an unambiguous discovery of quark matter with small gaps; there are other circumstances in which the direct URCA processes occur. However, if as data on neutron star temperatures improves in coming years the standard cooling scenario proves correct, indicating the absence of the direct URCA processes, this *would* rule out the presence of quark matter with gaps in the 10 keV range or smaller. The presence of a quark matter core in which *all* gaps are $\gg T$ can never be revealed by an analysis of the cooling history.

7.3. *Supernova Neutrinos*

We now turn from neutrino emission from a neutron star which is many years old to that from the protoneutron star during the first seconds of a supernova. Carter and Reddy[103] have pointed out that when this protoneutron star is at its maximum temperature of order 30-50 MeV, it may have a quark matter core which is too hot for color superconductivity. As such a protoneutron star core cools over the next few seconds, this quark matter will cool through T_c, entering the color superconducting regime of the QCD phase diagram. For $T \sim T_c$, the specific heat rises and the cooling slows. Then, as T drops further and Δ increases to become greater than T, the specific heat drops rapidly. Furthermore, as the number density of quark quasiparticles becomes suppressed by $\exp(-\Delta/T)$, the neutrino transport mean free path rapidly becomes very long.[103] This means that all the neutrinos previously trapped in the now color superconducting core are able to escape in a sudden burst. If a terrestrial neutrino detector sees thousands of neutrinos from a future supernova, Carter and Reddy's results suggest that there may be a signature of the transition to color superconductivity present in the time distribution of these neutrinos. Neutrinos from the core of the protoneutron star will lose energy as they scatter on their way out, but because they will be the last to reach the surface of last scattering, they will be the final neutrinos received at the earth. If they are released from the quark matter core in a sudden burst, they may therefore result in a bump at late times in the temporal distribution of the detected neutrinos. More detailed study remains to be done in order to understand how Carter and Reddy's signature, dramatic when the neutrinos escape from the core, is processed as the neutrinos traverse the rest of the protoneutron star and reach their surface of last scattering.

7.4. *R-mode Instabilities*

Another arena in which color superconductivity comes into play is the physics of r-mode instabilities. A neutron star whose angular rotation frequency Ω is large enough is unstable to the growth of r-mode oscillations which radiate away angular momentum via gravitational waves, reducing Ω. What does "large enough" mean? The answer depends on the damping mechanisms which act to prevent the growth of the relevant modes. Both shear viscosity and bulk viscosity act to damp the r-modes, preventing them from going unstable. The bulk viscosity and the quark contribution to the shear viscosity both become exponentially small in quark matter with $\Delta > T$ and as a result, as Madsen[104] has shown, a compact star made *entirely* of quark matter with gaps $\Delta = 1$ MeV or greater is unstable if its spin frequency is greater than tens to 100 Hz. Many compact stars spin faster than this, and Madsen therefore argues that compact stars cannot be strange quark stars unless some quarks remain ungapped. Alas, this powerful argument becomes much less powerful in the context of a neutron star with a quark matter core. First, the r-mode oscillations have a wave form whose amplitude is largest at large radius, outside the core. Second, in an ordinary neutron star there is a new source of damping: friction at the boundary between the crust and the neutron superfluid "mantle" keeps the r-modes stable regardless of the properties of a quark matter core.[105,104]

7.5. *Magnetic Field Evolution*

Next, we turn to the physics of magnetic fields within color superconducting neutron star cores.[64,106] The interior of a conventional neutron star is a superfluid (because of neutron-neutron pairing) and is an electromagnetic superconductor (because of proton-proton pairing). Ordinary magnetic fields penetrate it only in the cores of magnetic flux tubes. A color superconductor behaves differently. At first glance, it seems that because a diquark Cooper pair has nonzero electric charge, a diquark condensate must exhibit the standard Meissner effect, expelling ordinary magnetic fields or restricting them to flux tubes within whose cores the condensate vanishes. This is not the case, as we have seen. In both the 2SC and CFL phase, a linear combination of the $U(1)$ gauge transformation of ordinary electromagnetism and one (the eighth) color gauge transformation remains unbroken even in the presence of the condensate. This means that the ordinary photon A_μ and the eighth gluon G_μ^8 are replaced by the new linear combinations of (14) and (15), where $A_\mu^{\tilde{Q}}$ is massless and A_μ^X is massive. That is, $B_{\tilde{Q}}$ satisfies the ordinary Maxwell equations while B_X experiences a Meissner effect. $\sin(\theta)$ is proportional to e/g and turns out to be about $1/20$ in the 2SC phase and $1/40$ in the CFL phase.[64] This means that the $\tilde{Q}$-photon which propagates in color superconducting quark matter is mostly photon with only a small gluon admixture. If a color superconducting neutron star core is subjected to an ordinary magnetic field, it will either expel the X component of the flux or restrict it to flux tubes, but it can (and does[64]) admit the great majority of the flux in the form of a $B_{\tilde{Q}}$ magnetic field satisfying Maxwell's equations. The decay in time of this "free field" (i.e. not in flux tubes) is limited by the $\tilde{Q}$-conductivity of the quark matter. A color superconductor is not a $\tilde{Q}$-superconductor — that is the whole point — but it may turn out to be a very good $\tilde{Q}$-conductor due to the presence of electrons: if a nonzero density of electrons is required in order to maintain charge neutrality, the $B_{\tilde{Q}}$ magnetic field likely decays only on a time scale which is much longer than the age of the universe.[64] This means that a quark matter core within a neutron star can serve as an "anchor" for the magnetic field: whereas in ordinary nuclear matter the magnetic flux tubes can be dragged outward by the neutron superfluid vortices as the star spins down,[107] the magnetic flux within the color superconducting core simply cannot decay. Even though this distinction is a qualitative one, it will be difficult to confront it with data since what is observed is the total dipole moment of the neutron star. A color superconducting core anchors those magnetic flux lines which pass through the core, while in a neutron star with no quark matter core the entire internal magnetic field can decay over time. In both cases, however, the total dipole moment can change since the magnetic flux lines which do not pass through the core can move.

7.6. *Crystalline Color Superconductivity and Glitches in Quark Matter*

The final consequence of color superconductivity we wish to discuss is the possibility that (some) glitches may originate within quark matter regions of a compact star.

We do not yet know whether compact stars feature quark matter cores. And, we do not yet know whether, if they contain quark matter, that quark matter is color-flavor locked, meaning that quarks of all colors and flavors participate in BCS pairing, or whether the BCS condensate leaves some quarks unpaired. The lesson we take from the toy model analysis is that because the interaction between quarks in QCD is dominated by forward scattering, rather than being an s-wave point-like interaction, the difference in Fermi momenta between the unpaired quarks need not fall within a narrow window in order for them to form a crystalline color superconductor.

We wish now to ask whether the presence of a shell of crystalline color superconducting quark matter in a compact star (between the hadronic "mantle" and the CFL "inner core") has observable consequences. A quantitative formulation of this question would allow one either to discover crystalline color superconductivity, or to rule out its presence. (The latter would imply either no quark matter at all, or a single CFL-nuclear interface.[65])

Many pulsars have been observed to glitch. Glitches are sudden jumps in rotation frequency Ω which may be as large as $\Delta\Omega/\Omega \sim 10^{-6}$, but may also be several orders of magnitude smaller. The frequency of observed glitches is statistically consistent with the hypothesis that all radio pulsars experience glitches.[108] Glitches are thought to originate from interactions between the rigid neutron star crust, typically somewhat more than a kilometer thick, and rotational vortices in a neutron superfluid. The inner kilometer of crust consists of a crystal lattice of nuclei immersed in a neutron superfluid.[109] Because the pulsar is spinning, the neutron superfluid (both within the inner crust and deeper inside the star) is threaded with a regular array of rotational vortices. As the pulsar's spin gradually slows, these vortices must gradually move outwards since the rotation frequency of a superfluid is proportional to the density of vortices. Deep within the star, the vortices are free to move outwards. In the crust, however, the vortices are pinned by their interaction with the nuclear lattice. Models[110] differ in important respects as to how the stress associated with pinned vortices is released in a glitch: for example, the vortices may break and rearrange the crust, or a cluster of vortices may suddenly overcome the pinning force and move macroscopically outward, with the sudden decrease in the angular momentum of the superfluid within the crust resulting in a sudden increase in angular momentum of the rigid crust itself and hence a glitch. All the models agree that the fundamental requirements are the presence of rotational vortices in a superfluid and the presence of a rigid structure which impedes the motion of vortices and which encompasses enough of the volume of the pulsar to contribute significantly to the total moment of inertia.

Although it is premature to draw quantitative conclusions, it is interesting to speculate that some glitches may originate deep within a pulsar which features a quark matter core, in a region of that core which is in the crystalline color superconductor phase. If this phase occurs within a pulsar it will be threaded by an array of rotational vortices. It is reasonable to expect that these vortices will be pinned in a LOFF crystal, in which the diquark condensate varies periodically in space. The diquark condensate vanishes at the core of a rotational vortex, and for this reason the vortices will prefer to be located with their cores pinned to the nodes of the LOFF crystal.

A real calculation of the pinning force experienced by a vortex in a crystalline color superconductor must await the determination of the crystal structure of the LOFF phase. We can, however, attempt an order of magnitude estimate along the same lines as that done by Anderson and Itoh[111] for neutron vortices in the inner crust of a neutron star. In that context, this estimate has since been made quantitative.[112,113,110] For one specific choice of parameters,[85] the LOFF phase is favored over the normal phase by a free energy $F_{\rm LOFF} \sim 5 \times (10\ {\rm MeV})^4$ and the spacing between nodes in the LOFF crystal is $b = \pi/(2|\mathbf{q}|) \sim 9$ fm. The thickness of a rotational vortex is given by the correlation length $\xi \sim 1/\Delta \sim 25$ fm. The pinning energy is the difference between the energy of a section of vortex of length b which is centered on a node of the LOFF crystal vs. one which is centered on a maximum of the LOFF crystal. It is of order $E_p \sim F_{\rm LOFF}\, b^3 \sim 4$ MeV. The resulting pinning force per unit length of vortex is of order $f_p \sim E_p/b^2 \sim (4\ {\rm MeV})/(80\ {\rm fm}^2)$. A complete calculation will be challenging because $b < \xi$, and is likely to yield an f_p which is somewhat less than that we have obtained by dimensional analysis. Note that our estimate of f_p is quite uncertain both because it is only based on dimensional analysis and because the values of Δ, b and $F_{\rm LOFF}$ are uncertain. (We have a reasonable understanding of all the ratios Δ/Δ_0, $\delta\mu/\Delta_0$, q/Δ_0 and consequently $b\Delta_0$ in the LOFF phase. It is of course the value of the BCS gap Δ_0 which is uncertain.) It is premature to compare our crude result to the results of serious calculations of the pinning of crustal neutron vortices as in Ref. 112, 113, 110. It is nevertheless remarkable that they prove to be similar: the pinning energy of neutron vortices in the inner crust is $E_p \approx 1 - 3$ MeV and the pinning force per unit length is $f_p \approx (1 - 3\ {\rm MeV})/(200 - 400\ {\rm fm}^2)$.

The reader may be concerned that a glitch deep within the quark matter core of a neutron star may not be observable: the vortices within the crystalline color superconductor region suddenly unpin and leap outward; this loss of angular momentum is compensated by a gain in angular momentum of the layer outside the LOFF region; how quickly, then, does this increase in angular momentum manifest itself at the *surface* of the star as a glitch? The important point here is that the rotation of any superfluid region within which the vortices are able to move freely is coupled to the rotation of the outer crust on very short time scales.[114] This rapid coupling, due to electron scattering off vortices and the fact that the electron fluid penetrates throughout the star, is usually invoked to explain that the core nucleon superfluid speeds up quickly after a crustal glitch: the only long relaxation time is that of the vortices within the inner crust.[114] Here, we invoke it to explain that the outer crust speeds up rapidly after a LOFF glitch has accelerated the quark matter at the base of the nucleon superfluid. After a glitch in the LOFF region, the only long relaxation times are those of the vortices in the LOFF region and in the inner crust.

A quantitative theory of glitches originating within quark matter in a LOFF phase must await further calculations, in particular a three flavor analysis and the determination of the crystal structure of the QCD LOFF phase. However, our rough estimate of the pinning force on rotational vortices in a LOFF region suggests that this force may be comparable to that on vortices in the inner crust of a conventional neutron star. Perhaps, therefore, glitches occurring in a region of crystalline color superconducting quark matter may yield similar phenomenology to those occurring in the inner crust. This is surely strong motivation for further investigation.

There has been much recent progress in our understanding of how the presence of color superconducting quark matter in a compact star would affect five different phenomena: cooling by neutrino emission, the pattern of the arrival times of supernova neutrinos, the evolution of neutron star magnetic fields, r-mode instabilities and glitches. Nevertheless, much theoretical work remains to be done before we can make sharp proposals for which astrophysical observations can teach us whether compact stars contain quark matter, and if so whether it is in the 2SC or CFL phase and whether it is a crystalline color superconductor.

I am very grateful to the collaborators with whom I have been exploring the condensed matter physics of QCD: Mark Alford, Juergen Berges, Jeff Bowers, Joydip Kundu, Adam Leibovich, Cristina Manuel, Sanjay Reddy, Eugene Shuster and Frank Wilczek. I am also grateful to the organizers for the opportunity to lecture at such a stimulating school and hike in such a beautiful location. Research supported in part by the DOE under cooperative research agreement DE-FC02-94ER40818.

References

1. M. Alford, hep-ph/0102047; K. Rajagopal, F. Wilczek, hep-ph/0011333; T. Schäfer, E. Shuryak, nucl-th/0010049; K. Rajagopal, hep-ph/0009058; D. Rischke, R. Pisarski, hep-ph/0004016.
2. D. J. Gross and F. Wilczek, Phys. Rev. Lett. **30**, 1343 (1973); H. D. Politzer, Phys. Rev. Lett. **30**, 1346 (1973).
3. J. C. Collins and M. J. Perry, Phys. Rev. Lett. **34**, 1353 (1975).
4. J. Bardeen, L. N. Cooper and J. R. Schrieffer, Phys. Rev. **106**, 162 (1957); **108**, 1175 (1957).
5. B. Barrois, Nucl. Phys. **B129**, 390 (1977); S. Frautschi: "Asymptotic Freedom And Color Superconductivity In Dense Quark Matter". In: *Proceedings of 1978 Erice workshop "Hadronic matter at extreme density", Erice, Italy, Oct 13-21, 1978*, ed. by N. Cabibbo and L. Sertorio, (Plenum, New York, 1980) pp. 19-27
6. B. Barrois, "Nonperturbative effects in dense quark matter", Cal Tech PhD thesis, UMI 79-04847-mc (1979).
7. D. Bailin and A. Love, Phys. Rept. **107**, 325 (1984), and references therein.
8. M. Alford, K. Rajagopal and F. Wilczek, Phys. Lett. **B422**, 247 (1998) [hep-ph/9711395].
9. R. Rapp, T. Schäfer, E. V. Shuryak and M. Velkovsky, Phys. Rev. Lett. **81**, 53 (1998) [hep-ph/9711396].
10. S. Elitzur, Phys. Rev. **D12**, 3978 (1975).
11. E. Fradkin and S. Shenker, Phys. Rev. **D19**, 3682 (1979); T. Banks and E. Rabinovici, Nucl. Phys. **B160**, 349 (1979).
12. M. Alford, K. Rajagopal and F. Wilczek, Nucl. Phys. **B537**, 443 (1999) [hep-ph/9804403].
13. T. Schäfer and F. Wilczek, Phys. Rev. Lett. **82**, 3956 (1999) [hep-ph/9811473].
14. For a recent review and references, see O. Philipsen, hep-lat/0011019.
15. S. Chandrasekharan and U. Wiese, Phys. Rev. Lett. **83**, 3116 (1999) [cond-mat/9902128].
16. J. Berges and K. Rajagopal, Nucl. Phys. **B538**, 215 (1999) [hep-ph/9804233].
17. M. Alford, J. Berges and K. Rajagopal, Nucl. Phys. **B558**, 219 (1999) [hep-ph/9903502].
18. B. Vanderheyden and A. D. Jackson, Phys. Rev. **D62**, 094010 (2000) [hep-ph/0003150]; S. Pepin, A. Schäfer, hep-ph/0010225
19. G. W. Carter and D. Diakonov, Phys. Rev. **D60**, 016004 (1999) [hep-ph/9812445].
20. R. Rapp, T. Schäfer, E. V. Shuryak and M. Velkovsky, Annals Phys. **280**, 35 (2000) [hep-ph/9904353].
21. N. Evans, S. D. H. Hsu and M. Schwetz, Nucl. Phys. **B551**, 275 (1999) [hep-ph/9808444]; Phys. Lett. **B449**, 281 (1999) [hep-ph/9810514].
22. T. Schäfer and F. Wilczek, Phys. Lett. **B450**, 325 (1999) [hep-ph/9810509].
23. D. T. Son, Phys. Rev. **D59**, 094019 (1999) [hep-ph/9812287].

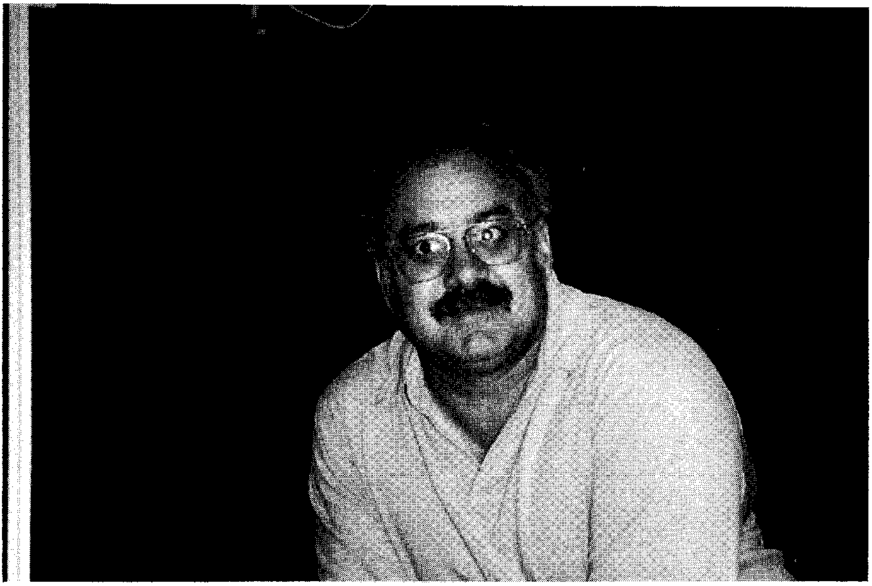
24. T. Schäfer and F. Wilczek, Phys. Rev. **D60**, 114033 (1999) [hep-ph/9906512].
25. R. D. Pisarski and D. H. Rischke, Phys. Rev. **D60**, 094013 (1999) [nucl-th/9903023]; Phys. Rev. **D61**, 051501 (2000) [nucl-th/9907041]; Phys. Rev. **D61**, 074017 (2000) [nucl-th/9910056];
26. D. K. Hong, Phys. Lett. **B473**, 118 (2000) [hep-ph/9812510]; Nucl. Phys. **B582**, 451 (2000) [hep-ph/9905523].
27. D. K. Hong, V. A. Miransky, I. A. Shovkovy and L. C. Wijewardhana, Phys. Rev. **D61**, 056001 (2000), erratum *ibid.* **D62**, 059903 (2000) [hep-ph/9906478].
28. T. Schäfer and F. Wilczek, Phys. Rev. **D60**, 114033 (1999) [hep-ph/9906512].
29. W. E. Brown, J. T. Liu and H. Ren, Phys. Rev. **D61**, 114012 (2000) [hep-ph/9908248]; Phys. Rev. **D62**, 054016 (2000) [hep-ph/9912409]; Phys. Rev. **D62**, 054013 (2000) [hep-ph/0003199].
30. S. D. Hsu and M. Schwetz, Nucl. Phys. **B572**, 211 (2000) [hep-ph/9908310].
31. T. Schäfer, Nucl. Phys. **B575**, 269 (2000) [hep-ph/9909574].
32. I. A. Shovkovy and L. C. Wijewardhana, Phys. Lett. **B470**, 189 (1999) [hep-ph/9910225].
33. N. Evans, J. Hormuzdiar, S. D. Hsu and M. Schwetz, Nucl. Phys. **B581**, 391 (2000) [hep-ph/9910313].
34. K. Rajagopal and E. Shuster, Phys. Rev. **D62**, 085007 (2000) [hep-ph/0004074].
35. C. Manuel, Phys. Rev. **D62**, 114008 (2000) [hep-ph/0006106].
36. Q. Wang and D. H. Rischke, Phys. Rev. D **65**, 054005 (2002) [nucl-th/0110016].
37. M. Iwasaki, T. Iwado: Phys. Lett. **B350**, 163 (1995); M. Iwasaki: Prog. Theor. Phys. Suppl. **120**, 187 (1995)
38. M. Srednicki and L. Susskind, Nucl. Phys. **B187**, 93 (1981).
39. T. Schäfer and F. Wilczek, Phys. Rev. **D60**, 074014 (1999) [hep-ph/9903503].
40. R. Casalbuoni and R. Gatto, Phys. Lett. **B464**, 111 (1999) [hep-ph/9908227].
41. M. Alford, J. Berges and K. Rajagopal, Phys. Rev. Lett. **84**, 598 (2000) [hep-ph/9908235].
42. D. T. Son and M. A. Stephanov, Phys. Rev. **D61**, 074012 (2000) [hep-ph/9910491]; erratum, *ibid.* **D62**, 059902 (2000) [hep-ph/0004095].
43. M. Rho, A. Wirzba and I. Zahed, Phys. Lett. **B473**, 126 (2000) [hep-ph/9910550].
44. D. K. Hong, T. Lee and D. Min, Phys. Lett. **B477**, 137 (2000) [hep-ph/9912531].
45. C. Manuel and M. H. Tytgat, Phys. Lett. **B479**, 190 (2000) [hep-ph/0001095].
46. M. Rho, E. Shuryak, A. Wirzba and I. Zahed, Nucl. Phys. **A676**, 273 (2000) [hep-ph/0001104].
47. K. Zarembo, Phys. Rev. **D62**, 054003 (2000) [hep-ph/0002123].

48. S. R. Beane, P. F. Bedaque and M. J. Savage, Phys. Lett. **B483**, 131 (2000) [hep-ph/0002209].
49. D. H. Rischke, Phys. Rev. **D62**, 054017 (2000) [nucl-th/0003063].
50. D. K. Hong, Phys. Rev. **D62**, 091501 (2000) [hep-ph/0006105].
51. T. Schäfer, nucl-th/0007021.
52. M. A. Nowak, M. Rho, A. Wirzba and I. Zahed, hep-ph/0007034.
53. V. A. Miransky, I. A. Shovkovy and L. C. Wijewardhana, hep-ph/0009173.
54. C. Manuel and M. Tytgat, hep-ph/0010274.
55. R. Casalbuoni, R. Gatto and G. Nardulli, hep-ph/0010321.
56. P. F. Bedaque and T. Schäfer, hep-ph/0105150. See also D. B. Kaplan and S. Reddy, hep-ph/0107265. Note that when $m_s \neq 0$, Δ is replaced by five different gap parameters which vary with m_s, at order m_s^2.[17] Since the Δ's occur quadratically in the free energy, there are contributions to the effective theory of the CFL mesons arising at order m_s^2 and m_s^4. Their effects (along with those of the electromagnetic and instanton-induced contributions to the meson masses) must be evaluated before firm conclusions about the effects of m_s^4 terms on K^0 condensation can be drawn.
57. V. A. Miransky and I. A. Shovkovy, hep-ph/0108178; T. Schafer, D. T. Son, M. A. Stephanov, D. Toublan and J. J. Verbaarschot, hep-ph/0108210; D. T. Son, hep-ph/0108260.
58. T. Schaefer, hep-ph/0109052.
59. D. H. Rischke, Phys. Rev. **D62**, 034007 (2000) [nucl-th/0001040].
60. G. Carter and D. Diakonov, Nucl. Phys. **B582**, 571 (2000) [hep-ph/0001318].
61. T. Schaefer, hep-ph/0201189.
62. D. B. Kaplan and A. E. Nelson, Phys. Lett. **B175**, 57 (1986).
63. K. Rajagopal and F. Wilczek, Phys. Rev. Lett. **86**, 3492 (2001) [hep-ph/0012039].
64. M. Alford, J. Berges and K. Rajagopal, Nucl. Phys. **B571**, 269 (2000) [hep-ph/9910254].
65. M. G. Alford, K. Rajagopal, S. Reddy and F. Wilczek, hep-ph/0105009.
66. E. Witten, Phys. Rev. **D30**, 272 (1984); E. Farhi and R. L. Jaffe, Phys. Rev. **D30**, 2379 (1984); P. Haensel, J. L. Zdunik and R. Schaeffer, Astron. Astrophys. **160**, 121 (1986); C. Alcock, E. Farhi and A. Olinto, Phys. Rev. Lett. **57**, 2088 (1986); Astrophys. J. **310**, 261 (1986).
67. C. Manuel and K. Rajagopal, Phys. Rev. Lett.
68. D. F. Litim and C. Manuel, hep-ph/0105165.
69. J. D. Jackson, "Classical Electrodynamics", 3rd Ed., (Wiley, New York, 1998).
70. G. 't Hooft, in "High Energy Physics," ed. A. Zichichi (Editrice Compositori, Bologna, 1976); S. Mandelstam, Phys. Rept. **23**, 245 (1976).
71. R. L. Jaffe, Phys. Rev. Lett. **38**, 195, 617(E) (1977).
72. D. Rischke, D. Son, M. Stephanov: hep-ph/0011379
73. R. D. Pisarski and D. H. Rischke, Phys. Rev. Lett. **83**, 37 (1999) [nucl-th/9811104].
74. V. A. Miransky, I. A. Shovkovy and L. C. Wijewardhana, Phys. Rev. **D62**,

085025 (2000) [hep-ph/0009129].
75. F. Sannino, Phys. Lett. **B480**, 280 (2000) [hep-ph/0002277]; R. Casalbuoni, Z. Duan and F. Sannino, hep-ph/0004207; S. D. Hsu, F. Sannino and M. Schwetz, hep-ph/0006059.
76. For a recent exploration of the μ-dependence of m_s, see M. Buballa and M. Oertel, hep-ph/0109095.
77. UKQCD Collaboration, Phys. Rev. **D59** (1999) 116002; S. Hands, J. B. Kogut, M. Lombardo and S. E. Morrison, Nucl. Phys. **B558**, 327 (1999) [hep-lat/9902034]; S. Hands, I. Montvay, S. Morrison, M. Oevers, L. Scorzato and J. Skullerud, hep-lat/0006018.
78. J. B. Kogut, M. A. Stephanov and D. Toublan, Phys. Lett. **B464**, 183 (1999) [hep-ph/9906346]; J. B. Kogut, M. A. Stephanov, D. Toublan, J. J. Verbaarschot and A. Zhitnitsky, Nucl. Phys. **B582**, 477 (2000) [hep-ph/0001171].
79. D. V. Deryagin, D. Yu. Grigoriev and V. A. Rubakov, Int. J. Mod. Phys. **A7**, 659 (1992).
80. E. Shuster and D. T. Son, Nucl. Phys. **B573**, 434 (2000) [hep-ph/9905448].
81. B. Park, M. Rho, A. Wirzba and I. Zahed, Phys. Rev. **D62**, 034015 (2000) [hep-ph/9910347].
82. R. Rapp, E. Shuryak and I. Zahed, hep-ph/0008207.
83. D. T. Son and M. A. Stephanov, hep-ph/0005225.
84. A. I. Larkin and Yu. N. Ovchinnikov, Zh. Eksp. Teor. Fiz. **47**, 1136 (1964) [Sov. Phys. JETP **20**, 762 (1965)]; P. Fulde and R. A. Ferrell, Phys. Rev. **135**, A550 (1964).
85. M. Alford, J. Bowers and K. Rajagopal, Phys. Rev. D **63**, 074016 (2001) [hep-ph/0008208].
86. J. A. Bowers, J. Kundu, K. Rajagopal and E. Shuster, Phys. Rev. D **64**, 014024 (2001) [hep-ph/0101067].
87. R. Casalbuoni, R. Gatto, M. Mannarelli and G. Nardulli, Phys. Lett. B **511**, 218 (2001) [hep-ph/0101326]; R. Casalbuoni, these proceedings, hep-th/0108195.
88. A. K. Leibovich, K. Rajagopal and E. Shuster, hep-ph/0104073.
89. J. Kundu and K. Rajagopal, hep-ph/0112206.
90. A. M. Clogston, Phys. Rev. Lett. **9**, 266 (1962); B. S. Chandrasekhar, App. Phys. Lett. **1**, 7 (1962).
91. P. F. Bedaque, hep-ph/9910247.
92. A. I. Buzdin and V. V. Tugushev Zh. Eksp. Teor. Fiz. **85**, 735 (1983) [Sov. Phys. JETP **58**, 428 (1983)]; A. I. Buzdin and S. V. Polonskii, Zh. Eksp. Teor. Fiz. **93**, 747 (1987) [Sov. Phys. JETP **66**, 422 (1987)].
93. K. Splittorff, D. T. Son and M. A. Stephanov, hep-ph/0012274.
94. A. Sedrakian, nucl-th/0008052. The related unpairing transition was discussed in the absence of LOFF pairing in A. Sedrakian and U. Lombardo, Phys. Rev. Lett. **84**, 602 (2000).
95. For a review, see H. Heiselberg and M. Hjorth-Jensen, Phys. Rept. **328**, 237 (2000).
96. N. K. Glendenning, Phys. Rev. D **46**, 1274 (1992).

97. M. Prakash, J. R. Cooke and J. M. Lattimer, Phys. Rev. D **52**, 661 (1995).
98. N. K. Glendenning and S. Pei, Phys. Rev. C **52**, 2250 (1995).
99. D. Blaschke, T. Klahn and D. N. Voskresensky, Astrophys. J. **533**, 406 (2000) [astro-ph/9908334]; D. Blaschke, H. Grigorian and D. N. Voskresensky, astro-ph/0009120.
100. D. Page, M. Prakash, J. M. Lattimer and A. Steiner, Phys. Rev. Lett. **85**, 2048 (2000) [hep-ph/0005094].
101. C. Schaab *et al*, Astrophys. J. Lett **480** (1997) L111 and references therein.
102. T. Schäfer, Phys. Rev. **D62**, 094007 (2000) [hep-ph/0006034].
103. G. W. Carter and S. Reddy, Phys. Rev. **D62**, 103002 (2000) [hep-ph/0005228].
104. J. Madsen, Phys. Rev. Lett. **85**, 10 (2000) [astro-ph/9912418].
105. L. Bildsten and G. Ushomirsky, Astrophys. J. **529**, L75 (2000) [astro-ph/9911155]; N. Andersson *et al.* in Ref. 114; L. Lindblom, B. J. Owen and G. Ushomirsky, astro-ph/0006242.
106. D. Blaschke, D. M. Sedrakian and K. M. Shahabasian, Astron. and Astrophys. **350**, L47 (1999) [astro-ph/9904395].
107. For reviews, see J. Sauls, in Timing Neutron Stars, J. Ögleman and E. P. J. van den Heuvel, eds., (Kluwer, Dordrecht: 1989) 457; and D. Bhattacharya and G. Srinivasan, in X-Ray Binaries, W. H. G. Lewin, J. van Paradijs, and E. P. J. van den Heuvel eds., (Cambridge University Press, 1995) 495.
108. M. A. Alpar and C. Ho, Mon. Not. R. Astron. Soc. **204**, 655 (1983). For a recent review, see A.G. Lyne in *Pulsars: Problems and Progress*, S. Johnston, M. A. Walker and M. Bailes, eds., 73 (ASP, 1996).
109. J. Negele and D. Vautherin, Nucl. Phys. **A207**, 298 (1973).
110. For reviews, see D. Pines and A. Alpar, Nature **316**, 27 (1985); D. Pines, in *Neutron Stars: Theory and Observation*, J. Ventura and D. Pines, eds., 57 (Kluwer, 1991); M. A. Alpar, in *The Lives of Neutron Stars*, M. A. Alpar et al., eds., 185 (Kluwer, 1995). For more recent developments and references to further work, see M. Ruderman, Astrophys. J. **382**, 587 (1991); R. I. Epstein and G. Baym, Astrophys. J. **387**, 276 (1992); M. A. Alpar, H. F. Chau, K. S. Cheng and D. Pines, Astrophys. J. **409**, 345 (1993); B. Link and R. I. Epstein, Astrophys. J. **457**, 844 (1996); M. Ruderman, T. Zhu, and K. Chen, Astrophys. J. **492**, 267 (1998); A. Sedrakian and J. M. Cordes, Mon. Not. R. Astron. Soc. **307**, 365 (1999).
111. P. W. Anderson and N. Itoh, Nature **256**, 25 (1975).
112. M. A. Alpar, Astrophys. J. **213**, 527 (1977).
113. M. A. Alpar, P. W. Anderson, D. Pines and J. Shaham, Astrophys. J. **278**, 791 (1984).
114. M. A. Alpar, S. A. Langer and J. A. Sauls, Astrophys. J. **282**, 533 (1984).

This page intentionally left blank



MARC SHER

GRAND UNIFICATION, HIGGS BOSONS, AND BARYOGENESIS

MARC SHER

Physics Dept., Nuclear and Particle Theory Group
College of William and Mary, Williamsburg, VA
E-mail: sher@physics.wm.edu

My task in these lectures is to discuss "Grand Unification and Higgs Bosons". Given that each of these subjects has had books written about them, this is a daunting task. My goal will be to introduce the basics of each topic, and provide references for those who wish to explore the topics further. I'll begin with a general motivation for grand unification, followed with an elementary review of $SU(N)$ group algebra. The seminal $SU(5)$ model will be discussed, followed by the supersymmetric version. On the second day, we'll look at other grand unified theories, and then look at the various methods of supersymmetry breaking in the context of grand unification. The third day, we'll turn to the Higgs mechanism, the effective potential, and mass bounds in the Standard Model and the MSSM. Finally, we'll look at baryogenesis, first in grand unified theories and then in the electroweak model.

1. Grand Unified Theories—$SU(5)$

1.1. *A History*

Let us begin with a history of grand unification.

•1973-4: The Standard Model of the strong and electroweak interactions is firmly established.[1]

•1974: The first grand unified theory, based on a product of $SU(4)$ groups, was proposed by Pati and Salam.[2] The first grand unified theory based on a simple gauge group, $SU(5)$, was proposed by Georgi and Glashow.[3] Its simplicity and economy made it an instant "favorite" for a unification of the strong and electroweak interactions.

•1975-1981: A model builder's paradise. Many versions of $SU(5)$, extensions such as $SO(10)$ and E_6, flavor symmetries are proposed. Very detailed analyses, through two loops, of the predictions of these theories for proton decay, the weak mixing angle and fermion masses are made. Proton decay experiments begin.

•1981: Prior to this time, supersymmetry was considered somewhat esoteric by GUT model-builders. A seminal paper by Witten[4] on the "dynamical breaking of supersymmetry" shows how supersymmetry can solve many problems of grand unified theories, and brings supersymmetric theories "to the masses".

•1981-1984: Many more models developed, now involving supersymmetric GUTs. These primarily focus on various methods of supersymmetry breaking.

•1984: The first superstring revolution. The field begins to bifurcate into those who work on string theory and those who do not (with a few exceptions, of course).

•1984-1989: Model-builders begin to get discouraged, as proton decay experiments just yield negative results and models become more and more complicated. A series of workshops on Grand Unified Theories ends with a conference entitled "Last Workshop on Grand Unification".

•1989-90: High precision data from LEP shows incredibly good agreement between the SUSY prediction of the weak mixing angle and the experimental value. The top quark is shown to be surprisingly heavy, leading to many more possibilities for models.

•1990's: Many development raise interest in the area. Neutrino masses, extra dimensions, a better understand of compactification, more precise calculations, and TeV scale quantum gravity.

•Now: The bifurcation that began in 1984 may be ending. TeV scale models indicate that the string sector may be phenomenologically relevant.

This TASI Summer School is primarily devoted to string theories, and the purpose of these lectures is to give students a feel for the important issue in grand unified theories, with the hope of closing this bifurcation.

There are a huge number of references introducing grand unified theories, and it is now a standard chapter in particle physics textbooks. One of the nicest early reviews, primarily of non-supersymmetric theories, is the Physics Reports of Langacker.[5] Many practitioners in the field learned the subject from that review article. Two later books are excellent: The book of Graham Ross,[6] "Grand Unified Theories" from 1984 and the more recent book of Mohapatra[7] from 1992.

1.2. *Introduction and Motivation*

The Standard Model is an $SU(3)\times SU(2)\times U(1)$ gauge theory. The fermions have quantum numbers:

$$\begin{aligned}
Q_i &= \begin{pmatrix} u_i \\ d_i \end{pmatrix} \sim (3, 2, +1/6) \\
\bar{u}_i &\sim (\bar{3}, 1, -2/3) \\
\bar{d}_i &\sim (\bar{3}, 1, +1/3) \\
L_i &= \begin{pmatrix} \nu_i \\ e_i \end{pmatrix} \sim (1, 2, -1/2) \\
\bar{e}_i &\sim (1, 1, +1)
\end{aligned} \tag{1}$$

All of these numbers are basically arbitrary. The gauge couplings are also arbitrary (why is $g_3 > g_2 > g_1$?). In a grand unified theory, all three interactions are put in a simple gauge group. There is only one coupling and only one or two quantum numbers for fermions. This represents a vast simplication of the Standard Model.

In order to set our notation, we begin with a simple review of the properties of $SU(N)$. This will be well-known to many of you. The fundamental representation is N-dimensional, and is represented by A_i, where $i = 1...N$). Conjugating this representation gives $\bar{N}$, written as $(A_i)^\dagger = (A^\dagger)^i$. The generators of the group are Hermitean and traceless. For $SU(2)$, there are the three generators:

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{2}$$

In general, for $SU(N)$, there are $N^2 - 1$ generators. To be even more explicit, we list the 15 generators of $SU(4)$:

$$\begin{pmatrix} 0&1&0&0 \\ 1&0&0&0 \\ 0&0&0&0 \\ 0&0&0&0 \end{pmatrix} + 5 \text{ with } 1'\text{s in the off} - \text{diagonal}$$

$$\begin{pmatrix} 0&i&0&0 \\ -i&0&0&0 \\ 0&0&0&0 \\ 0&0&0&0 \end{pmatrix} + 5 \text{ with i}, -\text{i in the off} - \text{diagonal}$$

$$\begin{pmatrix} 1&0&0&0 \\ 0&-1&0&0 \\ 0&0&0&0 \\ 0&0&0&0 \end{pmatrix} + \begin{pmatrix} 1&0&0&0 \\ 0&1&0&0 \\ 0&0&-2&0 \\ 0&0&0&0 \end{pmatrix} /\sqrt{3} + \begin{pmatrix} 1&0&0&0 \\ 0&1&0&0 \\ 0&0&1&0 \\ 0&0&0&-3 \end{pmatrix} /\sqrt{6} \tag{3}$$

Note that we have normalized the generators so that $Tr\ T_a T_b = 2\delta_{ab}$.

It is straightforward to combine representations. You are all familiar with combining two $SU(2)$ doublets, ψ_i and ψ_j into a symmetric triplet: $\psi_1\psi_1$, $\psi_2\psi_2$, and $(\psi_1\psi_2+\psi_2\psi_1)/\sqrt{2}$ plus an antisymmetric singlet: $(\psi_1\psi_2 - \psi_2\psi_1)/\sqrt{2}$. Schematically, this is written as $2 \times 2 = 3 + 1$.

A special note concerning $SU(2)$ only. The antisymmetric tensor, ϵ_{ij}, is also a generator of $SU(2)$, so $\bar{2} \equiv A^i = \epsilon^{ij} A_j$ is the same as 2. However, sometimes we still refer to $\bar{2}$ to keep track of signs. In a Higgs model, for example, one might have $2 = \begin{pmatrix} \phi^o \\ \phi^- \end{pmatrix}$ and $\bar{2} = (-\phi^o \phi^+)$. Note the minus sign. An important example is the μ-term of the MSSM. One has two fields of opposite hypercharge: $H_1 = \begin{pmatrix} \phi_1^+ \\ \phi_1^o \end{pmatrix}$ and $H_2 = \begin{pmatrix} \phi_2^o \\ \phi_2^- \end{pmatrix}$. The mixing term is written as:

$$\mu H_1 H_2 = \mu \epsilon_{ij} H_1^i H_2^j = \mu(\phi_1^+ \phi_2^- - \phi_1^o \phi_2^o) \tag{4}$$

but one can use $\bar{2}$ and then one doesn't need the ϵ_{ij}.

We can combine $SU(3)$ representations as well. $3 \times 3 = \bar{3} + 6$, where the $\bar{3}$ is a totally antisymmetric $3x3$ matrix and the 6 is a totally symmetric $3x3$ matrix. One can combine a 3 and a $\bar{3}$: $3 \times \bar{3} = 1 + 8$. In this case, the index structure is $3 = \psi_a$, $\bar{3} = \psi^b$, $1 = \psi_a \psi^a$ and $8 = \psi_a \psi^b - \frac{1}{\sqrt{3}} \psi_a \psi^a$. Note that $3 \times 3 \times 3 = 3 \times (\bar{3} + 6) = 1 + 8 + 8 + 10$ as is well known from the quark model.

For the prototype grand unified theory, $SU(5)$, we have

$$5 \times 5 = 10 + 15 \tag{5}$$

where the 10 (15) is an antisymmetric (symmetric) 5×5 matrix, and

$$5 \times \bar{5} = 1 + 24 \tag{6}$$

If one wishes to unify the Standard Model into a single gauge group, that group, G, must contain $SU(3) \times SU(2) \times U(1)$. Since the Standard Model has 4 diagonal generators, the rank of G must be greater than or equal to 4. Since the fermion representations are complex, the group G must have complex representations as well. There are an infinite number of possible groups satisfying these conditions, but the only rank 4 group is $SU(5)$.

1.3. *SU(5)*

The $SU(5)$ group has 24 generators, and thus there are 24 gauge bosons. Twelve of these make up the gluons, $W^\pm$, Z and photon. When $SU(5)$

breaks into $SU(3)\times SU(2)\times U(1)$, we can, without loss of generality, define the $SU(3)$ of color to operator on the first three indices, and the weak $SU(2)$ to operate on the last two indices. The matrix of gauge bosons then looks like

$$\begin{pmatrix} gluons\ others \\ others\ W^{\pm}, Z \end{pmatrix} \tag{7}$$

Note that the "others" have both $SU(3)$ and $SU(2)$ quantum numbers. The one generator which commutes with both $SU(3)$ and $SU(2)$, which will be the $U(1)$ generator is

$$T^{24} = c\begin{pmatrix} -1/3 & 0 & 0 & 0 & 0 \\ 0 & -1/3 & 0 & 0 & 0 \\ 0 & 0 & -1/3 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 0 & 1/2 \end{pmatrix} \tag{8}$$

where our normalization gives $c=\sqrt{3/5}$. This generator gives the hypercharge, with $T^{24}=cY$, so all of the known gauge bosons fit.

What about fermions? The fundamental representation of $SU(5)$ is a 5. We can decompose this into its $SU(3)\times SU(2)\times U(1)$ quantum numbers, using the above T^{24}:

$$5 = (3,1,-1/3)+(1,2,1/2)$$

$$\bar{5} = (\bar{3},1,1/3)+(1,\bar{2},-1/2) \tag{9}$$

We can now combine two 5's:

$$\begin{aligned} 5\times 5 = {} & (3,1,-1/3)\times(3,1,-1/3)+2(3,1,-1/3)\times(1,2,1/2) \\ & +(1,2,1/2)\times(1,2,1/2) \end{aligned} \tag{10}$$

or

$$\begin{aligned} 5\times 5 = {} & (\bar{3},1,-2/3)+(6,1,-2/3)+(3,2,1/6)+(3,2,1/6) \\ & +(1,1,1)+(1,3,1) \end{aligned} \tag{11}$$

The first, third and fifth of these terms are antisymmetric and contain 10 fields, while the second, fourth and sixth are symmetric and contain 15 fields. Comparing with the Standard Model quantum numbers, one sees that a $\bar{5}$ and a 10 contain **all** of the fermions, with no extra states.

One can show this explicitly by writing out the $\bar{5}+10$ for a single generation:

$$\psi^i = \begin{pmatrix} \bar{d}^r & \bar{d}^g & \bar{d}^b & e & -\nu \end{pmatrix}$$

$$\chi_{ij} = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & \bar{u}^g & -\bar{u}^b & u^r & d^r \\ -\bar{u}^g & 0 & \bar{u}^r & u^b & d^b \\ \bar{u}^b & -\bar{u}^r & 0 & u^g & d^g \\ -u^r & -u^b & -u^g & 0 & \bar{e} \\ -d^r & -d^b & -d^g & -\bar{e}^0 & \end{pmatrix} \tag{12}$$

Note that the gauge bosons in the upper 3×3 block of the gauge boson matrix will only affect the quarks, and those in the lower 2×2 block will not affect the $\bar{u}$ and $\bar{d}$, as expected. All of the $U(1)$ quantum numbers are fixed. This model automatically gives the proton the same charge as the positron, a fact that was input by hand in the Standard Model. So the multitude of quantum numbers has been replaced by just two: $\bar{5}$ and 10.

We now turn to the Higgs structure of the model. The symmetry breaking takes place in two stages: $SU(5) \to SU(3) \times SU(2) \times U(1) \to SU(3) \times U(1)_{em}$. The Higgs of the Standard Model is a $(1, 2, -1/2)$ of $SU(3) \times SU(2) \times U(1)$. This is also in a $\bar{5}$ of $SU(5)$, so the simplest Higgs to break the electroweak symmetry is a $\bar{5}$ (it is only the simplest–the Standard Model Higgs will also fit into a 45 of $SU(5)$). Note that the $\bar{5}$ also contains a $(\bar{3}, 1, 1/3)$, so the model will contain an isosinglet Higgs triplet.

The smallest representation that will break $SU(5)$ into the Standard Model is a 24 of Higgs, which has vacuum expectation value

$$\langle \Phi \rangle = V \begin{pmatrix} -1/3 & 0 & 0 & 0 & 0 \\ 0 & -1/3 & 0 & 0 & 0 \\ 0 & 0 & -1/3 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 0 & 1/2 \end{pmatrix} \tag{13}$$

The "minimal $SU(5)$" model has a Higgs sector consisting of a $\bar{5} + 24$. We now write the most general potential involving these two fields:

$$\begin{aligned} V(\Phi_{24}, H_5) &= m^2 Tr\Phi^\dagger\Phi + \frac{1}{2}aTr\Phi^4 + \frac{1}{4}b(Tr\Phi^\dagger\Phi)^2 \\ &+ \mu^2 H^\dagger H + \frac{1}{4}\lambda(H^\dagger H)^2 \\ &+ \alpha H^\dagger H Tr\Phi^\dagger\Phi + \beta H^\dagger \Phi^\dagger \Phi H \end{aligned} \tag{14}$$

To minimize the potential, we assume $m^2 < 0$. Then, if $15b + 7a > 0$, the minimum of the potential breaks $SU(5)$ into $SU(3) \times SU(2) \times U(1)$, whereas if $15b + 7a < 0$, it breaks into $SU(4) \times U(1)$. We assume the former region of parameter space. The minimum of the potential is then $\langle \Phi \rangle = v\ \mathrm{diag}(1, 1, 1, -3/2, -3/2)$ and $\langle H \rangle = \binom{0}{v_o}$. Thus, there is a range of parameter space in which the symmetry breaking is as desired.

But this does lead to the most serious problem underlying grand unified theories (not just $SU(5)$)–the hierarchy problem.[8] Minimizing the potential leads to two equations:

$$\begin{aligned} m^2 &= \frac{15}{2}bv^2 + \frac{7}{2}av^2 + \alpha v_o^2 + \frac{3}{10}\beta v_o^2 \\ \mu^2 &= \frac{1}{2}\lambda v_o^2 + 15\alpha v^2 + \frac{9}{2}\beta v^2 - 3\beta v_o^4/v^2 \end{aligned} \tag{15}$$

The second equation illustrates the problem. The left-hand side is of the order of the electroweak scale, yet the second and third terms on the right-hand side are of the order of the grand unification scale. This requires one to fine-tune (by, in this case, setting $15\alpha = -\frac{9}{2}\beta$), and the fine-tuning must be at least 28 orders of magnitude. Furthermore, even if one fine-tuned sufficiently, one-loop corrections would spoil the fine-tuning, and so one must fine-tune order by order in perturbation theory. This is the hierarchy problem. Note that supersymmetric models remove the second half of the fine-tuning, since the higher order corrections are zero, but not the main problem itself. It should also be noted that the Higgs triplet, H_3 acquires a mass-squared given by $-\frac{5}{4}\beta v^2$, and so it is extremely heavy. We will discuss the possible solutions to this problem in the next lecture.

So $SU(5)$ has a very nice structure, which perfectly accommodates all of the known fermions and gauge bosons. We now turn to the phenomenological implications.

1.4. *Phenomenology of SU(5)*

The most dramatic prediction of $SU(5)$ is proton decay. The gauge boson matrix can be written as

$$\tau^i W^i = \begin{pmatrix} G_r^r - \frac{2B}{\sqrt{30}} & G_r^b & G_r^g & X_r^\dagger & Y_r^\dagger \\ G_b^r & G_b^b - \frac{2B}{\sqrt{30}} & G_b^g & X_b^\dagger & Y_b^\dagger \\ G_r^g & G_b^g & G_g^g - \frac{2B}{\sqrt{30}} & X_g^\dagger & Y_g^\dagger \\ X^r & X^b & X^g & \frac{W^3}{\sqrt{2}} + \frac{3B}{\sqrt{30}} & W^+ \\ Y^r & Y^b & Y^g & W^- & -\frac{W^3}{\sqrt{2}} + \frac{3B}{\sqrt{30}} \end{pmatrix} \tag{16}$$

We see the gluons, $W^\pm$, and the fields that rotate to become the Z and photon. But there are new gauge bosons, the X and Y. They have both color and isospin and charges 4/3 and 1/3 respectively. The vertices can be easily found by looking at the $\bar{5} + 10$ representation of fermions, where one sees that the X, Y bosons mediate $\bar{d}e$; $\bar{d}\nu$; $u\bar{u}$ and $d\bar{e}$ transitions. These lead immediately to proton decay. For example, the u and d quarks in a proton can exchange an X boson, and become a $\bar{u}$ and an e^+, respectively. The $\bar{u}$ combines with the remaining u quark in the proton to become a π^o, and so we have the tree-level decay $p \rightarrow \pi^o e^+$. The lifetime, of course, depends on M_X and M_Y. We will discuss these masses shortly.

There are other predictions of minimal $SU(5)$. One can predict a relationship between fermion masses. This is because if one only has a $\bar{5}$ of Higgs and a $\bar{5}+10$ of fermions, there are only two Yukawa terms one can write down per generation: $\bar{5}_H\bar{5}_F10_F$ and $\bar{5}_H^\dagger 10_F10_F$. Since there are three masses per generation (u,d,e) one gets predictions. The predictions are:

$$\begin{aligned} m_d &= m_e \\ m_s &= m_\mu \\ m_b &= m_\tau \end{aligned} \tag{17}$$

All of these predictions are completely wrong.

Another prediction involves the various coupling constants. The covariant derivative is

$$D_\mu = \partial_\mu - ig_5(T_5)^\alpha A_\mu^\alpha \tag{18}$$

This leads to $g_5 = g_2 = g_3$. For g_1 one must be careful about the normalization. We have $g_5T^{24} = g_1Y$ and $T^{24} = cY$ from earlier, so that $g_1 = \sqrt{3/5}g_5$. Thus $\sin^2\theta_W = g_1^2/(g_1^2+g_2^2) = c^2/(1+c^2) = 3/8$. We thus predict

$$\begin{aligned} \sin^2\theta_W &= 3/8 \\ \alpha_s &= \alpha_w \end{aligned} \tag{19}$$

Both of these predictions are completely wrong.

Thus, we have a prediction of proton decay which depends on unknown masses, and a number of incorrect predictions involving masses and coupling constants. However, the gauge and Yukawa couplings vary with energy scale, and the above predictions only apply to the theory at the high energy scale. The renormalization group equations, at one-loop, give

$$\frac{dg_i}{dt} = -\frac{1}{16\pi^2}b_ig_i^3 \tag{20}$$

where $b_3 = 7$, $b_2 = \frac{19}{6}$ and $b_1 = -\frac{41}{6}$. We now take the previous results: $g_3 = g_2 = \sqrt{5/3}g_1$ at M_X, and run the couplings down to the weak scale. Note that the fact that $b_3 > b_2 > b_1$ automatically implies the correct ordering $g_3 > g_2 > g_1$. One can take the observed values of g_3 and g_2 at the weak scale, run them up until they meet at M_X, then run g_1 down to get g_1 at the weak scale. Thus one predicts[9] the value of M_X (and thus the proton lifetime) and the value of $\sin^2\theta_W$. The results are

$$\sin^2\theta_W = \frac{23}{134} + \frac{109}{201}\frac{\alpha_{em}(M)}{\alpha_3(M)}$$
$$\ln\frac{M_X}{M} = \frac{6\pi}{67}\left(\frac{1}{\alpha_{em}} - \frac{8}{3}\frac{1}{\alpha_3(M)}\right) \tag{21}$$

This is valid for any mass scale. Since we know α_{em} and α_3 experimentally, we can find M_X and $\sin^2\theta_W$. For $\alpha_3(M_Z) = 0.118$, the result is

$$M_X = 4 \times 10^{14}\text{GeV}$$

$$\sin^2\theta_W = 0.210 \pm 0.002 \tag{22}$$

This value of M_X gives a lifetime for $p \to \pi^o e^+$ of $10^{29\pm1.5}$ years.

Experimentally, the lifetime for this decay mode is now over 10^{32} years, and the value of $\sin^2\theta_W = 0.2324$. Thus, minimal $SU(5)$ is ruled out. We will see shortly that the new beta-functions of the supersymmetric Standard Model bring these values in very close agreement.

There is one other prediction[10] of minimal $SU(5)$. The mass relations also get modified by renormalization group running. At low energies, one finds that $m_b = 3.0m_\tau$, which works remarkably well. Although one can't get precise predictions for m_d and m_s, because QCD at these scales is strong, the ratio is insensitive to QCD, and one finds $m_d/m_s = m_e/m_\mu$. The left side of this is 1/20, and the right side is 1/200, so this prediction fails.

There are some relatively easy ways out. Georgi and Jarlskog[11] pointed out that if one uses a 45 of Higgs instead of a 5, then a Clebsch replaces the m_e/m_μ by $9m_e/m_\mu$, which is in agreement with experiment. Alternatively, it has been noted[12] that small, dimension-5 operators at the GUT scale, generated by Planck scale effects, give arbitrary $O(10)$ MeV corrections, which remove the prediction.

One final note about $SU(5)$. The Lagrangian has a global symmetry G:

$$\bar{5}_F \to e^{-3i\alpha}\bar{5}_F; \qquad 10_F \to e^{i\alpha}10_F; \qquad \bar{5}_H \to e^{-2i\alpha}\bar{5}_H \tag{23}$$

Although G is broken by $\langle H\rangle$, $U(1)_Y$ is also. A linear combination, $Y - \frac{G}{4}$ is not. This unbroken global symmetry turns out to be $B - L$. Thus a neutrino mass, which would violate B-L if Majorana, is not allowed. This symmetry will be very important in later discussion of baryogenesis.

Summarizing minimal $SU(5)$:
1. Group structure works perfectly. No extra fermions.
2. Get $g_3 > g_2 > g_1$.
3. $\sin^2\theta_W$ is off, but only by 5-10%.
4. τ_p is too small, but a few percent increase in M_X (logarithmically) makes it OK.
5. $M_b = 3.0M_\tau$ works very well.
6. $m_d/m_s = m_e/m_\mu$ fails, but there are easy fixes.

We now turn to SUSY $SU(5)$, and show how this will change the beta-functions. It will keep (1), (2) and (5) in this list, fit (3) extremely well, and (4) will be OK.

1.5. *SUSY SU*(5)

Supersymmetric $SU(5)$ is a straightforward extension of $SU(5)$. The Higgs sector (as in the MSSM) is more complicated, and this will be discussed in the third lecture. The primary effect here is on the phenomenological results noted above. The presence of superpartners in loops will change the beta functions, and thus the predictions for M_X and $\sin^2\theta_W$.

Prior to 1993, the prediction of $\sin^2\theta_W$ in SUSY $SU(5)$ was 0.2335 ± 0.002, which compares very well with the experimental value of 0.2324 ± 0.0006. In 1993, the top quark was discovered, and knowledge of its mass lowered the uncertainties. Rather than quote $\sin^2\theta_W$, people now assume the experimental value and determine the value of $\alpha_s(M_Z)$. Given the experimental range, the predicted value of $\alpha_s(M_Z) = .125 \pm .003$, compared with the latest lattice result of $.118 \pm .002$. This is a very slight discrepancy. It has apparently been made somewhat worse by recent LEP II results. However, threshold effects at the GUT scale, which are unknown (and possibly unknowable), introduce enough uncertainty to remove the problem.

What about the proton lifetime? In the SUSY model, M_X increases by a factor of 30, and this increases the proton lifetime by a factor of a million, beyond experimental reach. However, there are new potential contributions to proton decay:

•dimension 4: The most general superpotential contains terms that give proton decay even without unification. One must assume a symmetry, called R-parity, to remove them.

•dimension 5: The Higgs triplet, as discussed above, must exist. It is heavy, but in the supersymmetric version of the theory, there will be a

Higgsino triplet. Since it also (like the X and Y bosons) violates baryon and lepton number, it can generate a QQQL operator, and since it is a fermion, only one power of M_X will appear in the propagator. This is very dangerous, since the rate will scale as $1/M_X^2$ and not $1/M_X^4$ Fortunately, the QQQL operator has two scalars, and to turn them into fermions requires exchange of a gluino, W-ino or photino. This makes the diagram one-loop which suppresses the rate. The couplings of Higgs bosons to light fermions is also small, further suppressing the rate.

An important note about Higgs-mediated proton decay. If one explicitly includes the indices in the QQQL operator, one gets

$$\epsilon_{ijk}\epsilon_{ab}\epsilon cdQ^i_{Aa}Q^j_{Bb}Q^k_{Cc}L_{Dd} \tag{24}$$

where ijk are color indices, $abcd$ are weak indices and $ABCD$ label the generation number. If $A = B = C$, this vanishes, so all of the Q's must not be first generation. A strange quark is needed in the decay. Then, since uds is neutral, the lepton must be neutral. So the main decay mode is $p \to K^+\nu$.

The rate is given by (in years)[7]

$$\tau(p \to K\nu) = 7 \times 10^{28}[\frac{.01\mathrm{GeV}^3}{\beta}]^2[\frac{0.67}{A}]^2[\frac{\sin 2\beta}{1+y^t}]^2[\frac{M_{H_3}}{10^{16}\mathrm{GeV}}]^2[\frac{\mathrm{TeV}^{-}1}{f}]^2 \tag{25}$$

where β is a nuclear matrix element, which is between 0.003 and 0.03 GeV^3, A is a short distance renormalization factor, which is $0.6-0.7$, y^t depends on t-quark mixing angles ($.1 < y^t < 1.3$) and $f = M_{\tilde{W}}/M^2_{\tilde{Q}}$. The experimental bound is 1×10^{32} years, which pushes the parameter space, but does not rule out the model.

Finally, as shown by Carena et al.,[13] the ratio M_b/M_τ works very well for either small $\tan\beta$ or large $\tan\beta$ (where $\tan\beta$ is the ratio of the two vacuum expectation values in SUSY).

There remains one serious problem in SUSY $SU(5)$. It is called the doublet-triplet problem.[14,15,16,17] In SUSY models, two doublets, a 5 and a $\bar{5}$, are needed. There will be a term in the superpotential of the form $\lambda\bar{5}_H 24_H 5_H$. When the 24 gets a vev equal to $\mathrm{diag}(1,1,1,-3/2,-3/2)v$, the mass of the doublet fermion will be of order λv and that of the triplet will be of order $-\frac{3}{2}\lambda v$. These are the same order, but we know that the MSSM Higgsino (the doublet) must be of the order of the weak scale. One can add a singlet Higgs and fine-tune its vev to cancel this mass, but that is quite ugly. One can equally well fine-tune with a $\mu\bar{5}_H 5_H$ term, but that is also ugly.

One solution is the sliding singlet mechanism.[18,19,20] If one adds a singlet S, the superpotential is $W = \bar{5}_H(24_H + S)5_H$. The vev of the 24_H can be written (arising from other terms in the superpotential) as $v\,\mathrm{diag}(-2/3, -2/3, -2/3, 1, 1)$. Now at the minimum, $\frac{\partial W}{\partial \bar{5}_H} = 0$, so

$$(\langle 24_H \rangle + \langle S \rangle) \cdot \langle 5_H \rangle = 0 \tag{26}$$

This implies that $\langle S \rangle = -v$, so $\langle 24_H \rangle + \langle S \rangle = v(-5/3, -5/3, -5/3, 0, 0)$, and thus the doublet remains light. There may be a problem with radiative corrections to the singlet mass spoiling the cancellation. Other ideas involving higher-dimensional models have been proposed that will naturally eliminate this problem,[21,22,23] and there has been a recent analysis[24] in the context of heterotic string-inspired models.

2. Grand Unified Theories–Beyond $SU(5)$

Let us review the **satisfactory** features of SUSY $SU(5)$:

1. It has a simple gauge group.
2. Fermions fit neatly into a $\bar{5} + 10$, nothing else is needed.
3. $\sin^2\theta_W$ is accurate to better than a percent.
4. $\tau(p \to K^+\nu)$ is marginally OK, and with a very distinctive signature.
5. M_b/M_τ predicted.
6. Other fermion mass ratios can be accommodated.

However, let us now review the **unsatisfactory** features of SUSY $SU(5)$:

1. The fermions are not unified. There are two irreducible representations per family.
2. Parity asymmetry is put in by hand.
3. No neutrino masses are allowed.
4. No possibility of intermediate scales (useful for axions, Majorana neutrinos, cosmology).
5. There still remains a global symmetry (B-L).
6. Special symmetries needed to remove dimension-4 proton decay.
7. The doublet-triplet problem needs fine-tuning.

Since $SU(5)$ is the only rank-4 group, we first turn to rank-5 groups. What about $SU(6)$? If it breaks to $SU(5) \times U(1)$, then nothing is solved.

If it breaks into $SU(3) \times SU(3) \times U(1)$, then the $SU(3)$'s have the same coupling. No model works well, although there are some promising features of SUSY $SU(6)$, especially in solving the doublet-triplet problem.[25,26]

The Lie Groups are $SU(N)$, $SO(N)$, $Sp(N)$ and the five exceptional groups G_2, F_4, E_6, E_7 and E_8. The $Sp(N)$ and E_7 groups have no complex representations and must be discarded, and G_2 and F_4 have rank less than four. The only remaining rank 5 group is $SO(10)$. We will later look at E_6.

2.1. $SO(10)$

The $SO(N)$ group has $\frac{1}{2}N(N-1)$ generators. The fundamental representation is an N-dimensional real vector. One can only have complex representations if $N = 2m$, where m is odd. $SO(6)$ is too small, so this leaves $SO(10)$.

The fundamental representation is a 10, and

$$(D_\mu \Phi)_i = \partial_\mu \Phi_i - g A_\mu^{ij} \Phi_j \tag{27}$$

The gauge bosons are in the adjoint representation, which is a 45-dimensional representation. In addition, $SO(2m)$ has two complex spinor representations of dimension 2^{m-1}, which for $SO(10)$ is 16-dimensional.

Fundamental and adjoint representations are familiar, but spinor representations may not be. Consider N operators χ_i, satisfying

$$\{\chi_i, \chi_j^\dagger\} = \delta_{ij}; \qquad \{\chi_i, \chi_j\} = 0 \tag{28}$$

Then the operators $T_j^i \equiv \chi^{i\dagger}\chi_j$ satisfy a $U(N)$ algebra. Now define $2N$ operators, Γ_μ as

$$\Gamma_{2j-1} = -i(\chi_j - \chi_j^\dagger); \qquad \Gamma_{2j} = (\chi_j + \chi_j^\dagger) \tag{29}$$

Then it is easy to show that $\{\Gamma_\mu, \Gamma_\nu\} = 2\delta_{\mu\nu}$, i.e. the Γ's are generalized γ-matrices.

The generators of $SO(2N)$ can then be written as

$$\Sigma_{\mu\nu} = \frac{1}{2i}[\Gamma_\mu, \Gamma_\nu] \tag{30}$$

since these satisfy the algebra of $SO(2N)$. Note that if one defines $\Gamma_5 = (-1)^m \Gamma_1 \Gamma_2 ... \Gamma_{2N}$, then $[\Gamma_5, \Sigma_{ij}] = 0$, so that $\frac{1 \pm \gamma_5}{2}$ are projection operators. They project the space into two subspaces, one with an odd number of χ's and one with an even number.

Now define $|0\rangle$ to be the $SU(N)$ invariant vacuum. Then, using the χ's, one can build up the most general states: $|0\rangle$, $\chi_j^\dagger|0\rangle$, $\chi_j^\dagger\chi_k^\dagger|0\rangle$, etc. For

$N = 5$, corresponding to $SO(10)$, the general state can be written as

$$\begin{aligned}|\psi\rangle &= |0\rangle\psi_o + \chi_j^\dagger|0\rangle\psi_j \\ &+ \frac{1}{2}\chi_j^\dagger\chi_k^\dagger|0\rangle\psi_{jk} \\ &+ \frac{1}{12}\epsilon^{jklmn}\chi_l^\dagger\chi_m^\dagger\chi_n^\dagger|0\rangle\psi'_{jk} \\ &+ \frac{1}{24}\epsilon^{jklmn}\chi_k^\dagger\chi_l^\dagger\chi_m^\dagger\chi_n^\dagger|0\rangle\psi'_j \\ &+ \chi_j^\dagger\chi_k^\dagger\chi_l^\dagger\chi_m^\dagger\chi_n^\dagger|0\rangle\psi'_o\end{aligned} \tag{31}$$

This can be succinctly written as

$$\psi = \begin{pmatrix}\psi_o \\ \psi_j \\ \psi_{jk} \\ \psi'_{jk} \\ \psi'_j \\ \psi'_o\end{pmatrix} \tag{32}$$

The dimensionality of these states is (going top to bottom) $1, 5, 10, \bar{10}, \bar{5}, \bar{1}$. These are 32 states. But we noted above that the space can be divided into two spaces, one for even number of χ's and one for an odd number. The odd projection gives a 16, which is composed of a $1 + 10 + \bar{5}$ of $U(5)$.

So, for a spinor representation,

$$(D_\mu\psi)^a = \partial_\mu\psi^a + \frac{1}{2}ig\Sigma_{ab}^{ij}A_\mu^{ij}\psi^b \tag{33}$$

where $a, b = 1....16$ and $i, j = 1...10$. This representation contains a $\bar{5}$, 10 and singlet of $SU(5)$, which is **precisely the constituents of a single generation plus a right handed neutrino**. Thus, each generation of fermions fits into a single spinor representation of $SO(10)$.

Under the decomposition of $SO(10) \to SU(5)$

$$\begin{aligned}&10 \to 5 + \bar{5} \\ &16 \to \bar{5} + 10 + 1 \\ &45 \to 24 + 10 + \bar{10} + 1\end{aligned} \tag{34}$$

Note the the fundamental representation is real, as noted earlier.

Under the decomposition of $SO(10) \to SU(3)_c \times SU(2)_L \times SU(2)_R$, one has

$$16 \to (3, 2, 1) + (1, 2, 1) + (\bar{3}, 1, 2) + (1, 1, 2) \tag{35}$$

The first two parts are the left handed quark and lepton doublets. The third is $\begin{pmatrix} d^c \\ -u^c \end{pmatrix}$ and the fourth is $\begin{pmatrix} e^c \\ -\nu^c \end{pmatrix}$. When $SU(2)_R$ breaks, these latter two representations split. We thus see that $SO(10)$ is left-right symmetric, and that parity violation, which is put in by hand into $SU(5)$, will be broken here spontaneously.

How does the $SO(10)$ symmetry break? A 10 of Higgs will break $SO(10)$ into $SO(9)$, and a 45 of Higgs will not change the rank. There are three other representations, whose $SU(5)$ decomposition is shown:

$$\begin{aligned} 54 &\rightarrow 15 + \bar{15} + 24 \\ 120 &\rightarrow 5 + \bar{5} + 10 + \bar{10} + 45 + \bar{45} \\ 126 &\rightarrow 1 + \bar{5} + 10 + \bar{15} + 45 + \bar{50} \end{aligned} \tag{36}$$

To break $SO(10)$ into $SU(5)$, this decomposition must have a singlet, so the only possibilities are a 16 and/or a 126. The 54 will break $SO(10)$ into $SU(4) \times SU(2) \times SU(2)$ (and then a 45 can break the $SU(4)$ to $SU(3)_c \times U(1)$). There are many, many paths to the standard model, each involving a number of Higgses of rather high dimensionality. These paths are illustrated in Figure 1.

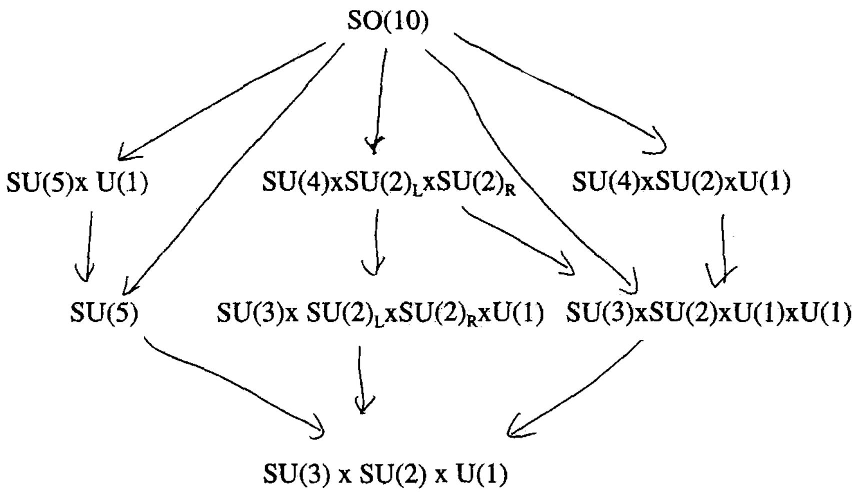


Figure 1. Many paths showing $SO(10)$ breaking into the Standard Model.

We see that $SO(10)$ can have several intermediate scales, which allows for heavy Majorana neutrinos, axions, right-handed W's, etc.

What about fermion masses? The product of two spinors is $16 \times 16 = 10 + 120 + 126$. With just a 10, one gets equal d and e masses, as in $SU(5)$, but also equal u and ν masses. Thus one needs a 120 or a 126. Recall that a 126 can break the gauge symmetry as well. The 126 will also give a Majorana mass to the $SU(5)$ singlet neutrino. Since the 10 would give a Dirac mass to the regular neutrino (which is identical to the $Q = 2/3$ quark mass), one gets

$$M = \begin{pmatrix} 0 & M_u \\ M_u & M_N \end{pmatrix} \rightarrow M_\nu = \frac{M_U^2}{M_N} \tag{37}$$

For $M_U = M_{top}$, and the experimentally indicated $M_{\nu_\tau} = 10^{-3} - 10^{-2}$ eV, this gives a Majorana neutrino mass of approximately 10^{16} GeV, which is precisely the grand unification scale!! Thus, the model can completely explain the light neutrino masses.

Finally, the use of a 45 of Higgs will automatically solve the doublet-triplet problem by allowing for the Higgs triplet to become superheavy. And the $B - L$ global symmetry of $SU(5)$ is now gauged.

Thus, we see that $SO(10)$ solves many of the problems of SUSY $SU(5)$, but at a cost of extra scales and much less predictive power.

2.2. E_6

E_6 is a rank six group. It was a popular unification group in the late 70's because it allowed for topless models. The fundamental representation of E_6 is a 27. Under $SO(10)$, the decomposition of a 27 is into a $16 + 10 + 1$. The hope was that the bottom quark and tau lepton could fit into the 10, giving a two-generation model. When it was recognized that the bottom quark has isospin 1/2, interest in this model faded.

It was revived in 1984 with the heterotic string theory. The $E_8 \times E_8'$ heterotic string requires that compactification leaves unbroken $N = 1$ SUSY. Compactification on a Calabi-Yau manifold results in $E_8 \rightarrow E_6 \times SU(3)$, where the $SU(3)$ is the spin connection. The other E_8 could account for the "hidden sector" needed to break SUSY (which will be discussed shortly). The matter fields are in

$$n_g\ 27 + \delta(27 + \bar{27}) \tag{38}$$

where n_g is the Euler number of the Calabi-Yau manifold and δ is the Betti-Hodge number. It was very exciting that the most attractive string theory automatically, when compactified on a C-Y manifold, gave a plausible grand unified theory plus the right fermion structure. Although somewhat less appealing following discovery of orbifold compactifications, models with bigger holonomy groups, etc., it is still the GUT which seems closest to being realizable in string theory compactifications.

The 27 breaks down as follows:

$$\begin{aligned} SO(10) &: 16+10+1 \\ SU(5) &: \bar{5}+10+1+\bar{5}+5+1 \\ SU(3)_c \times SU(3)_L \times SU(3)_R &: (3,3,1)+(3^c,1,3^c)+(1,3^c,3) \end{aligned} \tag{39}$$

The third decomposition can be displayed directly:

$$\begin{pmatrix} u \\ d \\ g \end{pmatrix} + \begin{pmatrix} \bar{u} & \bar{d} & \bar{g} \end{pmatrix} + \begin{pmatrix} N & \bar{E} & \nu \\ E & \bar{N} & e \\ \bar{\nu} & \bar{e} & S \end{pmatrix} \tag{40}$$

We see that there are new isosinglet quarks, g, a right-handed neutrino, and a pair of new lepton doublets: $\binom{N}{E}$ and $\binom{\bar{E}}{\bar{N}}$. In SUSY, these could become the light Higgs doublets.

There are many, many breaking chains down to the Standard Model. Since the rank of E_6 is six, there could be two additional Z-bosons, and there has been extensive study of the phenomenology of these extra Z's and the extra fermions by Hewett and Rizzo.[27]

What about the Higgs structure? The product of two 27's is $27 \times 27 = \bar{27} + 351_S + 351_A$. The 351's can break the E_6 into the standard model, and the 27 can break the electroweak symmetry. While this may look like a huge number of fields, one should note that ALL of the Higgs bosons necessary to break the symmetry are either in the 27 or the 27×27, and thus a composite type of model might be very attractive.

2.3. *SUSY Breaking*

In the remaining five minutes of this lecture, I will mention some aspects of supersymmetry breaking. The reader is referred to the review articles of Nilles[29] and of Haber and Kane[28] for details.

Supersymmetry can be broken spontaneously (through a potential whose ground state does not have zero energy, thereby breaking supersymmetry, which requires zero ground state energy), dynamically (though a gluino condensate) or softly (through soft mass terms). Spontaneous breaking fails, since it leads to no scalar-gaugino-gaugino couplings and requires very light colored and charged scalar fields. Dynamical symmetry breaking also has difficults with index theorems. Softly broken SUSY seems very arbitrary, but is very natural in so-called "hidden sector" models.

In these models, there are two distinct sectors of the theory. One is the visible sector, consisting of the MSSM fields and interactions. The other is the "hidden sector", under which the standard model fields are singlets. The basic idea is to break supersymmetry spontaneously or dynamically **in the hidden sector**. The SUSY breaking is then transmitted to the visible sector by mediating particles (which could simply be gravitons, which will certainly mediate information from one to the other). The resulting low-energy visible sector will be the MSSM with soft supersymmetry breaking. An important advantage of this scenario is that if the mediating interactions are flavor-blind (such as would be the case for gravity), then the soft scalar quark masses (at the GUT scale) will be identical, which solves the problem of flavor-changing neutral currents in SUSY models.

Suppose the messenger interaction is gravity. Then if SUSY is broken spontaneously in the hidden sector, the resulting soft terms are of the order of $\langle F\rangle/M_{Pl}$, where $\sqrt{F}$ is the breaking scale; this leads to the SUSY breaking scale of 10^{11} GeV. If it is broken dynamically, one has a gluino condensate with a scale of 10^{13} GeV. Since these scales are obtained by loops or by running couplings, they are easy to generate without fine-tuning.

Another option for the messenger particles are fields that have standard model gauge interactions. The soft breaking terms then arise through loops. Here $\langle F\rangle$ must be $(10^{4-5}\text{GeV})^2$. Using a $5+\bar{5}$ of $SU(5)$ will be acceptable and will not mess up the success of coupling constant unification.

How do these soft terms break the electroweak symmetry? The simplest approach (this is a very imprecise discussion—see the review articles for details) is radiative breaking. At the GUT scale, one has $m_H^2 = m_Q^2 = m_T^2$, where these are the soft supersymmetry breaking terms for the Higgs and scalar top quarks. From the λHQT superpotential term, one gets

$$\begin{aligned}\frac{dm_H^2}{dt} &= 3\frac{\lambda}{4\pi}(m_H^2+m_Q^2+m_T^2)\\ \frac{dm_T^2}{dt} &= 2\frac{\lambda}{4\pi}(m_H^2+m_Q^2+m_T^2)\\ \frac{dm_Q^2}{dt} &= 1\frac{\lambda}{4\pi}(m_H^2+m_Q^2+m_T^2)\end{aligned} \tag{41}$$

The coefficients $3, 2, 1$ arise from simple counting: the correction to m_H^2 has a QT loop, and there is a 3 for color, the correction to m_T^2 has a QH loop, and there is a 2 for isospin, while the correction to m_Q^2 has a TH loop. This means that m_H^2 runs faster, and so it will go negative before the other two. Thus, the $SU(2)$ symmetry will break, but **not** the $SU(3)$ or $U(1)$ symmetry. Thus we can understand the nature of electroweak symmetry breaking.

3. The Higgs Mechanism

3.1. *Introduction*

The Standard Model consists of three sectors: the gauge sector is determined by the $SU(2) \times U(1)$ gauge symmetry, the fermion sector consists of left-handed isodoublets and right-handed isosinglets, and the Higgs sector consists of a complex isodoublet. The gauge and fermion sectors, and the interactions of fermions with gauge bosons, has been very well-studied. However the Higgs boson has not yet been discovered, and no elementary scalar bosons are known in Nature. This, plus the *ad hoc* nature of the Higgs mechanism, has made it the most uncertain part of the Standard Model. The search for the Higgs boson, and an understanding of the mechanism of electroweak symmetry breaking, has been the highest priority of the particle physics community for decades.

In a sense, some Higgs bosons have already been discovered. The longitudinal components of the W^+, W^- and Z form three members of a four-member complex isodoublet. The Higgs boson will be the missing fourth component. Thus, some sort of scalar state must exist. Prior to the discovery of the top quark, it was known that it had to exist since the isospin of the left-handed bottom quark was known to be 1/2, with $I_3 = -1/2$, and so the other member of the isodoublet had to exist, to preserve the gauge symmetry and renormalizability. Similarly, a scalar state must exist to preserve the underlying gauge symmetry and renormalizability.

This scalar state can be an elementary scalar, or it can be a composite scalar. Models in which the Higgs boson is a strongly-bound pair of fermions are a direct analogy with BCS superconductivity, in which the scalar Cooper pair, which gives the photon an effective mass in a superconductor, is a bound state of two electrons. Such models, called Technicolor models, have a new quantum number which binds the fundamental fermions, the "techniquarks", into a Higgs boson. They also predict the existence of a large number of states involving techniquarks. This procedure can work to break the electroweak symmetry, but can't give mass to fermions. To give fermions mass, one must introduce another interaction, Extended Technicolor, which mixes the regular fermions and the technifermions to give the regular fermions mass. These problems then tend to yield large flavor-changing neutral currents, which are not observed, as well as electroweak radiative corrections which are much larger than observed. Each of these problems can be fixed, but no model is particularly compelling. For a detailed review of these models, the reader is referred to the TASI-2000 lectures of Chivukula.[30] In these lectures, only elementary scalars will be considered. With the popularity of supersymmetry, the presence of elementary scalars is not as worrisome as it once was.

We will begin with the Higgs mechanism in the Standard Model, rapidly reviewing the basic structure of the Higgs sector. After a discussion of the R_ξ gauges, we will turn to the effective potential, and the resulting bounds on the Higgs mass. Extensions of the scalar sector will then be discussed, and we will concentrate on the Higgs structure of the minimal supersymmetric model. This popular model has an upper bound to the Higgs mass that is well within reach of the Tevatron during the next four years.

3.2. *The Higgs Mechanism in the Standard Model*

The simplest gauge theory is a $U(1)$ gauge theory, where

$$\begin{aligned}\mathcal{L} &= -\frac{1}{4}F_{\mu\nu}F^{\mu\nu}\\ F_{\mu\nu} &= \partial_\mu A_\nu - \partial_\nu A_\mu.\end{aligned} \tag{42}$$

This Lagrangian is invariant under $A_\mu(x) \to A_\mu(x) - \partial_\mu\eta(x)$, where $\eta(x)$ is arbitrary. A mass term, $\frac{1}{2}m^2A_\mu A^\mu$ violates this invariance and thus breaks the symmetry. Simply putting in an arbitrary mass term results in a non-renormalizable theory. Mass terms must arise by breaking the symmetry spontaneously, i.e. arranging so that the symmetry of the Lagrangian is not the same as the symmetry of the ground state of the theory.

To do this, one adds a complex scalar field, ϕ:

$$\begin{aligned}\mathcal{L} &= -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + |D_\mu\phi|^2 - V(\phi)\\ D_\mu &= \partial_\mu - ieA_\mu\\ V(\phi) &= \mu^2(\phi^*\phi) + \lambda(\phi^*\phi)^2.\end{aligned} \tag{43}$$

This Lagrangian is invariant under

$$\begin{aligned}A_\mu &\to A_\mu - \partial_\mu\eta(x)\\ \phi(x) &\to e^{-ie\eta(x)}\phi(x)\end{aligned} \tag{44}$$

This $U(1)$ invariance is spontaneously broken when μ^2 is negative. The potential is then given by the familiar "Mexican hat" potential shown in Figure 2.

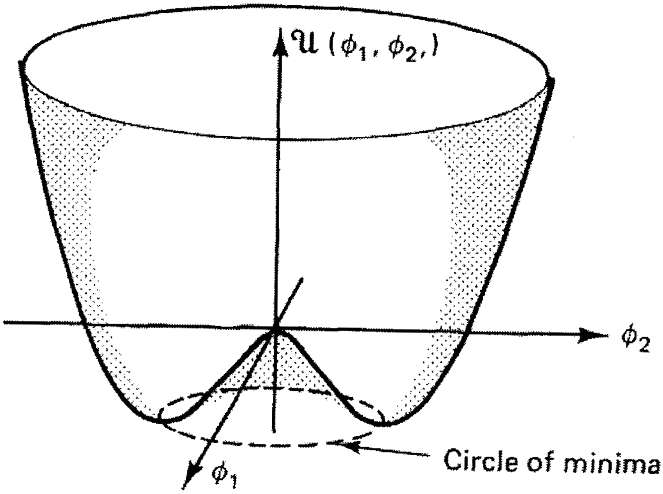


Figure 2. The Higgs potential $U(\phi_1, \phi_2)$, where ϕ_1 and ϕ_2 are the real and imaginary parts of the Higgs field.

The potential is obviously dependent only on $\phi^*\phi$, and one can thus rotate the real and imaginary axes with impunity. However, choosing a ground state requires one to pick out a direction in field space, breaking this invariance. Without loss of generality, we choose the ground state to lie along the real ϕ axis. The minimum is then at

$$\langle\phi\rangle = \sqrt{\frac{-\mu^2}{\lambda}} \equiv \frac{v}{\sqrt{2}}, \tag{45}$$

and one can perturb about the ground state by defining

$$\phi \equiv \frac{1}{\sqrt{2}} e^{i\chi(x)/v}(v + h(x)) \tag{46}$$

where $\chi(x)$ and $h(x)$ are fields that have no vacuum expectation values. Note that the $h(x)$ field is radial and would be expected to have a mass, while the χ field is angular and should be massless.

The Lagrangian in terms of these new fields is

$$\begin{aligned}\mathcal{L} = &-\frac{1}{4}F_{\mu\nu}F^{\mu\nu} - evA_\mu\partial^\mu\chi + \frac{e^2v^2}{2}A_\mu A^\mu \\ &+ \frac{1}{2}\partial_\mu h\partial^\mu h) + \mu^2h^2 + \frac{1}{2}\partial_\mu\chi\partial^\mu\chi \\ &+ h \text{ and } \chi \text{ interactions}\end{aligned} \tag{47}$$

We see that this Lagrangian contains a photon mass, ev, an h scalar mass-squared of $-2\mu^2$, a χ scalar mass-squared of zero, and a strange $A_\mu\partial^\mu\chi$ term.

Although one can proceed with the mixed term, resulting in mixed $A-\chi$ propagators, it is easier to use the gauge freedom to fix the gauge, by choosing

$$A'_\mu \equiv A_\mu - \frac{1}{ev}\partial_\mu\chi \tag{48}$$

Plugging this into the Lagrangian above, the χ-field disappears completely, along with the mixed term. The original four degrees of freedom: a massless vector and a complex scalar doublet, have been changed into a massive vector and a real massive scalar. One says that the vector boson has "eaten" the χ field in order to get mass. The vector propagator becomes

$$\Delta_{\mu\nu} = -\frac{i}{k^2-M_A^2}\left(g_{\mu\nu} - \frac{k_\mu k_\nu}{M_A^2}\right) \tag{49}$$

This propagator has very bad high-energy behavior due to the second term, and violates one-loop renormalizability. To alleviate this problem, one chooses a different gauge, by adding a term to the Lagrangian

$$\mathcal{L}_{GF} = -\frac{1}{2\xi}(\partial_\mu A^\mu + \xi ev\chi)^2 \tag{50}$$

Note that the cross term cancels the mixed propagator term (after integrating by parts). In the limit $\xi \to \infty$, one gets the previous gauge, called the unitary gauge.

The important features of the new Lagrangian are the vector propagator, which is

$$D_{\mu\nu} = -\frac{i}{k^2-M_A^2}\left(g_{\mu\nu} - \frac{(1-\xi)k^\mu k^\nu}{k^2-\xi M_A^2}\right) \tag{51}$$

the χ field mass, given by $M_\chi^2 = \xi M_A^2$, and the coupling of ξ to h which scales as ξ. In the unitary gauge, $\xi = \infty$, the propagator is that discussed in the preceding paragraph, the Goldstone field, χ has infinite mass and decouples, but the theory is not renormalizable. In the Feynman gauge,

$\xi = 1$. Here, the vector propagator is very simple, $D_\mu \nu = -\frac{i}{k^2 - M_A^2} g_{\mu\nu}$, but the Goldstone field has the same mass of the vector boson. In the Landau gauge, $\xi = 0$. The propagator is messier than in the Feynman gauge, of course, but the Goldstone field is massless and doesn't couple to the Higgs bosons. Because of this latter property, Higgs calculations are usually done in the Landau gauge.

Physically, ξ must drop out of any calculation. To see how this happens, consider $e^+e^- \to \mu^+\mu^-$ by exchange of a heavy photon. The diagram with heavy photon exchange gives

$$\bar{u}_e \gamma_\mu v_e \frac{\left(g^{\mu\nu} - \frac{(1-\xi)k^\mu k^\nu}{k^2 - \xi M_A^2} \right)}{k^2 - M_A^2} \bar{v}_\mu \gamma_n u u_\mu \tag{52}$$

The extra piece is proportional to $\bar{u}\gamma_\mu k^\mu v \bar{v} \gamma_\mu k^\mu u$, and with $k = p + p'$, one can use the Dirac equations ($\gamma_\mu p^\mu v = m$ and $\bar{u}\gamma_\mu p'^\mu = -m$) to get zero for the extra piece. In other calculations, such as the electron-muon scattering, the extra piece does not give zero, but when one includes χ exchange as well, the ξ-dependence drops out.

The extension to the Standard Model is straightforward. One introduces a complex isodoublet

$$\Phi = \begin{pmatrix} \phi^+ \\ \frac{h + i\chi}{\sqrt{2}} \end{pmatrix} \tag{53}$$

and gets a ground state of

$$\langle \Phi \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} \tag{54}$$

The particle masses-squared become $M_W^2 = \frac{1}{4} g^2 v^2$, $M_Z^2 = \frac{1}{4}(g^2 + g'^2)v^2$ and $M_h^2 = 2\lambda v^2$. Since we know that $\frac{G_F}{\sqrt{2}} = \frac{g^2}{8M_W^2} = \frac{1}{2v^2}$, we know that $v = 246.3$ GeV. However, since the scalar self-coupling is unknown, the Higgs mass is arbitrary. This is the catch-22 of particle physics. We have a particle whose couplings are completely determined, and yet whose mass is arbitrary; and whose coupling to light particles is much smaller than its coupling to heavy particles, and (alas!) accelerators are made of light particles...

In general, one would say that the Higgs mass could be anything between 0 and a TeV. (At a TeV, the self-coupling and width become so large enough that the Higges is no longer an elementary object.) However, it can be bounded rather severely by effective potential arguments.

3.3. *Effective Potential and Mass Bounds*

Although the Higgs mass appears arbitrary, one can bound it by considering radiative corrections to the Higgs potential. Here, we discuss how one calculates such corrections. The results turn out to have wide-ranging implications, and are much more important than just bounding the Higgs mass in the Standard Model. In string theory, there are many flat directions to potentials, and radiative corrections can be crucial. Models in which the electroweak scale is generated from the GUT scale by a scalar mass-squared becoming negative at low energies, discussed in the last lecture, also use the effective potential. So the results here are much more general.

An extensive review of the effective potential and bounds from vacuum stability appeared in 1989.[31] Since then, the potential has been improved, including a proper renormalization-group improvement of scalar loops, and the bounds have been refined to much higher precision. In addition, the discovery of the top quark has narrowed the region of parameter space that must be considered. In this section, we discuss the effective potential and its renormalization-group improvement. This section is somewhat expanded over the presentation given in the TASI lectures.

It is easy to see how bounds on masses can arise. One can get bounds on fermion masses given the Higgs mass, which translate into bounds on Higgs masses given the fermion (in this case, the top quark) mass. The one-loop effective potential, as originally written down by Coleman and Weinberg[32] can be written, in the direction of the physical ϕ field, as

$$V(\phi) = V_0 + V_1 \tag{55}$$

where

$$V_0 = -\frac{1}{2}\mu^2\phi^2 + \frac{1}{4}\lambda\phi^4 \tag{56}$$

and V_1 is the one-loop potential. One can get a rough idea of the form of the one-loop potential as follows. The tree-level potential has a ϕ_c^2 and a ϕ_c^4 term, but with a single loop, one can generate $\phi_c^6, \phi_c^8 ...$ terms. Here, ϕ_c is the classical field. The ϕ_c^{2n} can be generated by drawing a ϕ_c loop, and putting n ϕ_c^4 interactions, with $2n$ external legs, on the loop. The full one-loop potential is found by summing all of these terms. Consider the case where the quadratic term vanishes, so the scalar field is massless. The n-th diagram has n propagators, $\left(\frac{1}{k^2}\right)^n$, n vertices, $(3\lambda)^n$, $2n$ fields, ϕ_c^{2n} and a $\frac{1}{2n}$ combinatoric factor. Thus each diagram is of the form x^n/n, which sums to a logarithm. In this case, up to an overall constant, they sum to

$$V_1 = \frac{1}{2}\int \frac{d^4}{(2\pi)^4}\ln(k^2 + 3\lambda\phi_c^2) \tag{57}$$

This seems eminently reasonable if one does the integral over k_o and obtains $\int \frac{d^3k}{(2\pi)^3}\sqrt{k^2 + 3\lambda\phi_c^2}$, which is the sum of all zero-point fluctuations about the vacuum $\phi = \phi_c$.

A much more detailed analysis, including gauge boson loops and fermion loops, was first performed by Coleman and Weinberg,[32] and they found that

$$V_1 = \frac{1}{64\pi^2}\sum_i (-1)^F \eta_i \mathcal{M}^4(\phi_c)\ln\frac{\mathcal{M}^2(\phi_c)}{M^2} \tag{58}$$

where the sum is over all particles in the model, F is the fermion number, η_i is the number of degrees of freedom of the field i, and $\mathcal{M}^2(\phi_c)$ is the mass that the field has in the vacuum in which the scalar field has a value ϕ_c. In the expression for V_1, we have ignored terms which can be absorbed into V_0– these will be fixed by the renormalization procedure. In the standard model, we have for the W-boson, $\mathcal{M}^2(\phi_c) = \frac{1}{4}g^2\phi_c^2$, for the Z-boson, $\mathcal{M}^2(\phi_c) = \frac{1}{4}(g^2+g'^2)\phi_c^2$, for the Higgs boson, $\mathcal{M}^2(\phi_c) = -\mu^2+3\lambda\phi_c^2$, for the Goldstone bosons, $\mathcal{M}^2(\phi_c) = -\mu^2 + \lambda\phi_c^2$ and for the top quark $\mathcal{M}^2(\phi_c) = \frac{1}{2}h^2\phi_c^2$. For a very large values of ϕ, quadratic terms are negligible and the potential becomes

$$V = \frac{1}{4}\lambda\phi^4 + B\phi^2\ln(\phi^2/M^2) \tag{59}$$

where

$$B = \frac{3}{64\pi^2}[4\lambda^2 + \frac{1}{16}(3g^4 + 2g^2g'^2 + g'^4) - h^4] \tag{60}$$

One can see that if the top quark is very heavy, then h is large and thus B is negative. In this case, the potential is unbounded from below at large values of ϕ. This is the origin of the instability of the vacuum caused by a heavy quark.

Although this form of the effective potential is well known, it is NOT useful in determining vacuum stability bounds. The reason is as follows. Suppose one denotes the largest of the couplings in a theory by α, in the standard model,for example, $\alpha = [\max(\lambda, g^2, h^2)]/(4\pi)$. The loop expansion is an expansion in powers of α, but is also an expansion in powers of logarithms of ϕ_c^2/M^2, since each momentum integration can contain a single logarithmic divergence, which turns into a $\ln(\phi_c^2/M^2)$ upon renormalization. Thus the n-loop potential will have terms of order

$$\alpha^{n+1}[\ln(\phi^2/M^2)]^n \tag{61}$$

In order for the loop expansion to be reliable, the expansion parameter must be smaller than one. M can be chosen to make the logarithm small for any specific value of the field, but if one is interested in the potential over a range from ϕ_1 to ϕ_2, then it is necessary for $\alpha \ln(\phi_1/\phi_2)$ to be smaller than one. In examining vacuum stability, one must look at the potential at very large scales, as well as the electroweak scale, and the logarithm is generally quite large. Thus, any results obtained from the loop expansion are unreliable (and, in fact, the bound on the top quark mass can be off by more than a factor of two).

A better expansion, which does not have large logarithms, comes from solving the renormalization group equation (RGE) for the effective potential. This equation is nothing other than the statement that the potential cannot be affected by a change in the arbitrary parameter, M, i.e. $dV/dM = 0$. Using the chain rule, this is

$$[M\frac{\partial}{\partial M} + \beta(g_i)\frac{\partial}{\partial g_i} - \gamma\phi\frac{\partial}{\partial\phi}]V = 0 \tag{62}$$

where $\beta = M dg_i/dM$ and there is a beta function for every coupling and mass term in the theory. The γ function is the anomalous dimension.

It is important to note that the renormalization group equation is exact and no approximations have been made. If one knew the beta functions and anomalous dimensions exactly, one could solve the RGE exactly and determine the full potential at all scales. Although we do not know the exact beta functions and anomalous dimensions, we do have expressions for them as expansions in couplings. Thus, by *only* assuming that the couplings are small, the beta functions and γ can be determined to any level of accuracy and $V(\phi)$ can be found. The resulting potential will be accurate if $g_i << 1$ and will not require $g_i \ln(\phi/M) << 1$.

For example, in massless $\lambda\phi^4$ theory, the RGE can be solved exactly to give

$$V = \frac{1}{4}\lambda'(t,\lambda)G^4(t,\lambda)\phi^4 \tag{63}$$

where $t = \ln(\phi/M)$ and $\lambda'(t,\lambda)$ is defined to be the solution of the equation

$$\frac{d\lambda'}{dt} = \frac{\beta(\lambda')}{(1+\gamma(\lambda'))} \tag{64}$$

with the boundary condition being determined by the renormalization condition. $G(t,\lambda)$ is defined as $\exp(-4\int_0^t dt'(\gamma(\lambda')/(1+\gamma(\lambda')))$. Note that this potential gives the same result as before in the limit that $\gamma = 0$ and $\beta =$

constant. Then $G = 1$ and $\lambda' = \beta t$ + constant. With $t = \ln(\phi/M)$ this gives the $\phi^4 \ln(\phi/M)$ terms as above.

What about the massive case? The RGE is given by

$$[M\frac{\partial}{\partial M} + \beta_\lambda \frac{\partial}{\partial \lambda} + \beta(g_i)\frac{\partial}{\partial g_i} + \beta_{\mu^2}\mu^2 \frac{\partial}{\partial \mu^2} - \gamma\phi\frac{\partial}{\partial \phi}]V = 0 \qquad (65)$$

One is tempted to reduce this equation to a set of ordinary differential equations as before, giving

$$V(\phi) = \frac{1}{2}\mu^2(t)g^2(t)\phi^2 + \frac{1}{4}\lambda(t)G^4(t)\phi^4, \qquad (66)$$

where the coefficients are running couplings obeying first order differential equations as in the massless case.

However, this is not correct. By considering small excursions in field space, one does not, as in the massless case, reproduce the unimproved one-loop potential. This is not surprising. In the massless theory, the only scale is set by ϕ, and thus all logarithms must be of the form $t = \ln(\phi^2/M^2)$. In the massive theory, there is another scale, and there will be logarithms of the form $\ln((-\mu^2 + 3\lambda\phi^2)/M^2)$. Thus one can not easily sum all of the leading logarithms. In addition, the scale dependence of the constant term in the potential (the cosmological constant) can be relevant.

In earlier work (and in the review of Sher[31]), it was argued that the bounds only depend on the structure of the potential at large ϕ, and thus the mass term and constant term are irrelevant. However, in going from λ to the Higgs mass, the structure of the potential near its minimum is important, and thus using the naive expression above is not as accurate (although it is fairly close). This will be discussed more in the next section.

More recently, Bando, et al.[33] and Ford, et al.,[34] following some earlier work by Kastening,[35] found a method of including the additional logarithms found in the massive theory. In general, they showed that if one considers the L-loop potential, and runs the parameters of that potential using $L+1$ beta and gamma functions, then all logarithms will be summed up to the Lth-to-leading order. The standard model potential, including all leading and next-to-leading logarithms, is then (in the 't Hooft Landau gauge)

$$\begin{aligned} V(\phi) = -\tfrac{1}{2}\mu^2\phi^2 + \tfrac{1}{4}\lambda\phi^4 + \tfrac{1}{16\pi^2}\left[\tfrac{3}{2}W^2\left(\ln\tfrac{W}{M^2} - \tfrac{5}{6}\right) + \tfrac{3}{4}Z^2\left(\ln\tfrac{Z}{M^2} - \tfrac{5}{6}\right)\right. \\ \left. +\tfrac{1}{4}H^2\left(\ln\tfrac{H}{M^2} - \tfrac{3}{2}\right) + \tfrac{3}{4}G^2\left(\ln\tfrac{G}{M^2} - \tfrac{3}{2}\right) - 3T^2\left(\ln\tfrac{T}{M^2} - \tfrac{3}{2}\right)\right] \end{aligned} \qquad (67)$$

with $W \equiv g^2\phi^2/4$, $Z \equiv (g^2+g'^2)\phi^2/4$, $H \equiv -\mu^2+3\lambda\phi^2$, $G \equiv -\mu^2+\lambda\phi^2$ and $T = h^2\phi^2/2$. All of the couplings in this potential run with $t = \ln\phi/M$.

Use of two-loop beta and gamma functions will then give a potential in which all leading and next-to-leading logarithms are summed over. It was shown by Casas, et al.[36] that the resulting minima and masses are relatively independent of the precise choice of M, *as long as* this potential is used (use of earlier potentials was inaccurate due to a sensitive dependence on the choice of scale). It is this potential that will now be used to determine bounds on the top quark and Higgs masses.

The first paper to notice that fermionic one-loop corrections could destabilize the effective potential was by Krive and Linde,[37] working in the context of the linear sigma model. Later, independent investigations by Krasnikov,[38] Hung,[39] Politzer and Wolfram[40] and Anselm[41] all looked at the one-loop, non-renormalization group improved potential of Eqs. 56 and 59., and required that the standard model vacuum be stable for all values of ϕ. The first of these was that of Krasnikov[38] who noted that the bound would be of O(100) GeV, rising to O(1000) GeV if scalar loops were included. The works of Politzer and Wolfram[40] and Anselm[41] gave much more precise numerical results, but ignored scalar loop contributions–thus they obtained upper bounds of $80-90$ GeV on the top quark mass. Hung[39] gave detailed numerical results and did include scalar loops, thus his upper bound ranged from 80 GeV to 400 GeV as the Higgs mass ranged from 0 to 700 GeV.

All of these results are unreliable because the potential used is not valid for large values of ϕ. In these papers, the instability would occur for large values of ϕ, and thus $\ln(\phi/\sigma)$ is large enough that only a renormalization group improved potential is reliable. The first attempt to use an improved potential was the work of Cabibbo, Maiani, Parisi and Petronzio.[42] They included the scale dependence of the Yukawa and gauge couplings, and required that the effective scalar coupling be positive between the weak scale and the unification scale. Although they didn't use the language of effective potentials, this procedure turns out to be very close to that used by considering the full renormalization group improved effective potential. Similar results, using the language of effective potentials, was later obtained by Flores and Sher.[61]

Use of the renormalization-group improved potential will weaken the bounds. The beta function for the top quark Yukawa coupling is negative, and thus the coupling falls as the scale increases. Thus, the effects of fermionic corrections will decrease at larger scales. Compared with the bounds that one would obtain by ignoring the renormalization-group improvement, the decrease in the Yukawa coupling at large scales will weaken the upper bounds. This effect is not small; the Yukawa coupling for a quark will fall by roughly a factor of three between the weak and unification scales. Note that for additional leptons, the Yukawa coupling does not fall significantly, thus the bounds obtained by the non-renormalization-improved potential will not be greatly changed.

The first attempt to bound fermion masses using the full renormalization group improved effective potential (earlier works, for example, never mentioned anomalous dimensions) was the 1985 work of Duncan et al.[44] Their results, however, used tree level values for the Higgs and top masses, in terms of the scalar self-coupling and the "$\overline{MS}$ Yukawa coupling", and found a bound which, to within a couple of GeV, can be fit by the line

$$M_{top} \; < \; 80 \text{ GeV} + 0.54 M_{Higgs} \tag{68}$$

As we will see shortly, however, corrections to the top quark mass can be sizeable, as much as 10 GeV. A much more detailed analysis, using two loop beta functions and one-loop corrections to the Higgs and top quark masses (defined as the poles of the propagator), was carried out in 1989 by Lindner, Zaglauer and Sher,[45] and followed up with more precise inputs in 1993 by Sher.[46] In all of these papers, the allowed region in the Higgs-top mass plane was given–the allowed region was always an upper bound on the top mass for a given Higgs mass, or a lower bound on the Higgs mass for a given top mass. The allowed region depended on the cutoff Λ at which the instability occurs. For example, if the instability occurs for values of ϕ above 10^{10} GeV, then one concludes that the standard model vacuum is unstable IFF the standard model is valid up to 10^{10} GeV (should the lifetime of the metastable vacuum be less than the age of the Universe, one would conclude that the standard model can not be valid up to 10^{10} GeV). Thus, all of the bounds depend on the value of Λ.

In the above papers, the effective potential used was the renormalization group improved tree-level potential, Eq.66. As discussed in the previous section, this would be as precise as the precision of the beta functions and anomalous dimensions (two-loop were used) if the only logarithms were of the form $\ln(\frac{\phi^2}{M^2})$; the resulting potential is exact in terms of the beta and gamma functions. However, when scalar loops are included, terms of the form $\ln(\frac{\mu^2+\lambda\phi^2}{M^2})$ arise, and these terms are not summed over. In the earlier papers, it was argued that when ϕ is large, the scalar terms are effectively of the form $\ln(\frac{\phi^2}{M^2})$, and thus the difference is irrelevant. But, in determining the Higgs boson mass in terms of the potential, the structure of the potential at the electroweak scale is relevant, and thus the difference in the form of the scalar loops is relevant. It turns out that this difference is especially crucial when the value of Λ is relatively small ($1 - 10$ TeV), and less important when Λ is large (10^{15-19} GeV), thus the results of the above papers are valid in the large Λ case.

To include the proper form of the scalar loops, one must use the form of Ford, et al.,[34] discussed in the last section. This analysis was carried out very recently by Casas, Espinosa and Quiros[47] and by Espinosa and Quiros.[48,a] A very pedagogical review of the analysis can be found in Espinosa's Summer School Lectures.[51] We now briefly review that analysis and present their results.

Consider the tree level renormalization group improved potential, Eq. 66. At large values of ϕ, the quadratic term becomes negligible, and the question of whether the standard model vacuum is stable is essentially identical to the question of whether $\lambda(t)$ ever goes negative. If $\lambda(t)$ goes negative at some scale Λ, then the instability will occur at that scale.

Casas, et al.[36,47] analyzed the question using the full one-loop renormalization group improved potential, with two-loop beta and gamma functions, of Eq. 67. They showed that the instabilty sets in when $\tilde{\lambda}$ becomes negative, where $\tilde{\lambda}$ is slightly different from λ:

$$\begin{aligned}\tilde{\lambda} = \lambda - \tfrac{1}{16\pi^2}\Big[3h^4\left(\ln\tfrac{h^2}{2} - 1\right) - \tfrac{3}{8}g^4\left(\ln\tfrac{g^2}{4} - \tfrac{1}{3}\right) \\ -\tfrac{3}{16}(g^2+g'^2)^2\left(\ln\tfrac{g^2+g'^2}{4} - \tfrac{1}{3}\right)\Big]\end{aligned} \tag{69}$$

All that remains is to relate the parameters in the potential to the physical masses of the Higgs boson and of the top quark.

It is not a trivial matter to extract the Higgs and top quark masses from the values of $h(t)$ and $\lambda(t)$ used in the potential. One can write

$$\begin{aligned} m_{top}(\mu) &= m_{top}^{pole}(1+\delta_{top}(\mu)) = \tfrac{1}{\sqrt{2\sqrt{2}G_F}}h(\mu) \\ m_H(\mu) &= m_H^{pole}(1+\delta_H(\mu)) = \sqrt{\tfrac{\sqrt{2}}{G_F}\lambda(\mu)} \end{aligned} \tag{70}$$

where the pole masses are the physical masses of the top and Higgs, and $\delta_{top}(\mu)$ is the radiative corrections to the $\overline{MS}$ top quark mass. Note that the physical Higgs mass is NOT simply the second derivative of the effective potential, since the potential is defined at zero external momentum and the pole of the propagator is on-shell; $\delta_H(\mu)$ accounts for the correction.

The correction $\delta_{top}(\mu)$ receives contributions from QCD, QED and weak radiative effects, with the QCD corrections being the largest. The QCD

[a]See Alterelli and Isidori[50] for a similar and independent analysis.

corrections have been calculated to $O(g_3^2)$[52] and to $O(g_3^4)$,[53,54] the other corrections have been determined.[55,56,57] The correction $\delta_H(\mu)$ has also been found.[36,58] The detailed expressions for these quantities, which correct several typographical errors in the published works, are summarized in an extensive review article by Schrempp and Wimmer.[59] The largest correction is to the top quark mass; the leading order term is $\frac{4}{3}\frac{\alpha_3}{\pi}$, which is 5%, or almost 10 GeV.

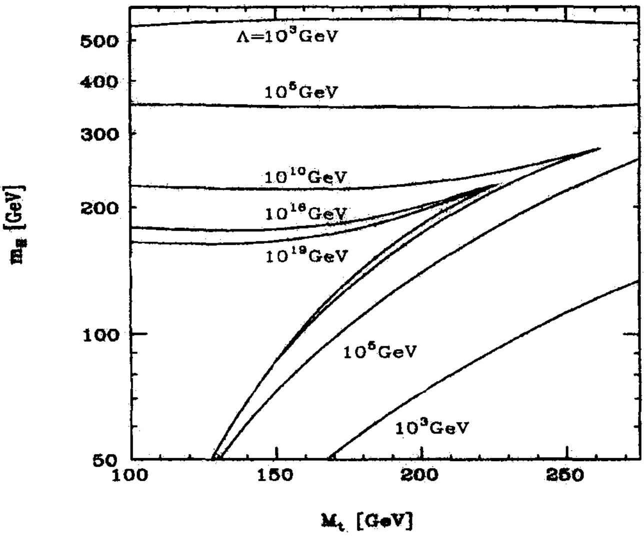


Figure 3. Perturbativity and stability bounds on the SM Higgs boson. Λ denotes the energy scale where the particles become strongly interacting.

All of these corrections were included by Casas, Espinosa and Quiros.[47,48,49] If one requires stability of the vacuum up to a scale Λ, then there is an excluded region in the Higgs mass-top mass plane. The result, for various values of Λ, is given in Figure 3. This figure, in addition, also includes the region excluded by the requirement that the scalar and Yukawa coupling don't become nonperturbative by the scale Λ; these bounds will be discussed in the next section. The lower part of each curve is the vacuum stability bound; the upper part is the perturbation theory bound. The excluded region is outside the solid lines. Thus, for a top quark mass of 170 GeV, we see that a discovery of a Higgs boson with a mass of 90 GeV would imply that the standard model vacuum is unstable at a scale of 10^5 GeV, i.e. if we live in a stable vacuum, the standard model must break down at a scale below 10^5 GeV. The curves in Figure 3 are approximately straight lines in the vicinity of $M_{top} \sim 170$ GeV, thus the top mass dependence can be given analytically.[49] For $\Lambda = 10^{19}$ GeV, we must have

$$M_H(GeV) \quad > \quad 133 + 1.92(M_{top}(GeV) - 175) - 4.28\frac{\alpha_3(M_Z) - .12}{0.006} \qquad (71)$$

and for $\Lambda = 1$ TeV,

$$M_H(GeV) \quad > \quad 52 + 0.64(M_{top}(GeV) - 175) - 0.5\frac{\alpha_3(M_Z) - .12}{0.006} \qquad (72)$$

It is estimated[47,49] that the error in the result, primarily due to the two-loop correction in the top quark pole mass and the effective potential, is less than 5 GeV. In Figure 4, the stability and perturbation theory bounds are given explicitly as a function of Λ for $M_{top} = 175$ GeV.

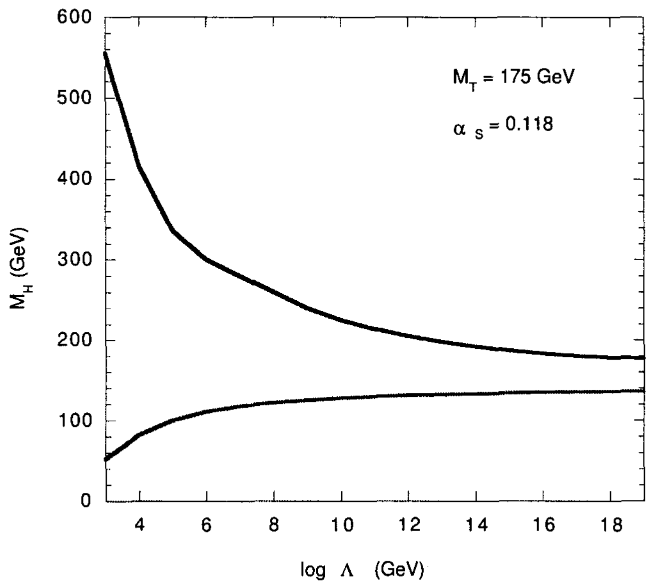


Figure 4. Perturbativity and stability bounds on the SM Higgs boson as a function of Λ for $M_{top} = 175$ GeV.

Of course, it is not formally necessary that we live in a stable vacuum. Should another deeper vacuum exist, it is only necessary that the Universe goes into our metastable vacuum and then stay there for at least 10 billion years. A detailed discussion of the finite temperature effective potential and tunnelling probabilities is beyond the scope of this review; the reader is referred to Casas[47] and Quiros[49] for the details, as well as a comprehensive list of references. In short, the bound in the above paragraph for $\Lambda = 10^{19}$ GeV weakens by 8 GeV, and for $\Lambda = 1$ TeV, weakens by about 25 GeV. In all cases, the bound obtained by requiring that our vacuum have a lifetime in excess of 10 billion years is weaker than the bound obtained by requiring that the Universe arrive in our metastable vacuum.

3.4. *Extensions of the Higgs Sector and the MSSM*

The tree-level potential of the Standard Model is very simple, however the simple addition of a second Higgs doublet complicates the potential greatly. The most general potential with two doublets is given by

$$\begin{aligned}
V &= m_1^2 H_1^\dagger H_1 + m_2^2 H_2^\dagger H_2 + m_3^2(H_1 H_2 + h.c.) \\
&+ \frac{1}{2}\lambda_1 |H_1^\dagger H_1|^2 + \frac{1}{2}\lambda_2 |H_2^\dagger H_2|^2 + \lambda_3 (H_1^\dagger H_1)(H_2^\dagger H_2) + \lambda_4 H_1 H_2 H_1^\dagger H_2^\dagger \\
&+ [\frac{1}{2}\lambda_5 (H_1 H_2)^2 + \lambda_6 (H_1^\dagger H_1)(H_1 H_2) + \lambda_7 (H_2^\dagger H_2)(H_1 H_2) + \mathrm{h.c.}] \qquad (73)
\end{aligned}$$

where the $SU(2)$ indices for "$H_1 H_2$" are $\epsilon_{ij} H_1^i H_2^j$. The details of this potential and the effective potential and bounds on Higgs masses resulting from it are given by Sher.[31] Note that with two complex doublets, we have eight fields. Three are eaten by the $W^\pm$ and Z, leaving five physical scalars: a charged pair and three neutral scalars (if CP is conserved, one of the three neutral scalars is a pseudoscalar).

In supersymmetry, the quartic terms of a potential are restricted. This is because quartic terms are dimensionless, and with softly or spontaneously broken supersymmetry, they must reflect the underlying supersymmetry. Thus, they can only come from F-terms or D-terms. Since we have only Higgs doublets, there are no quartic F-terms, and the D-terms are given in terms of gauge couplings. It can easily be shown that the couplings, in supersymmetry, are given by

$$\begin{aligned}
\lambda_1 = \lambda_2 = \frac{1}{4}(g^2 + g'^2)\,; \qquad \lambda_3 = \frac{1}{4}(g^2 - g'^2)\,; \\
\lambda_4 = -\frac{1}{4}g^2\,; \qquad \lambda_5 = \lambda_6 = \lambda_7 = 0 \qquad (74)
\end{aligned}$$

so the potential becomes

$$V = m_1^2 H_1^\dagger H_1 + m_2^2 H_2^\dagger H_2 + m_3^2(H_1 H_2 + h.c.)$$

$$+\frac{1}{8}g^2(H_1^\dagger\vec{\sigma}H_1 + H_2^\dagger\vec{\sigma}H_2)^2 + \frac{1}{8}g'^2(H_2^\dagger H_2 - H_1^\dagger H_1)^2 \tag{75}$$

This potential has only three unknown parameters. Yet there are five masses (the W mass, and the masses of the charged scalar, $H^\pm$ and three neutral scalars, h, H and χ). As a result, there are predictions:

$$\begin{aligned} M_{H^\pm}^2 &= M_W^2 + M_\chi^2 \\ M_{h,H}^2 &= \frac{1}{2}\left[M_\chi^2 + M_Z^2 \mp \sqrt{(M_\chi^2 + M_Z^2)^2 - 4M_\chi^2 M_Z^2 \cos^2 2\beta}\right] \end{aligned} \tag{76}$$

where $\tan\beta$ is the ratio of the two vacuum expectation values. The second of these gives the critical relationship[60,61]

$$M_h < M_Z \cos 2\beta \tag{77}$$

Since the lower bound on the Higgs mass is already greater than M_Z, it would appear that the MSSM is ruled out. However, radiative corrections are crucial.[62–68] Although dependent on many parameters of the MSSM, the results can be summarized as follows. One defines X_t to be $A_t - \mu/\tan\beta$, where A_t is the top quark Yukawa term A-parameter and μ is the coefficient of the $H_1 H_2$ term in the superpotential. The higher order correction to the mass-squared is then

$$\Delta M_h^2 = \frac{3}{4\pi^2 v^2} M_{top}^4 \log\frac{M_{SUSY}^2}{M_{top}^2} + \frac{3}{8\pi^2}\frac{M_{top}^4}{v^2}\tilde{A}^2(2 - \tilde{A}^2/6) \tag{78}$$

where $\tilde{A}$ is X_t/M_{SUSY}. For a top quark mass of 175 GeV, and $M_{SUSY} < 1$ TeV, the upper bound increases to 125 GeV, which is still phenomenologically acceptable.

Currently, LEP[69] has cited a lower bound of 114 GeV on the Higgs mass (and weak evidence for a signal just above this). However, that bound applies only to the Standard Model, not the MSSM. Nonetheless, the couplings in the two models are similar, and so the MSSM bound will not be too much lower. There is very little room left for the MSSM. If the Higgs mass, even an MSSM Higgs mass, is below 125 GeV, it will be discovered during Run III at the Tevatron.[70] The moment of truth is arriving for the MSSM. If the Tevatron does not discover the Higgs within three or four years, the MSSM will be dead.

One wouldn't necessarily have to give up on supersymmetry. If a singlet is added to the MSSM, then there is an extra parameter in the superpotential. This widens the allowed parameter space. If one insists that the theory remain perturbative up to a large scale, this extra parameter is limited, and the upper bound can't increase much more than 150 GeV.[71,72,73]

Depending on the integrated luminosity at the Tevatron, this might evade detection until the onset of the LHC.

4. Baryogenesis

Why are we here? Since the discovery of antimatter in the 1930's, the fact that our universe appears to be made almost entirely of matter has been a great puzzle. Without baryon number violation, it is impossible for a matter-antimatter symmetric universe to develop an asymmetry, and prior to the discovery of GUTs, the observed asymmetry had to be put in *ab initio*. The size of the asymmetry is characterized by the dimensionless quantity n_B/s, where n_B is the baryon number density and s is the entropy density of the universe (which primarily comes from the cosmic microwave background radiation). Since both n_B and s vary as T^3 as the universe cools, this ratio is constant. Its value today is observed to be

$$\frac{n_B}{s} = 4 \times 10^{-9}\Omega_B h^2 \sim 7 \times 10^{-11} \tag{79}$$

Since at temperatures well above 1 GeV, quarks, antiquarks and photons are in thermal equilibrium, $n_q = n_{\bar{q}} = n_\gamma$ (neglecting some factors of order 1), and n_γ of a photon gas is related to the entropy, this can be rewritten as

$$\frac{n_q - n_{\bar{q}}}{n_q} \sim 3 \times 10^{-8} \tag{80}$$

so for every 30 million antiquarks in the early universe, there must have been 30 million and one quarks. As the universe cooled, the 30 million antiquarks annihilated with 30 million quarks, producing photons, and leaving the residual small asymmetry. This small ratio had to be put in by hand.

In a seminal paper, long before the discovery of grand unification, Sakharov[74] showed that three conditions must be met for a generation of a baryon asymmetry (from an initially symmetric universe).

•1) Baryon number must be violated.

•2) C and CP must be violated. If this condition is not met, the B-violating interactions will produce baryons and antibaryons at the same rate (baryon number is odd under C and CP).

•3) The universe must go out of equilibrium. In equilibrium, the density of baryons and antibaryons are equal, since their masses are equal, i.e. if one has $A \leftrightarrow B$, then whatever gets made gets unmade.

When GUTs appeared in the mid-to-late 70's, a natural source of baryon number violation appeared. C and CP violation could also occur, and the mass of the X bosons caused the universe to go out of equilibrium. With the Sakharov conditions satisfied, a baryon asymmetry could be generated. Many years were spent analyzing various models for generating an asymmetry, for combining the need for an asymmetry with inflation, etc. Then in the late 80's, it was recognized that the **electroweak** interaction, at high temperature, also violates baryon number, and could easily wash out any asymmetry produced in grand unified models. Interest began to focus on electroweak baryogenesis.

We will first discuss baryogenesis in the context of grand unified theories, and then turn to electroweak baryogenesis. There are a number of excellent reviews on the subject (far more detailed than this single lecture). For a general discussion, there is the book of Kolb and Turner,[75] which gives an excellent summary up to the beginning of the 90's, just after the important of electroweak baryogenesis was discovered. More recently, an excellent review by Riotto and Trodden[76] includes most of the latest results, as well as a very extensive list of references. Other reviews have also appeared.[77–81]

4.1. *GUT Baryogenesis*

Since GUTs have quarks and leptons in the same irreducible representations, they provide a natural source of baryon number generation. To see how this works, consider a simple case of a particle X with two decays: $X \to qq$ and $X \to \bar{q}\bar{l}$. The CPT theorem requires that the decay rates of X and $\bar{X}$ are equal, but C and CP violation allow

$$\frac{X \to qq}{X \to \bar{q}\bar{l}} \neq \frac{\bar{X} \to \bar{q}\bar{q}}{\bar{X} \to ql} \tag{81}$$

Define $r \equiv BR(X \to qq)$ and $\bar{r} \equiv BR(\bar{X} \to \bar{q}\bar{q})$, then CP violation allows $r \neq \bar{r}$. Suppose you have a volume with equal numbers of X and $\bar{X}$. Then the baryon number from a decay of X is $\frac{2}{3}r - \frac{1}{3}(1-r)$ and that from $\bar{X}$ is $-\frac{2}{3}\bar{r} + \frac{1}{3}(1-\bar{r})$. The net baryon asymmetry is the sum of these, or $\epsilon = r - \bar{r}$.

At very high temperatures, $T >> M_X$, there is complete thermal equilibrium, so $n_X = n_{\bar{X}} = n_\gamma$. As the temperature drops below M_X, however, X and $\bar{X}$ decay without backreactions. Since each decay produces a baryon number ϵ, the resulting baryon density is $n_B = \epsilon n_X \sim \epsilon n_\gamma$. The entropy density is related to n_γ by $s = g_* n_\gamma$, where g_* depends on the number of states in the theory and is roughly a couple of hundred. Thus the final n_B/s is given by ϵ/g_*, and so an ϵ of around 10^{-8} is needed to generate a sufficient baryon asymmetry.

But one still will not be able to produce a baryon asymmetry without a departure from equilibrium. This is provided by the expansion of the universe. Once the processes that maintain thermal equilibrium, Γ, fall below the expansion rate, H, they will not be able to keep up. H is the Hubble constant, given by $\sqrt{g_*}T^2/M_{Pl}$, where M_{Pl} is the Planck mass. Define

$$K \equiv \left(\frac{\Gamma_{decay}}{2H}\right)_{T=M_X} = \frac{\alpha M_{Pl}}{3.3\sqrt{g_*}M_X} \tag{82}$$

where α is a typical coupling constant in the decay. If $K << 1$, then $\Gamma_{decay} << H$, and the number density of X and $\bar{X}$ do not decrease, so they become overabundant. They eventually decay, producing the asymmetry of the above paragraph. The condition $K < 1$ implies that $M_X > \alpha(g_*)^{-1/2}M_{Pl}$, or

$$M_X > \left(\frac{\alpha}{10^{-2}}\right) 10^{16} \text{ GeV} \tag{83}$$

If X is a Higgs boson, α is generally small, so this is generally satisfied. Even if it is a gauge boson, it can also be satified.

Details, of course, require solving the Boltzman equations. In minimal $SU(5)$, where the only CP-violation comes from the CKM sector, the largest value of n_B/s that can be obtained is 10^{-19}, which is far too small. For other models, even the two-Higgs $SU(5)$ model, there are other sources of CP violation which can be much larger. It is important to note that $SU(5$ has an accidental $B-L$ symmetry, and thus a non-zero $B-L$ cannot be generated in $SU(5)$. This is critical, since non-perturbative processes, as we will see below, can wash out a baryon asymmetry, but will not change $B-L$. Thus, models such as $SO(10)$, which can produce a non-zero $B-L$, can survive this washout.

So we see that GUTs can achieve the desired asymmetry. However, the advent of inflationary cosmology causes some problems. In "conventional" inflation, the universe enters a supercooled phase transition, in which it expands by an enormous factor. This will erase any pre-existing asymmetry. The universe will then reheat as the transition ends, but the reheating temperature is often well below M_X, and thus no X bosons get produced, and thus no baryon asymmetry is produced. There are a number of ways out of this problem, the most attractive being the "pre-heating" scenario described in the review article, with many references, of Riotto and Trodden.[76]

The attractiveness of GUT baryogenesis faded somewhat with the discovery that anomalous electroweak interactions will erase any pre-existing B asymmetry. We now consider these processes.

4.2. *Electroweak Baryon Number Violation*

Consider an $SU(N)$ gauge theory with a single massless fermion

$$\mathcal{L} = i\bar{Q}\gamma_\mu D^\mu Q - \frac{1}{4g^2} F_{\mu\nu}F^{\mu\nu} \tag{84}$$

The theory has two symmetries: (1) $Q \to e^{i\alpha}Q, j^\mu = \bar{Q}\gamma^\mu Q$, and (2) $Q \to e^{i\alpha\gamma_5}Q, j^{\mu 5} = \bar{Q}\gamma^\mu\gamma_5 Q$. At the quantum level, the axial current is *not* conserved, due to a triangle graph with the current at one vertex and vector bosons at the other two. This triangle graph gives $\frac{1}{16\pi^2}\epsilon_{\mu\nu\rho\sigma}p_1^\mu p_2^\nu \epsilon_1^\rho \epsilon_2^\sigma$, and thus the divergence in the axial current is

$$\partial_\mu j^{\mu 5} = \frac{1}{16\pi^2} F^a_{\mu\nu}\tilde{F}^{a\mu\nu} \tag{85}$$

where $\tilde{F}_{\mu\nu} \equiv \frac{1}{2}\epsilon_{\alpha\beta\mu\nu}F^{\alpha\beta}$

Note, however that $\frac{1}{2}F^a_{\mu\nu}\tilde{F}^{a\mu\nu} = \partial_\mu K^\mu$ where $K^\mu = \epsilon^{\mu\nu\rho\sigma}(F_{\nu\rho}A_\sigma + \frac{2}{3}A_\nu A_\rho A_\sigma)$, and thus it is a total derivative. One sees that the quantity $j_5^\mu - \frac{g^2}{8\pi^2}K^\mu$ is conserved. For Abelian theories, this redefinition has no physical effect.

In non-Abelian theories, however, the total derivative term can have an effect, due to the non-trivial vacuum structure. In these theories, the vacuum state consists of many configurations of gauge fields which can't be continuously transformed into each other. Each of these configurations is characterized by a topological charge, called a Chern-Simons number

$$n = \frac{ig^3}{24\pi^2}\int d^3x Tr(\epsilon_{ijk}A^i A^j A^k) \tag{86}$$

The integrand is the Jacobian of the transformation from S_3 to the hypersphere in group space. The vacuum state

$$|\theta\rangle \equiv \sum_n e^{-in\theta}|n\rangle \tag{87}$$

is gauge invariant, and different θ's give different theories.

One can now see how the existence of these different θ-vacua can cause a physical effect from the triangle graph. Choosing a gauge where $A^o = 0$, one cnan see that the Chern-Simons number is $n = \frac{g^2}{16\pi^2}\int d^3x K^o$. Now, go back to the divergence of the axial current. It is $\partial_\mu j^{\mu 5} = \frac{g^2}{16\pi^2}\partial_\mu K^\mu$. Integrating the o-component by parts gives $\Delta Q_5 = \Delta\int d^3x K^o = n(\infty) - n(-\infty)$. So if a process causes a system to transform from one n to another, the axial charge will change. A configuration of gauge fields which changes n by one unit is called a sphaleron.

An explicit example of a sphaleron can be found in an $SU(2)$ model with a Higgs doublet in the $A_o^a = 0$ gauge. The solution is

$$\begin{aligned} U(\vec{r} &= \frac{1}{r}\begin{pmatrix} z & x+iy \\ x-iy & z \end{pmatrix} \\ \phi(\vec{r}) &= h(r)U(\vec{r})\begin{pmatrix} 0 \\ \mu \end{pmatrix} \\ \sigma^a A_i^a &= -\frac{2i}{g}f(r)\frac{\partial}{\partial x^i}U(\vec{r})(U(\vec{r})^{-1} \end{aligned} \qquad (88)$$

where $f(r)$ and $h(r)$ are calculable functions.

In the standard model, the baryon and lepton currents are also chiral: $\partial_\mu j_B^\mu = \frac{3}{16\pi^2}F^W\tilde{F}^W$ and $\partial_\mu j_L^\mu \frac{3}{16\pi^2}F^W\tilde{F}^W$. Note that $B-L$ is anomaly free, but B and L individually are not. Sphalerons can cause violation of both, with $\Delta B = \Delta L = 3$. The sphaleron is shown schematically in Figure 5. The energy of the sphaleron is $M_W/\alpha_W \sim 10$ TeV, thus the barrier size in the figure is very high. A transition[82] from one vacuum to another must climb this barrier, with a typical transition rate of $e^{-4\pi/\alpha_W} \sim e^{-400}$ which is very weak (the process would be $p+p \rightarrow \bar{p} + 3e^+$). As a result, this was thought to be phenomenologically irrelevant.

However, Kuzmin, Rubakov and Shaposhnikov[83] showed that at high temperatures (in excess of 1 TeV), thermal fluctuations remove the suppression factor. There is some controversy over the precise rate (see Riotto and Trodden[76] for an extensive set of references), but all agree that a pre-existing baryon asymmetry will be removed by sphaleron effects.

So no net $B+L$ can remain from GUT baryogenesis. There are three possible solutions to this problem:

1. B-L is nonzero, so even if B+L=0, one can have a nonzero B. This fails for $SU(5)$, but is allowed in $SO(10)$.

2. B+L is generated above the electroweak scale and not completely neutralized. There are no good models (to my knowledge) which can implement this.

3. Electroweak baryogenesis.

We will now consider the latter case. Here, one can't rely on the expansion of the universe to cause certain fields to drop out of thermal equilibrium, since the expansion rate is too slow at the 100 GeV scale. One needs an out-of-equilibrium process, i.e. an electroweak phase transition.

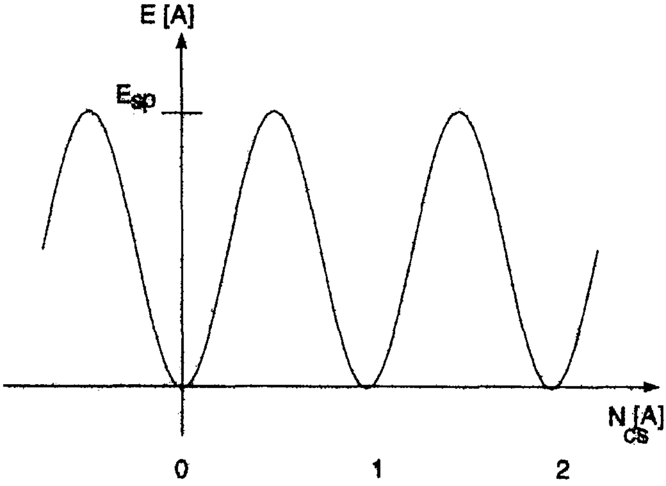


Figure 5. The energy of the gauge configurations as a function of the Chern-Simons number. The adjacent vacua differ in baryon and lepton number by $\Delta B = \Delta L = 3$.

4.3. *Electroweak Baryogenesis*

To discuss the electroweak theory at high temperature, we must first consider the Higgs potential at high temperature. At nonzero temperature, the Higgs potential is (writing $\phi^\dagger\phi$ as ϕ^2)

$$\begin{aligned} V &= \frac{1}{2}\mu^2\phi^2 + \frac{1}{4}\lambda\phi^4 \\ &+ \sum_i \frac{n_i(-1)^F}{64\pi^2}\mathcal{M}_i^4(\phi)[\ln\left(\frac{\mathcal{M}_i^2(\phi)}{M^2}\right) - c_i] \\ &+ \sum_i \frac{(-1)^F n_i T^4}{2\pi^2}\int_0^\infty dx\, x^2 \ln(1 - (-1)^F e^{-(x^2+\frac{\mathcal{M}_i^2(\phi)}{T^2})^{1/2}}) \end{aligned} \tag{89}$$

where n_i is the number of degrees of freedom of particle i, $\mathcal{M}$ is the mass of the particle i in the shifted vacuum and F is the fermion number. There are many ways to derive this:

1) It is easy to show that field theory at finite temperature is the same as at zero temperature but periodic in Euclidean time. Thus, the integral over k_o in the derivation of the one-loop potential becomes a sum giving the extra term.

2) The finite-temperature propagator of a field is the usual zero-temperature

propagator plus a piece of the form (for a boson) $2\pi\delta(p^2-m^2)/(e^{E/kT}-1)$, which says that if the particle is on shell, it must obey the statistics of the heat gas.
3) Look up the free energy of an ideal Bose or Fermi gas in Landau and Lifshitz statistical mechanics text. It is the same.

All of these give the above result. It becomes much more transparent when one looks at the high temperature limit. There is a ϕ–independent T^4 term, and (dropping that) the temperature dependent part becomes

$$V_T^{bosons} = \sum_i n_i\left[\frac{1}{24}\mathcal{M}^2T^2 - \frac{\mathcal{M}^3T}{12\pi} + ...\right]$$
$$V_T^{fermions} = \sum_i n_i\left[\frac{1}{48}\mathcal{M}^2T^2 + ...\right] \qquad (90)$$

So, in general

$$V(\phi,T) = D(T^2-T_o^2)\phi^2 - ET\phi^3 + \frac{\lambda}{4}\phi^4... \qquad (91)$$

At high temperature, the coefficient of the ϕ^2 term is positive, so there is only a minimum at the origin and the symmetry is restored. As the temperature drops, a new asymmetric minimum occurs. At a temperature T_c, both the symmetric and asymmetric minima have the same value, and as the temperature drops below T_c, the asymmetric minimum is lower, and the symmetric minimum disappears. A first-order phase transition will occur near this temperature, T_c. It is first order because the vacuum expectation value of the Higgs field will change discontinuously (due to thermal or quantum fluctuations) from the symmetric to the asymmetric vacuum. One can show that at T_c, $\phi(T_c)/T_c = E/\lambda$.

The existence of a first-order phase transition provides the necessary departure from equilibrium. It is important for the vev of the Higgs field to be sufficiently large at the time of the transition. If it is not, the universe will be close enough to equilibrium that any baryon number generated will be destroyed immediately. The necessary condition is approximately that $v(T_c) \geq T_c$. This condition has been demonstrated both perturbatively and non-perturbatively (on the lattice); a very long and comprehensive list of references is given in the recent article by Quiros.[84]

In the Standard Model, the value of E is $\frac{2}{3}\frac{2M_W^3+M_Z^3}{\sqrt{2}\pi v^3}$, and that leads to a bound of $M_h < 40$ GeV. This is in conflict with experiment, and so the Standard Model can not generate a sufficient baryon number.

However, the MSSM can provide sufficient baryogenesis. There are a number of CP violating phases in the model, and over a dozen references (all within the past three years) can be found in the article of Quiros.[84] The general conclusion is that the MSSM can work, provided

1. There are light right-handed scalar top quarks (less than 160 GeV).
2. There is a heavy left-handed scalar top quark (greater than 1 TeV).
3. The A parameter for the top Yukawa coupling is not too large
4. $\tan\beta$ is greater than 5.
5. Some CP-violating phases are large (this may conflict with electric dipole measurements).

These conditions might seem to point to a small region of parameter space, but recall that this is the MSSM. Somewhat non-minimal models will have substantially more freedom.

When there is a first order phase transition, one generally has a bubble wall whose thickness depends on the order of the transition. In these models, the baryon violation occurs in the symmetric phase, outside the bubble, the CP violation occurs in the asymmetric phase, inside the bubble, and the wall itself is out of equilibrium. One can see that the details of these calculations are quite complicated, and the reader is referred to the review of Riotto and Trodden[76] for a detailed discussion and a set of references.

4.4. *Alternatives*

With the recent discovery of neutrino masses, greater emphasis has been placed on the possibility of generating baryon number through leptogenesis. Most models that explain the smallness of the neutrino masses have a heavy Majorana neutrino, which then generates a small neutrino mass through the see-saw mechanism.[85] Majorana neutrinos automatically violate lepton number, and so a lepton asymmetry is generated. The idea is that it is generated well before the electroweak transition, and then sphaleron interactions, which conserve B-L but violate B+L, distribute that lepton asymmetry into a baryon asymmetry.

The heavy Majorana neutrino, N has interactions

$$\mathcal{L} = \bar{l}_L \Phi h_\nu N_L^c + M \bar{N}_L^c N_L^c \tag{92}$$

where Φ is the Higgs doublet. This gives an off-diagonal term of the order of the weak scale in the neutrino mass-matrix. M is assumed to be very large, and thus small masses of order M_W^2/M are generated for the light neutrinos. The Majorana neutrino can decay into a Higgs boson and a light neutrino, either through $N \rightarrow \bar{\Phi}\nu$ or $N \rightarrow \Phi\bar{\nu}$. These decays obviously violate lepton number. The interference between the tree level decay and the absorptive part of the one-loop vertex generates a baryon asymmetry of the right order of magnitude.[86,87,88] A comprehensive review can be found in the article by Pilaftsis.[89]

Another method of generating a baryon asymmetry is the Affleck-Dine scenario.[90] The idea is to consider a potential with a flat direction for some component of a scalar field χ. Such flat directions appear often in supersymmetric and superstring theories. The flat direction will become slightly curved due to supersymmetry breaking and nonrenormalizable operators. Initially, one would expect the χ field to start at some nonzero value of $\langle\chi\rangle$. We assume that the χ field has baryon or lepton number. Once the Hubble rate is the size of the curvature asymmetry will be generated as the χ field slowly rolls down to its current minimum and begins oscillating about the minimum. The χ field then decays, and the baryon number in its condensate gets transmitted to the decay products.

At first, this might seem far-fetched, since it requires scalar fields with baryon or lepton number. However, it is not only natural, but expected, in supersymmetric theories. Scalar quarks and leptons are excellent candidates for these fields. Even the MSSM has flat directions in the potential (which are broken by SUSY-breaking terms). A detailed review of the original Affleck-Dine scenario can be found in the review of Dolgov,[91] and the implementation in the MSSM can found in the papers of Dine, Randall and Thomas.[92] A nice discussion of recent work is in the review of Riotto and Trodden.[76]

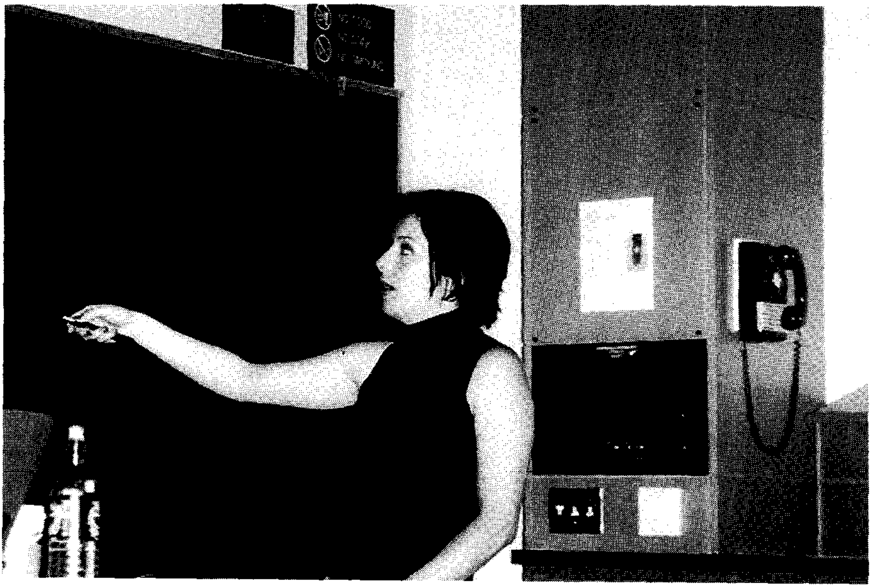
References

1. S.L. Glashow, Nucl. Phys. **B22**, 529 (1961); S. Weinberg, Phys. Rev. Lett. **19**, 1264 (1967); A. Salam, Proc. of Nobel Symposium, 1968.
2. J. Pati and A. Salam, Phys. Rev. **D10**, 275 (1974).
3. H. Georgi and S.L. Glashow, Phys. Rev. Lett. **32**, 438 (1974).
4. E. Witten, Nucl. Phys. **B188**, 513 (1981).
5. P. Langacker, Physics Reports **72**, 185 (1981).
6. G.G. Ross, "Grand Unified Theories", Benjamin/Cummings, 1984.
7. R.N. Mohapatra, "Unification and Supersymmetry", Springer-Verlag, 1992.
8. E. Gildener and S. Weinberg, Phys. Rev. **D13**, 3333 (1976).
9. H. Georgi, H.R. Quinn and S. Weinberg, Phys. Rev. Lett. **33**, 451 (1974).
10. A. Buras, J. Ellis, M.K. Gaillard and D.V. Nanopoulos, Nucl. Phys. **B135**, 66 (1978).
11. H. Georgi and C. Jarlskog, Phys. Lett. **B86**, 297 (1979).
12. J. Ellis and M.K. Gaillard, Phys. Lett. **B88**, 315 (1979).
13. M. Carena, M. Olechowski, S. Pokorski and C.E.M. Wagne, Nucl. Phys. **426**, 269 (1994).

14. E. Witten, Phys. Lett. **B105**, 267 (1981).
15. D.V. Nanopoulos and K. Tamvakis, Phys. Lett. **B113**, 151 (1981)
16. L. Ibanez and G. Ross, Phys. Lett. **B110**, 715 (1981).
17. D. Nemeschansky, Nucl. Phys. **B234**, 379 (1984).
18. S. Dimopoulos and H. Georgi, Nucl. Phys. **B150**, 193 (1981).
19. M. Sakai, Z. Phys. **C11**, 153 (1981).
20. E. Witten, Nucl. Phys. **B188**, 573 (1981).
21. N. Maru, Phys. Lett. **B522**, 117 (2001).
22. M. Kakizaki and M. Yamaguchi, hep-ph/0104103.
23. Y. Kawamura, Prog. Theor. Phys. **105**, 999 (2001).
24. A.E. Faraggi, Phys. Lett. **B520**, 337 (2001).
25. R. Barbieri, G. Dvali, A. Strumia, Z. Berezhiani and L. Hall, Nucl. Phys. **B432**, 49 (1994).
26. Z. Berezhiani, Phys. Lett. **B355**, 481 (1995).
27. J. Hewett and T. Rizzo, Phys. Reports **183**, 193 (1989).
28. H.E. Haber and G.L. Kane, Phys. Reports **117**, 75 (1985).
29. H.P. Nilles, Phys. Reports **110**, 1 (1984).
30. R.S. Chivukula, Lectures on Technicolor and Compositeness, TASI 2000, hep-ph/0011264 (2000).
31. M. Sher, Phys. Reports **179**, 273 (1989).
32. S. Coleman and E. Weinberg, Phys. Rev. **D7**, 1888 (1974).
33. M. Bando, T. Kugo, N. Maekawa and H. Nakano, Phys. Lett. **B301**, 83 (1993) ; Prog. Theor. Phys. **90**, 405 (1993).
34. C. Ford, D.R.T. Jones, P.W. Stephenson and M.B. Einhorn,Nucl. Phys. **B395**, 17 (1993).
35. B. Kastening,Phys. Lett. **B283**, 287 (1992) .
36. J.A. Casas, J.R. Espinosa, M. Quiros and A. Riotto, Nucl. Phys. **B436**, 3 (1995); (E)Nucl. Phys. **B439**, 466 (1995).
37. I.V. Krive and A.D. Linde, Nucl. Phys. **B117**, 265 (1976).
38. N.V. Krasnikov, Yad. Fiz **28**, 549 (1978).
39. P.Q. Hung, Phys. Rev. Lett. **42**, 873 (1979) .
40. H.D. Politzer and S. Wolfram, Phys. Lett. **B82**, 242 (1979) ; Phys. Lett. **B83**, 421 (1979) (E).
41. A.A. Anselm, JETP Lett **29**, 590 (1989).
42. N. Cabibbo, L. Maiani, G. Parisi and R. Petronzio, Nucl. Phys. **B158**, 295 (1979).
43. R.A. Flores and M. Sher, Phys. Rev. **D27**, 1679 (1983).
44. M.J. Duncan, R. Philippe and M. Sher, Phys. Lett. **B153**, 165 (1985)
45. M. Lindner, M. Sher and H.W. Zaglauer,Phys. Lett. **B228**, 139 (1989)
46. M. Sher,Phys. Lett. **B317**, 159 (1993) ; Phys. Lett. **B331**, 448 (1994) Nucl. Phys. **B395**, 17 (1993).
47. J.A. Casas, J.R. Espinosa and M. Quiros, Phys. Lett. **B342**, 171 (1995) ; Phys. Lett. **B382**, 374 (1996) .
48. J.R. Espinosa and M. Quiros, Phys. Lett. **B353**, 257 (1995) .
49. M. Quiros, Perspectives in Higgs Physics II, Ed. G.L. Kane, World Scientific, Singapore, hep-ph/9703412.

50. G. Altarelli and I. Isidori,Phys. Lett. **B337**, 141 (1994) .
51. J.R. Espinosa, Lectures presented at the XXIV ITEP Winter School, Snegiri, Russia, February 1996, hep-ph/9606316.
52. J. Gasser and H. Leutwyler, Phys. Reports **87**, 77 (1982).
53. S.G. Gorishny, A.L. Kataev, S.A. Larin and L.R. Surguladze, Mod. Phys. Lett **A5**, 2703 (1990).
54. N. Gray, D.J. Broadhurst, W. Grafe and K. Schilcher, Z. Phys. **C48**, 673 (1990).
55. R. Barbieri, M. Beccaria, P. Ciafaloni, G. Curci and A Vicere, Nucl. Phys. **B409**, 105 (1993).
56. A.I. Bochkarev and R.S. Willey, Phys. Rev. **D51**, 2049 (1995).
57. R. Hempfling and B.A. Kniehl, Phys. Rev. **D51**, 1386 (1995).
58. A. Sirlin and R. Zucchini,Nucl. Phys. **B266**, 389 (1986).
59. B. Schrempp and M. Wimmer, Prog. Part. Nucl. Phys. **37**, 1 (1997).
60. K. Inoue, A. Kakuto, H. Komatsu and S. Takeshita, Prog. Theor. Phys. **67**, 1889 (1982).
61. R.A. Flores and M. Sher, Ann. Phys. **148**, 95 (1982).
62. H. E. Haber and R. Hempfling, Phys. Rev. Lett. **66**, 1815 (1991).
63. A. Yamada, Phys. Lett. **B263**, 233 (1991)
64. P.H. Chanowski, S. Pokorski and J. Rosiek, Phys. Lett. **B274**, 191 (1992).
65. Y. Okada, M. Yamaguchi and T. Yanagida, Prog. Theor. Phys. **85**, 1 (1991).
66. J. Ellis, G. Ridolfi and F. Zwirner, Phys. Lett. **B257**, 83 (1991) and **B262**, 477 (1991).
67. R. Barbieri, M. Frigeni and F. Caravaglios, Phys. Lett. **B258**, 167 (1991).
68. J.R. Espinosa and M. Quiros, Phys. Lett. **B266**, 389 (1991).
69. LEP Working Group for Higgs Boson Searches, CERN-EP/2001-55
70. M. Carena, J. Conway, H. Haber, J. Hobbs, et al., hep-ph/0010338.
71. J.R. Espinosa and M. Quiros, Phys. Lett. **B279**, 92 (1992).
72. J.R. Espinosa and M. Quiros, Phys. Lett. **B302**, 51 (1993).
73. G.L. Kane, C. Kolda and J.D. Wells, Phys. Rev. Lett. **70**, 2686 (1993).
74. A.D. Sakharov, JETP Lett. **91B**, 24 (1967)
75. E.W. Kolb and M.S. Turner, "The Early Universe", Perseus Books, 1990.
76. A. Riotto and M. Trodden, Ann. Rev. Nucl. Part. Sci. **49**, 35 (1999).
77. A.G. Cohen, D.B. Kaplan and A.E. Nelson, Ann. Rev. Nucl. Part. Sci. **43**, 27 (1993).
78. M. Quiros, Helv. Phys. Acta **67**, 451 (1994).
79. V.A. Rubakov and M.E. Shaposhnikov, Phys. Usp. **39**, 461 (1996).
80. M. Carena and C.E.M. Wagner, Perspectives in Higgs Physics II, Ed. G.L. Kane, World Scientific, Singapore, hep-ph/9704347.
81. M. Quiros and M. Seco, Nucl. Phys. Proc. Supp. **B81**, 63 (2000), hep-ph/9703274.
82. G. 't Hooft, Phys. Rev. Lett. **37**, 37 (1976); Phys. Rev. **D14**, 3432 (1976).
83. V.A. Kuzmin, V.A. Rubakov and M.E. Shaposhnikov, Phys. Lett. **155B**, 36 (1985).
84. M. Quiros, Nucl. Phys. Proc. Suppl. **B101**, 401 (2001), hep-ph/0101230.
85. M. Gell-Mann, P. Ramond and R. Slansky, in *Supergravity*, North-Holland

Publishing, 1979.
86. P. Langacker, R.D. Peccei and T. Yanagida, Mod. Phys. Lett. **A1**, 541 (1986)
87. M. A. Luty, Phys. Rev. **D45**, 455 (1992).
88. C.E. Vayonakis, Phys. Lett. **B286**, 92 (1992).
89. A. Pilaftsis,Intl. J. Mod. Phys. **A14**, 1011 (1998).
90. I. Affleck and M. Dine, Nucl. Phys. **B249**, 249 (1985).
91. A. Dolgov, Physics Reports **222**, 381 (1992).
92. M. Dine, L. Randall and S. Thomas, Phys. Rev. Lett. **75**, 398 (1995) and Nucl. Phys. **B458**, 291 (1996).



MARIA SPIROPULU

This page intentionally left blank

COLLIDER EXPERIMENT: STRINGS, BRANES AND EXTRA DIMENSIONS

MARIA SPIROPULU
Enrico Fermi Institute
5640 S. Ellis Ave. Chicago IL
E-mail: smaria@fnal.gov

1. Preface

The subject of the Theoretical Advanced Study Institute summer 2001 school was "Strings, branes and extra dimensions". Although not too many years ago a course on collider physics would had been an unorthodox inclusion to the curriculum of a such titled school, today there are attempts to connect strings, branes and extra dimensions with experiment. Indeed string phenomenology is not an oxymoron, to the delight of all. I was particularly content to see young theorists spending time calculating the beam power in a future linear electron accelerator and the dollar sum necessary to pay the electric bill in a year. I was also very impressed that there were students who already knew what a trigger is and why we use it.

It took at least a generation of inspired physicists and also math experts to move from field theories to string theory, and a concurrent of technowizzes and inquiring minds to discover and measure with exquisite precision the theory describing nature at the most fundamental length scales yet explored: the Standard Model. Although it seemed briefly that theory and experiment are fast growing apart we witness today a remarkable exchange between the two, and a tendency to almost believe that by putting separate small bricks of knowledge together, one day preferably soon, the complete edifice of nature will be exactly blueprinted and raised.

The lectures were organized in three sections. The first one is devoted to accelerators. There is no doubt that these machines gave birth to the field we call High Energy Physics and filled in turn the Particles and Fields databooks. And there is no substitute for these machines. They are evolving, with the purpose of exploring the physics at the highest energy reachable or equivalently the physics at the shortest length scale. The second is an overview of the kind of experiments that lead us from atoms all the way down to quarks and a brief summary of the particle physics jargon used when we report results. And the third is analysis examples and interpretation of results when looking for new physics. Since the school's interests is on string theory I will give a supersymmetry search example and an example of the search for extra dimensions with collider data.

2. Accelerators

There are at least 10,000 accelerators in the world, most of them put to action in solving every day problems. A world-wide search by W.H. Scharf and O.A. Chomicki[1] reports 112 accelerators of more than 1 GeV. A third of these are used in high energy physics research. The rest are mostly synchrotron light sources. About 5000 accelerating machines of lower energy are used in medicine (radiotherapy, biomedical research and isotope production). About as many are used in industry, usually for ion implantation and surface physics. The latest sterilization via irradiation made the news in the post September 11 time when all the letters to Washington were collected to be irradiated for fear of anthrax. A great review of accelerators for medical applications has been written by Ugo Amaldi.[2]

We are familiar with the use of high energy lepton and hadron accelerators for particle physics. Present investigations are focused on the search for the Higgs boson, neutrino oscillations, heavy quark physics, the production of supersymmetric particles and even the geometry and geography of spacetime. There is little doubt that this type of research is connected intimately with cosmology in re-creating particles and interactions from the first instant of the Big Bang through the era when nuclei were formed by the more fundamental particles. The data accumulated from high energy particle collisions are essential in formulating and stimulating cosmological models and in helping understand the origin of the dark matter and perhaps even the dark energy in the universe.

I would like at this point to remark that it seems to me that when the director of the Office of Science and Technology Policy, Dr. Marburger, wrote[3] that at "some point we will have to stop building accelerators" he meant it in a similar way that NASA will stop building space shuttles. This is to say that new technologies will necessarily be employed in endeavors of tremendous magnitude such as the exploration of the universe in all scales; the exploration itself will not cease.

Table 1. Statistics of operating accelerators (1995).

Accelerators	Number in use
(1) High-energy more than 1GeV	112
(2) Radiotherapy	>4000
(3) Research/Biomedical Research	~800 ~5000
(4) Medical radioisotope production	~ 200
(5) Industry	~1500
(6) Ion implanters	>2000
(7) Surface modification centers and research	~1000
(8) Synchrotron radiation sources	50
total (in 1995) ~10000	

2.1. *DC accelerators*

In the beginning of the century there was chemistry, philosophy (cosmic ray research was being published in journals of philosophy), and by 1932 we had nuclear physics when the neutron was discovered by Chadwick. Three more particles were known by then; the electron (J.J. Thompson in 1897 passed electrons through crossed E and B fields, measured the velocity and then e/m of the electron), the proton (Rutherford, 1913 - he called the proton the hydrogen atom nucleus) and the photon (Max Planck, black body radiation in quanta of $h\nu$, Einstein photoelectric effect, Compton X-ray scattering from electrons as if particles. The photon was actually named by a chemist.)

During the first part of the century natural radioactivity and cosmic rays were the source of energetic particles for atomic physics research. In 1906 Rutherford bombarded a mica sheet with alpha particles from a natural radioactive source (Rutherford scattering). Natural sources are limited in energy and intensity. In 1928 Cockroft and Walton started thinking about building an accelerator to use at the Cavendish Laboratory. In 1932 the apparatus was finished and used to split lithium nuclei with 400 keV protons. The measurement of the binding energy in this experiment provided the first experimental verification of Einstein's mass-energy relationship (known since 1905). This was the start of particle accelerators for nuclear research. The very first ones were direct voltage accelerators like the Cockroft and Walton (rectifier generator up to 1 MV), the Van de Graff generator (up to 10 MV in high-pressure tank containing dry nitrogen or freon to avoid sparking) and the tandem electrostatic accelerator (the accelerator is known as tandem because the ions, at the beginning negative, undergo a double acceleration: they are attracted by the positive central electrode, pass in a cleaner who makes them positive, and they are then pushed back by the electrode; up to maybe 35 MV, Vivitron Strasbourg, in operation now at 20 MV).

2.2. *AC accelerators*

Ising in 1924 proposed the first particle accelerator that would give the particles more energy than the maximum voltage in the system. He proposed an electron linear accelerator with drift tubes (but did not build it). In 1928 Wideroe used an alternating 25 kV voltage with 1 MHz frequency applied over two gaps and produced 50 keV potassium ions. The Wideroe type linac comprises a series of conducting drift tubes. Alternate drift tubes are connected to the same terminal of the RF generator. The frequency is such that when a particle goes through the gap it sees the accelerating field and when the field becomes decelerating the particle is shielded inside the drift tube. As the particle gains in energy and velocity the structure periods must be longer in order to be in sync. At very high frequencies (so that the structure does not become inconveniently long) the open drift-tube scheme needs to be enclosed to form a cavity or series of cavities.

If one applies Ising's resonant principle in a homogeneous magnetic field the particle would be bent back to the same RF gap twice for each period. This is Lawrence's and Livingston's fixed-frequency *cyclotron* (the initial was less than one foot in diameter and could accelerate protons to 1.25 MeV). The resonance condition in the cyclotron is obtained by choosing the RF period equal to the cyclotron period, which is independent of the particle velocity and the orbit radius ; it depends on the q/m ratio and the magnetic field. Particles that pass the gap near the peak of the RF voltage would continue to do so every half turn moving in ever increasing half-circles (spiraling) until they reach the edge of the magnetic field or until they become relativistic and slip back with respect to the gap voltage. The intrinsic limit was confronted in the late thirties at about 25 MeV for protons and 50 MeV for deuterons and alpha particles. The cyclotron consisted of two "D" shaped regions in vacuum, called *dees* with a gap separating them and a magnetic field applied perpendicular to the dees. As the proton beam crosses the gap it experiences an electric field which gives the proton a kick and increases its energy. It gets an energy increase every time it crosses the gap. The frequency of the applied electric field is constant, while the radius of the proton beam keeps increasing.

Synchro-cyclotron -or the frequency-modulated cyclotron- was the remedy for the relativistic limit in which the revolution frequency decreases with increasing energy, and the frequency of the accelerating voltage must also be correspondingly decreased. Although this is necessary, it in not sufficient to maintain sync, because the natural energy spread in a bunch of relativistic ions causes a spread in their cyclotron frequencies, and thus longitudinal focusing is required to maintain the "bunch". The problem was overcome by McMillan and Veksler who discovered the principle of phase stability in 1944. The effect of phase stability is that a bunch of charged particles with an energy spread can be kept bunched throughout the acceleration cycle by injecting them at a suitable phase of the RF cycle. Synchro-cyclotrons can be used to accelerate protons up to 1 GeV. The higher energy is obtained at the expense of intensity (number of particles in the bunch) since the pulsed beam has less intensity compared to a continuous beam. To achieve transverse stability of the beam the field should be decreasing with radius according to an inverse power law. At Berkeley they found (the hard way) that the magnetic field has to decrease slightly with increasing orbit radius to prevent the particles from getting lost.

For electrons the cyclotron is useless as they are very quickly relativistic. The solution was the *betatron*, a device that has found applications in laboratories and hospitals. It was conceived again by Wideroe; Kernst built the first betatron in 1941 and Kernst and Serber published a paper on "betatron oscillations" the same year. In 1950 Kernst built the world's largest betatron. In a betatron, the electromagnet is powered with an AC current at 50 to 200 Hz. The magnetic field guides the particles in a circular orbit, but because it is a changing magnetic field, it induces a circumferential voltage which accelerates the particles. The guide field was carefully shaped and given a radial gradient in order to provide vertical and horizontal beam stability. If the electron path is to remain at constant radius the magnetic field must increase as the electron energy increases. The increasing field results in increasing magnetic flux through the orbits which then induces the force that increases the energy of the electrons. The magnetic flux through the orbit must be twice the bending field in order to keep the beam at the same radius.

The betatron was soon replaced by the *synchrotron* which is an accelerator that combines the properties of the cyclotron and those of the betatron. McMillan and Veksler already discusses the idea of synchronous acceleration in their cyclotron papers. The first synchrotron (Cosmotron) was a 3 GeV proton accelerator built in 1952 at Brookhaven National Laboratory (BNL). The machine had straight sections and a guide field similar to the betatron with a bending field to keep the particles on a circular orbit and appropriate radial gradient to achieve vertical and horizontal stability. Acceleration was achieved by RF voltage at the revolution frequency. As the particle energy increases the field is also increased at a rate that keeps the particles in approximately the same orbit at all energies. This means that the RF voltage frequency should be increasing (compare with cyclotron). Many large synchrotron machines followed the Cosmotron. The Tevatron at Fermilab is one of them, accelerating protons and antiprotons to approximately 1000 GeV.

In all accelerators build before 1952 the transverse stability of the beam depended on what is now called *weak focusing* in the magnet system. The guide field decreases slightly with increasing radius in the vicinity of the particle orbit and this gradient is the same all around the circumference of the magnet. The tolerance on the gradient is severe and sets limits to such an accelerator (believed to be around 10 GeV in the early fifties). The aperture needed to contain the beam becomes very large and the magnet becomes very bulky and costly. The invention of the *alternating-gradient* principle by Christofilos and independently by Courant, Livingston and Snyder in 1952 changed this picture. As a matter of fact, machines in the range of 10-100 TeV seem already technically possible. The limitation is the cost. Alternating-gradient or "strong" focusing is directly analogous to a result in geometrical optics, that the combined focal length of two appropriately spaced lenses of equal strength will be overall focusing when one lens is focusing and the other defocusing. Such a system remains focusing for quite some range of the focal length values of the two lenses. A quadrupole lense focuses on one plane and defocuses on the orthogonal plane. An appropriate arrangement of quadrupoles can be altogether focusing in both planes. Structures based on this principle are called *alternating gradient (AG) structures.*[a]

Although the circular machines took over for some time, the linear machines were revived after WW II with advances in ultra-high frequency technologies. At Berkeley a proton linac was built by Alvarez (1946) who employed the war-developed radar technology and enclosed the entire drift tube structure in a resonant cavity to reduce losses. Since then this type of accelerator has been widely used as an injector (with injection energies reaching 200 MeV) for large proton and heavy-ion synchrotrons (e.g. the

[a]The first such machine, the Brookhaven AGS, is still in operation today. The AGS was used to discover the J/Ψ (Nobel 1976 S. Ting), CP violation in the Kaon system (Nobel 1980 J. Cronin and V. Fitch) and the muon neutrino (Nobel 1988 J. Steinberger, L. Lederman, M. Schwartz).

Fermilab linac and the HERA linac at Deutches Electronen Synchrotron (DESY) in Hamburg, Germany). The largest proton linear accelerator today is the the 800 MeV proton linac at the Los Alamos Neutron Science Center (LANSE). The largest electron linear accelerator in operation is the 50 GeV linac at the Stanford Linear Accelerator Center (SLAC). Linear accelerators, like betatrons previously, have become very popular in fields outside particle physics (e.g materials science, biomedical research, and medicine).

An increase in beam energy of about 1.5 orders of magnitude per decade is illustrated by the Livingston chart shown in Figure 1. While the energy increases by 8 orders of magnitude the cost per GeV of a typical accelerator has been drastically reduced. But what is most amazing is that while each type of acceleration is coming to saturation relatively fast, there is an advance in technology that allows a different idea on acceleration to kick in. Examples: the invention of the alternating gradient focusing in the fifties, the *colliding beams* in the sixties, superconducting magnet technology and stochastic cooling in the seventies and eighties.

2.3. *Colliders/storage rings*

In high energy physics experiments, the type and number of particles brought into collision and their center of mass energy characterize an interaction. The center of mass energy E_{CM} made available when a synchrotron beam of energy E_b hits a fixed target is approximately

$$\sqrt{s} = E_{CM} = (2m_{target}E_b)^{1/2} \tag{1}$$

The center of mass energy when colliding head-on two beams of energy E'_b is

$$\sqrt{s} = E_{CM} = 2E'_b \tag{2}$$

The purpose of storage rings is to make head-on collisions possible with a useful interaction rate. At the Tevatron the maximum proton energy is about 1000 GeV. In the $p\bar{p}$ collider mode the center of mass energy is $\sqrt{s} = E_{CM} = 2000$ GeV. In the fixed target mode it is $\sqrt{s} = E_{CM} = 41$ GeV.

The fixed target collisions produce a variety of secondary particles from the target that can be collected and focused into secondary beams. Also in fixed target collisions the achieved luminosities are extremely high. The reaction rate R is $R = L \times \sigma$, where L is the luminosity and σ is the cross section. This implies that $L = (N_{beam}/\text{sec})(\mathrm{N_{target}}/\text{cm}^2)$. Luminosity is

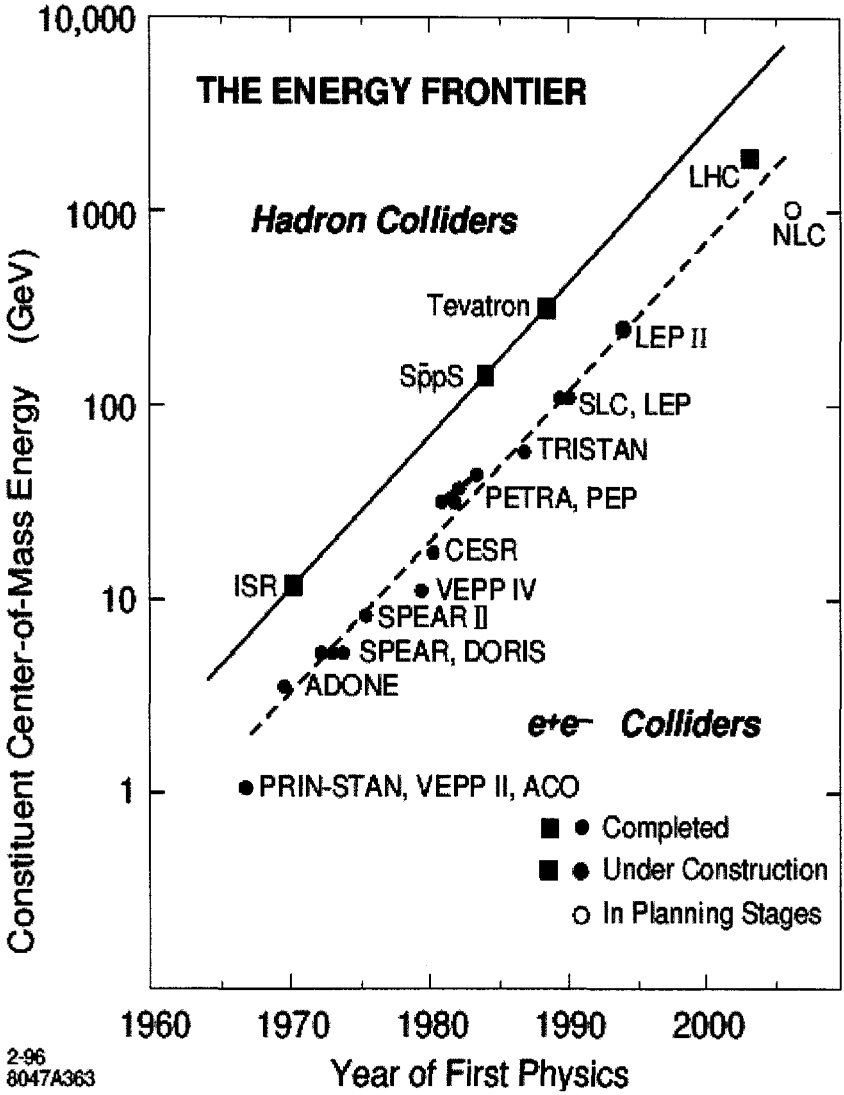


Figure 1. The effective constituent energy of existing and planned colliders and the year of first physics results from each (after Ref. 4).

measured usually in $\mathrm{cm}^{-2}\mathrm{sec}^{-1}$. A fixed target experiment can have luminosity as high as 10^{37} $\mathrm{cm}^{-2}\mathrm{sec}^{-1}$. In a high luminosity collider experiment the luminosity is of the order of 10^{32} $\mathrm{cm}^{-2}\mathrm{sec}^{-1}$.

What is the limitation on the energy for an accelerator storage ring? For proton accelerators it is the maximum magnetic field to bend the particles ($R_{ring}(m) = P(GeV)/0.3B(Tesla)$). For electron storage rings it is the energy lost to synchrotron radiation per revolution, $U \approx 0.0885E(GeV)^4R(m)^{-1}$. (For relativistic proton beams this is $\approx$ $7.79^{-15}E(\mathrm{GeV})^4\mathrm{R(m)}^{-1}$). The largest electron-positron storage ring was the 27 km LEP ring at CERN in Geneva. It started running at a beam energy of about 50 GeV: this meant that 200 MeV per turn should be supplied by the RF to make up for the synchrotron radiation loss. When it ran at 100 GeV, a factor of two higher in energy, the energy loss per turn was 3.2 GeV, a factor of 16 increase in losses! A linear electron-positron accelerator (such as the SLC, the Stanford Linear Collider) avoids the problem of synchrotron radiation. It is expected that all future electron-positron machines at higher energy than LEP will be linear.

Note that a monoenergetic proton (antiproton) beam is equivalent to a wide-band parton beam (where parton≡quarks, antiquarks, gluons), described by momentum distribution dn_i/dp (structure function, i specifies the parton type, i.e. $u, d, g, \bar{u}, \bar{d}...$). Proton structure functions are measured in *deep inelastic scattering* experiments.

Some of the advantages of hadron collisions include the simultaneous study of a wide energy interval, therefore there is no requirement for precise tuning of the machine energy; the greater variety of initial state quantum numbers e.g. $u+\bar{d} \to W^+$, $\bar{u}+d \to W^-$; the fact that the maximum energy is much higher than the maximum energy of e^+e^- machines; and finally that hadron collisions are the only way to study parton-parton collisions (including gluon-gluon). Some of the disadvantages are the huge cross sections for uninteresting events; the multiple parton collisions in the same hadron collision that result in complicated final states; and that the center-of-mass frame of the colliding partons is not at rest at the lab frame.

Table 2 lists some present, historic and proposed colliders.

2.4. *Future colliders*

We saw that the highest energy circular electron machine has been LEP. In order to go any higher in energy in electron-positron collisions, we need to build the collider on a straight line. The luminosity in terms of the con-

Table 2. Present, historic(including cancelled) and proposed colliders

Name	Lab	Location	GeV	status
Proton Fixed Target Accelerators				
PS	CERN	Geneva	28	inactive
AGS	BNL	Long Island	32	active
SPS	CERN	Geneva	450	active
Tevatron	FNAL	Chicago	1000	active
Electron Fixed Target Accelerators				
LINAC	SLAC	Stanford	50	active
Proton Colliding Beam Machines				
ISR	CERN	Geneva	$31p\times 31p$	inactive
$Sp\bar{p}S$	CERN	Geneva	$310p\times 310\bar{p}$	inactive
Tevatron	FNAL	Chicago	$1000p\times 1000\bar{p}$	active
SSC	SSCL	Dallas	$20000p\times 20000\bar{p}$	canceled
LHC	CERN	Geneva	$7000p\times 7000p$	2007
Electron Colliding Beam Machines				
SPEAR	SLAC	Stanford	$4e^- \times 4e^+$	inactive
DORIS	DESY	Hamburg	$5e^- \times 5e^+$	inactive
BES	BES	Beijing	$4e^- \times 4e^+$	active
CESR	Cornell	Ithaca	$8e^- \times 8e^+$	active
PEP	SLAC	Stanford	$15e^- \times 15e^+$	inactive
PEPII	SLAC	Stanford	$9e^- \times 3e^+$	active
PETRA	DESY	Hamburg	$23e^- \times 23e^+$	inactive
TRISTAN	KEK	Tsukuba	$30e^- \times 30e^+$	inactive
KEK B	KEKB	Tsukuba	$8e^- \times 3.5e^+$	active
SLC	SLAC	Stanford	$50e^- \times 50e^+$	active
LEPII	CERN	Geneva	$200e^- \times 200e^+$	inactive
Electron-Proton colliding Beam Machines				
HERA	DESY	Hamburg	$30e^- \times 820p$	active
Future Electron-Positron Collider Study Collaborations				
JLC	-	KEKB	(stage I) $125e^- \times 125e^+$	R&D
NLC	-	SLAC	$400e^- \times 400e^+$	R&D
TESLA	-	DESY	$400e^- \times 400e^+$	R&D
CLIC	-	CERN	$1500e^- \times 1500e^+$	R&D
Future Hadron Collider Study Collaborations				
VLHC	FNAL	Chicago	$40000p\times 40000p$	R&D

figuration of the beams is $L = \frac{f_{coll}N^2}{S}$ where f_{coll} is the number of bunch collisions per second, N is the number of particles in a bunch and S is the beam cross-sectional area at the collision point. In a linear collider clearly f_{coll} is much smaller than in a circular collider. The number of particles N in a bunch is similar or smaller. Therefore to achieve worthwhile luminosities it is imperative that the beam emittance S be much smaller. A linear collider needs very high accelerating gradient to keep the site length reasonable and very high power efficiency to keep the cost under control. As we noted the beams need to be generated with extremely small emittance which has to be preserved during acceleration. At the collision point the beams have to be tightly focused. Small emittances can be achieved by preparing the beam through what is called a damping ring and making use of the synchrotron radiation. The acceleration is achieved by microwave cavities which can be superconducting or normal conducting, and the frequency choice is based on the achievable accelerating gradient. There are at least two different designs under development: The NLC/JLC [b] design that uses normal conducting copper cavities at low frequency (11.424 GHz or X-band) which is also referred to as *warm* design, and the TESLA [c] design that uses supeconducting niobium cavities, a technology that is power consumption effective (there is extremely small power dissipation in the cavity walls) but thought to be very expensive until recently. This is referred to as *cold* design because the cavities need to be kept at very low temperatures. Linear collider accelerating structures can also be used to drive free electron lasers, with important applications for chemistry, materials science, plasma research and life sciences research.

A post-LHC and probably post-LC collider that accelerates and collides electron and positron beams with center of mass 3-5 TeV is envisioned in the CLIC[d] design. It is based on the two-beam acceleration method in which the RF power for sections of the main linac is extracted from a secondary, low-energy, high-intensity electron beam running parallel to the main linac.

The next hadron collider proposed is the VLHC [e]. A design study group has developed the basic parameters, technology/construction challenges and cost of a proton-proton collider with center of mass energy greater than 30 TeV, that would allow the eventual operation of a collider with center of mass energy greater than 150 TeV in the same tunnel. The machine would involve two phases, one with low-field magnets and one with high-field magnets for the energy upgrade in the same 233 km ring.

It is not included in the table, but extensive study and research is geared towards the feasibility and potential of high energy high luminosity muon colliders operating at a center-of-mass energy in the range 100 GeV - 4

---

[b] Next Linear Collider, Japan Linear Collider
[c] Tera Electronvolt Superconducting Linear Accelerator
[d] Compact Linear Collider
[e] Very Large Hadron Collider

TeV. The high intensity muon source needed for muon colliders can also be used to feed a muon storage ring neutrino source (neutrino factory). The enthusiasm in recent years for a muon collider is due to ionization cooling that allows for very bright muon beams.

One lesson from the Livingston plot is that a new technology can push the collision energy and drive high energy research. Indeed accelerator technology is also moving towards alternative directions. High-gradient plasma wakefield acceleration is one. The plasma-wave acceleration process is dramatically strong and may lead to very high energy beams with reasonable size accelerators.

3. From atoms to quarks, *Rutherford redux*

The kind of experiments we perform to probe the structure of matter belong to three categories: scattering, spectroscopy and break-up experiments. The results of such experiments brought us from the atom to the nucleus, to hadrons, to quarks.

Geiger and Marsden reported in 1906 the measurements of how α particles (the nuclei of He atoms) were deflected by thin metal foils. They wrote "it seems surprising that some of the α particles, as the experiment shows, can be turned within a layer of 6×10^{-5} cm of gold through an angle of 90°, and even more". Rutherford later said he was amazed as if he had seen a bullet bounce back from hitting a sheet of paper. It took two years to find the explanation. The atom was known to be neutral and to be containing negatively charged electrons of mass very much less than the mass of the atom. So how was the positively charged mass distributed within the atom? Thompson had concluded that the scattering of the α particles was a result of "a multitude of small scatterings by the atoms of the matter traversed". This was called the *soft model.* However "a simple calculation based on probability shows that the chance of an α particle being deflected through 90° is vanishingly small". Rutherford continued "It seems reasonable to assume that the deflection through a large angle is due to a single atomic encounter, for the chance of a second encounter of the type to produce a large deflection must be in most cases exceedingly small. A simple calculation shows that the atom must be a seat of an intense electric field in order to produce such a large deflection in a single encounter". According to the soft model the distribution of scattered particles should fall off exponentially with the angle of deflection; the departure of the experimental distribution from this exponential form is the signal for hard scattering.

The soft model was wrong. Geiger and Marsden found that 1 in 20,000 α particles was turned at 90° or more in passing through a thin foil of gold; The calculation of the soft model predicted one in 10^{3500}. The nuclear atom was born with a hard constituent , the small massive positively charged nucleus. Rutherford calculated the angular distribution expected from his nuclear model and he obtained the famous $\sin^4(\phi/2)$ law, where ϕ is the scattering angle.

At almost the same time Bohr (1913) proposed the model for the dynamics of the nuclear atom, based on a blend of classical mechanics and the early quantum theory. This gave an excellent account of the spectroscopy of the hydrogen. Beginning with the experiments of Frank and Hertz the evidence for quantized energy levels was confirmed in the inelastic scattering of electrons from atoms.

The gross features of atomic structure were described well by the non-relativistic quantum mechanics of point-like electrons interacting with each other and with a point-like nucleus, via Coulomb forces.

As accelerator technology developed it became possible to get beams of much higher energy. From the de Broglie $\lambda = h/p$ relationship it is clear that the resolving power of the beam becomes much finer and deviations from the Rutherford formula for charged particles scattering (which assumed point-like α's and point-like nucleus) could be observed. And they did! At SLAC in 1950 an electron beam of 126 MeV (instead of α's) was used on a target of gold and the angular distribution of the electrons scattered elastically fell below the point-nucleus prediction. (qualitatively this is due to wave mechanical diffraction effects over a finite volume on the nucleus). The observed distribution is a product of two factors: the scattering from a single point-like target (*a la Rutherford* with quantum mechanical corrections, spin, recoil etc.) and a "form factor" which is characteristic of the spatial extension of the target's charge density.

From the charge density distribution it became clear that the nucleus has a charge radius of about 1-2 fm (1 fm=10^{-15} m); In heavier nuclei of mass number A the radius goes as $A^{1/3}$ fm. If the nucleus has a finite spatial extension, it is not point-like. As with atoms, inelastic electron scattering from nuclei reveals that the nucleus can be excited into a sequence of quantized energy states (confirmed by spectroscopy).

Nuclei therefore must contain constituents distributed over a size of a few fermis, whose internal quantized motion lead to the observed nuclear spectra involving energy differences of order a few MeV.

Chadwick discovered the neutron in 1932 establishing to a good approximation that nuclei are neutrons and protons (generically called nucleons). Since the neutron was neutral, a new force with range of nuclear dimensions was necessary to bind the nucleons in the nuclei. And it must be very much stronger than the electromagnetic force since it has to counterbalance the "uncertainty" energy ($\approx$20 MeV; the repulsive electromagnetic potential energy between two protons at 1 fm distance is ten times smaller, well below the nucleon rest energy).

Why all this? The typical scale of size and energy are quite different for atoms and nuclei. The excitation energies in atoms are in general insufficiently large to excite the nucleus: hence the nucleus appears as a small inert point-like core. Only when excited by appropriately higher energy beams it reveals that it has a structure. Or in other words the nuclear degrees of freedom are frozen at the atomic scale.

To figure out whether the nucleons are point-like the picture we already developed is repeated once again. First elastic scattering results of electrons from nucleons (Hofstadter) revealed that the proton has a well defined form factor, indicating approximately exponential distribution of charge with an *rms* radius of about 0.8 fermi. Also the magnetic moments of the nucleons have a similar exponential spatial distribution. Inelastic scattering results, as expected, showed signs of nucleon spectroscopy which could be interpreted as internal motions of constituents. For example in the scattering of energetic electrons from protons there is one large elastic peak (at the energy of the electron beam) and other peaks that correspond to excitations of the recoiling system. The interpretation of the data is a hairy story: several excited recoil states contribute to the same peak and even the apparently featureless regions conceal structure. In one of the first experiments that used close to 5 GeV electrons, only the first of the two peaks beyond the elastic had a somewhat simple interpretation: It corresponded exactly to a long established resonant state observed in pion-nucleon scattering and denoted by Δ. Four charge combinations correspond to the accessible pion-nucleon channels: π^+p, $\pi^+n(\pi^0p)$, π^-p (π^0n), π^-n. The results of such "baryon spectroscopy" experiments revealed an elaborate and parallel scheme as the atomic and nuclear previously. One series of levels comes in two charged combinations (charged and neutral) and is built on the proton and neutron as ground states; the other comes in four charge combinations (–,0,+,++) with the Δs as ground state.

Yukawa predicted the pion as the quantum of the short-range nuclear forces. In the 60s the pion turned out to be the ground state of excited states forming charge triplets. Lets note again that we are looking at the excitations of the constituents of composite systems. Gell-Mann and Zweig proposed that the nucleon-like states (baryons), are made of three 1/2 spin constituents: the quarks. The mesons are quark-antiquark bound states. Baryons and mesons are called collectively hadrons. As in the nuclear case the simple interpretation of the hadronic charge multiplets is that the states are built out of two types of constituents differing by one unit is charge - hence its constituents appeared fractionally charged. the assignment was that the two constituents were the up (u) and down (d) with 2/3 and -1/3 charge. The series of proton would then be uud and the neutron udd while the Δ^{++} would be uuu. The forces between the quarks must be charge independent to have this kind of excited states. Lets point here that the typical energy level differences in nuclei are measured in MeV. For the majority of nuclear phenomena the neutrons and protons remain in their ground unexcited states and the hadronic excitations are being typically of order MeV; the quark-ian, hadronic degrees of freedom are largely frozen in nuclear physics. People did not like the quarks; They really thought it was nonsense. Nevertheless a very simple "shell model" approach of the nucleon was able to give an excellent description of the hadronic spectra in terms of three quarks states and bound states of quark-antiquark. So now, how would we try to see those quarks directly? We need to think Rutherford scattering again and extend the inelastic electron scattering measurements to larger angles. The elastic peak will fall off rapidly due to the exponential fall off of the form factor. The same is true for the other distinct peaks; this indicates that the excited nucleon states have some finite spatial extension. The bizarre thing is that for large energy transfer the curve does not fall as the angle increases. In other words for large enough energy transfer the electrons bounce backwards just as the αs did in the Geiger-Marsden experiment. This suggests *hard* constituents! This basic idea was applied in 1968 in so-called *deep inelastic scattering* experiments by Friedman, Kendall and Taylor in which very energetic electrons were scattered off of protons. The energy was sufficient to probe distances shorter than the radius of the proton, and it was discovered that all the mass and charge of the proton was concentrated in smaller components, spin 1/2 hadronic constituents called "partons" which were later identified with quarks. A lot of spectroscopy and scattering data became available in the '70s and still people used the quarks as mathematical elements that help systematize a bunch of complicated data. In fact the following quote is attributed to Gell-Mann: "A search for stable quarks of charge -1/3 or +2/3 and/or stable di-quarks of charge -2/3 or +1/3 at the highest energy accelerators would help reassure us of the non-existence of real quarks". Indeed quarks have not been seen as single isolated particles. When you smash hadrons at high energies, where you expect a quark what you observe downstream is a lot more hadrons - not fractionally charged quarks. The explanation of this quarky behavior – that they don't exist as single isolated particles but only as groups "confined" to hadronic volumes – lies in the nature of the interquark force. November 1974 was a revolution of quarks: A new series of mesonic (quark-antiquark) spectra, the J/ψ particles, were discovered, with quantum number characteristics of fermion-antifermion states. The J/ψ spectrum is very well described in terms of $c\bar{c}$ states where c is a new quark: the *charm*. The $c\bar{c}$ is called *charmonium* after the e^+e^- *positronium*. There is a funny resemblance between the energy states of the charmonium with those of the positronium given that the positronium is bound via electromagnetic forces while the charmonium via strong forces.

Back to Rutherford again. We saw how the large angle large energy transfer electron scattering from nucleons provided evidence for "hard constituents". What if we collide two nucleons? With a "soft" model of the nucleons we expect some sort of an exponential fall off of the observed "reaction products" as a function of their angle to the beam direction. On a "hard" model we should see prominent "events" at wide angles, corresponding to collisions between the constituents. The hard scattered quarks are converted into two roughly collimated *'jets' of hadrons*. These jets and their angular distributions provide indirect evidence of quarks. At $p\bar{p}$ collisions at CERN jets were observed in the 80s when CERN achieved with SPS the largest momentum transfers. Clear evidence of hadronic jets associated with primary quark processes were observed earlier in electron-positron collisions.

If quarks are not point-like we expect to see at higher energies, where the sub-quarkian degrees of freedom unfreeze perhaps at the Tevatron and the LHC, deviations from the theory similar to the deviations observed in the deep inelastic scattering experiments.

3.1. *Technical handbook*

The fragments of a high energy collision in matters of nanoseconds have decayed and/or left the detectors. In the early times of particle experiments an event was a picture in a bubble chamber for example, of the trail the particles left when ionizing a medium. Now an event is an electronic collection of the trails many particles left in a multiple complex of detectors.

We are taking a little detour here to define some jargon particle physicists use, to discuss briefly what is the modern process of particle detection and what is the data analysis after all. To start, there is a physics collision for which we have a theoretical model to describe the particle interaction (we draw a Feynman graph). The fragments of the collision decay and interact with the detector material (we have for example multiple scattering). The detector is responding (there is noise, cross-talk, resolution, response function, alignment, temperature, efficiency...) and the real-time data selection (trigger) together with the data acquisition system give out the raw data (in the form of bytes; we read out addresses, ADC and TDC values and bit patterns). The analysis consists of converting the raw data to physics quantities. We apply the detector response (e.g. calibration, alignment) and from the interaction with the detector material we perform pattern recognition and identification. From this we reconstruct the particles' decays and get results based on the characterization of the physics collision. We compare the results with the expectations by means of a reverse path that simulates the physics process calculated theoretically and driven through a computational model of the detector, the trigger and data acquisition path (Monte Carlo). The challenge is to select the useful data and record them with minimum loss (deadtime) when the detector and accelerator is running properly; and of course to analyze them and acquire results that are statistically rather than systematically limited. When the statistical uncertainty becomes smaller than the systematic uncertainty, it is time to build a new experiment for this measurement.

To give an example, in a hadronic collision at CMS in 2007, the interaction rate is 40 MHz (corresponding to data volume of 1000 TB/sec). The first level of data selection is hardware implemented and by using specific low level analysis reduces the data to 75 kHz rate (corresponding to data volume of 75 GB/sec). The second level of judging whether an event is going to be further retained is implemented with embedded processors that reduce the data rate to 5 Hz (5 GB/sec). The third level of the trigger is a farm of commodity CPUs that records data at 100 Hz (100 MB/sec). These data are being recorded for offline analysis. The final data volume depends on the physics selection trigger and for example we expect at the first phase of the Tevatron Run II to have between 1 and 8 petabytes per experiment. The improvement in high energy experiments is multifold. Better accelerator design and controls give higher energies and collision rates. Better trigger architecture is making best use of the detector subsystems. Better storage, networks and analysis algorithms contribute to precise and statistically significant results. Large CPU and clever algorithms improve the simulations and theoretical calculations that in turn enable the discovery of new physics in the data.

To summarize the jargon: *trigger* is a fast, rigid and primitive, usually hardware implemented selection applied on raw data or even analog signals from the detectors. Triggers have *levels* that an event passes through or fails. The trigger system and algorithms are emulated so that the signals are driven through a computational model of the trigger path, and the results are compared with the acquired data for diagnostic purposes. The efficiency of a multilevel trigger path is measured in datasets coming from orthogonal trigger paths (e.g. you want to measure the efficiency of a multilevel missing energy trigger in a dataset of events that come from a trigger that has no missing energy in its requirements). The duration of time when the data acquisition system cannot accept new data (usually because it is busy with current data) is called *deadtime.* To go faster modern experiments are running parallel data acquisition on sub-detector systems that are ending in fanned out triggers (*data streams*); the combination of the fragments of data from the detector subsystems is the *event building.* A *filter* is a later selection after the trigger, which is usually software implemented and can be sophisticated. *Reconstruction* is the coding that converts the sparsified hardware bytes to physics objects (tracks, vertices, tags etc.) A computer *farm* is a dedicated set of processors and associated networks used to run filters, event reconstruction, simulation etc. *Efficiency* is the probability to pass an event (signal or background). *Enhancement* is the enrichment of the data sample after a selection is applied.

4. Bottom-top experimental approach

There is a lot of unfinished business before we confront strings and extra dimensions in colliders. However, we investigate phenomena in fundamental physics not diagram by diagram, but scale by scale as it has been pointed out by Joe Polchinski in his string colloquium .[5] We do this in theory and unavoidably in experiment too: a fact that is punctuating the role of higher energy accelerators. In the section discussing the path from atoms to quarks, we saw that different degrees of freedom are frozen at different energy regimes and unveiled in others. We also saw that the energy scale at which these degrees of freedom show up is only discovered by means of experimental data. It is very difficult to theoretically determine the energy scale where new phenomena emerge (in the words of theorist Joe Lykken[6]), in particular when one has no idea what these phenomena are, and even if one has a good scenario of what they may be. In the past few years space itself is approached as degrees of freedom that may unfreeze and emerge at a particular characteristic energy scale.

In particle physics we talk about two major scales, one of which we have extensively studied theoretically and experimentally. It is the electroweak scale (10^{-17} cm or 10^{11} eV) where three of the four forces of nature have comparable strengths and the masses of the W and Z bosons are generated. The other distinct scale is the Planck scale which is very distant in value from the electroweak scale (10^{-33} cm or 10^{27} eV) and is the scale at which gravity would have comparable strength with the rest of the known forces of nature. At the electroweak scale all physical phenomena except for gravity are very well described by the Standard Model of the elementary particles and their interactions. At the scale of 100μm gravity follows Newton's law[36] and at large length scales the general theory of relativity takes over. With the present data and understanding of physics at the electroweak scale we can argue that a compelling completion of the standard model is found in supersymmetry, discovered by Pierre Ramond[7] Wess, Zumino, and others in the seventies as a side-effect of theoretical attempts to include fermions in string theories. Supersymmetry is a good theory that gives back more than the inputs it requires. The theory is supremely decorated, among others with the best candidate for the dark matter of the universe, the path towards the inclusion of gravity in a unified way with the rest of the forces, the exquisite prediction of the GUT scale where all couplings meet, the accurate and precise (within 1% of its measured value) predicted value of the weak angle at the electroweak scale from the GUT scale, the heaviness of the top quark and its function in electroweak symmetry breaking through radiative corrections, and even clues towards the understanding of the negative pressure that drives the recently discovered accelerating universe. We must keep in mind the fact that all the above can be realized with supersymmetry at the electroweak scale, which is within the reach of experiment.

A variety of supersymmetric models have been developed and evolved in the past decade.[8] The differences lying in the assumed SUSY breaking mechanism and the identity of the lightest supersymmetric particle. Models with photons and leptons in the final state are usually gauge mediated, models with, "disappearing" tracks are signatures of anomaly mediation while jets, leptons and missing energy are signatures of generic minimal supersymmetric models. Although supersymmetry is the best bet for the next great scheme that will include and further the Standard Model description of nature, there is hardly a good realizable model, *the* model, to be put to the test. Clearly this will be accomplished when we discover the first supersymmetric particles. However it is of critical importance to use the data we have in the same way the blind use their walking stick to go about. Let me give an example of what I mean. In the time when LEP results were pouring in, the visionaries expected supersymmetric particles right around the corner. Corner by corner there was no sign of them. Within the supersymmetric models developed at that time, and using the data acquired, they discovered false theoretical assumptions and corrected them. With the new assumptions in the models the vision was moved to a different value of the similar scale. Now we expect supersymmetric particles to be found at the Tevatron and the LHC. In other words there is an almost divine feedback mechanism between data, their studious interpretation and the tracking of the theory. The data is sculpting the theory. Experimentalists are comparing the observed data with the standard model predictions and use the result of the comparison to probe particular theoretical models. The colorful complicated exclusion and reach plots that are generated, although attempted to be as model independent as possible, are always increasing our understanding of a particular theory and why it might not be the proper one. If supersymmetry- despite being the most educated formulation of the biggest theory we are looking for- is wrong, the data will show it and will point to a better theory. Rather curious as it may seem at first look I completely agree with E. Witten that ". . . One of the biggest adventures of all is the search for supersymmetry".[8]

The bottom-top experimental approach in the title of this section refers to the upwards evolving energy of the machines we build and correspondingly the energy scale we explore. However nature is tricky, and so for example we have measured the mass of the heaviest of the quarks before the masses the tiniest of the leptons, the neutrinos. There is always a chance that although machine-wise we work our way up in energy, the physics of the highest energies shows up unexpectedly. This is the case in the scenarios of large extra dimensions that have been extensively studied in this summer school, where the real Planck scale is indeed allowed to take values close to the electroweak scale. The same is true for the physics of strings and black holes in high energy collisions.

In the next section I will give an example of a a data analysis that was interpreted in two SUSY models, one more and one less constrained, and the ingredients that enter this analysis. And similarly for an extra dimensions analysis. I would also like to refer you to a marvelous report on particle physics for theorists that appeared in 1996 TASI proceeding by Persis Drell.[9]

4.1. *Example 1: From a trigger, to a model, to a signature, to the data, and back to a model*

The "missing energy plus jets" signature is referred to as one of the golden signatures when searching for SUSY in hadron colliders. The reason is the large rate at which squarks and gluinos are produced and the abundance of the lightest supersymmetric particle in the decay chain of supersymmetric particles. The large missing energy would originate from the two LSPs in the final states of the squark and gluino decays. The three or more hadronic jets would result from the hadronic decays of the $\widetilde{q}$ and/or $\widetilde{g}$. It is not only SUSY that gives you such a signature. Leptoquark or technicolor models can have the same or similar final state. The general analysis direction for the search is similar for different models and the data reduction steps are kept as inclusive as possible. However, when optimizing an analysis a model is chosen, in this case supersymmetry.

We do use missing energy collider data for measurements of the Z boson invisible decay rates,[10] the top quark cross section,[11] searches for the Higgs boson[12] and other non-Standard Model physics processes.[13,14,15]

In a detector with hermetic 4π solid angle coverage the measurement of missing energy is the measurement of neutrino energy plus the energy of any unknown weakly interacting particle.. In a real detector it is also a measurement of energy that escapes detection due to uninstrumented regions. Jets dominate the cross section for high P_T proton-antiproton scattering and contribute to a final all-hadronic state with missing energy from the heavy flavor decays but mostly from jet energy resolution and mis-measurements. QCD production (including $t\bar{t}$) and W/Z QCD associated production dominate in a sample of events with large missing energy and multiple jets. This is because of the invisible Z decays and the decays of the W to a charged lepton and a neutrino. These processes can be understood with the data before narrowing the search for any exotic signal with the same detection signature.

In this particular example the *blind box* method was employed. It was discussed by R. Cousins and others in the past decade and various versions and improvisations of the method have been used in different measurements and searches.

Important data analysis methods to identify in this example are a) all data samples are revertexed and reclustered offline according to the best determined hard scattering vertex; b) on all Monte Carlo samples a tracking degradation algorithm is applied to appropriately account for the tracker aging c) a two stage cleanup algorithm is applied to reduce a sample that has a two-to-one noise to physics ratio and make it useful for further analyses; d) orthogonal data samples are used to normalize the theory predictions for most of the Standard Model backgrounds. In all predictions the appropriate trigger efficiencies measured in the data are folded in before the comparison with the data and extraction of the normalization factor. Notice that when you normalize to the data the uncertainty on the normalized predictions comes mostly from the statistics of the data samples used, e) in the case of a blind box analysis, we form control regions around the box by inverting the requirements which define it and compare the data in these regions with the normalized prediction to make sure that the normalization of the background samples are accurate and to diagnose potential pathologies before opening the box. The same is true if the analysis in not blind: control regions should be checked for the validity of the standard model predictions. The main objective of a blind analysis is to avoid biased human decisions involving the data selection. We achieve this by insulating the signal candidate data sample until we estimate the total background. We then *a priori* define the signal candidate data sample based on the signal signature and the total background estimate and precision. In this analysis we use three variables to define the signal candidate region : The missing transverse energy, $\not\!\!E_T$, the scalar sum $H_T \equiv E_{T(2)} + E_{T(3)} + \not\!\!E_T$, and the isolated track multiplicity, N_{trk}^{iso}.[43] The blind box contains events with $\not\!\!E_T \geq 70$ GeV, $H_T \geq 150$ GeV, and N_{trk}^{iso}=0. Several questions arise pertaining to the choice of the variables and their values. The value 70 GeV is chosen based on the MET trigger efficiency. We need to know how many standard model background events and how many for example SUSY events would pass through the same triggering scheme and be registered in the data sample that we are examining, in this case the $\not\!\!E_T$ sample. We normalize the standard model backgrounds using differently triggered samples (e.g lepton triggered samples), we apply the normalization to the prediction and we also fold in the $\not\!\!E_T$ trigger efficiency. More than 95% of events with reconstructed $\not\!\!E_T$ above 70 GeV would have passed the $\not\!\!E_T$ trigger during data taking. We extract this from jet data samples that do not require missing energy in their trigger path. Is this the optimal value for searching for new physics? Starting with any value less than 70 GeV would introduce larger uncertainties from the trigger efficiency. For values above 70 GeV we need to optimize the signal to background ratio for a particular signal. The H_T variable is constructed so as to play a discriminant role between the signal and the standard model backgrounds. Notice that the sum does not include the first jet which is the well measured jet. This is so that we discriminate events with real missing energy from events where the missing energy is a result of jet (second and third usually) mismeasurements. At LHC just the sum of all jets in the event would be a good variable. The track isolation variable is a counter of the number of high p_T isolated tracks. By requiring it be zero we indirectly veto event with leptons, this way focusing the search to all-hadronic states, and f) for processes where the measured or theoretical cross sections have large uncertainties such as most QCD processes we need to find adequate "standard(izable) candles"; a good example is the dielectron+jets event data sample where the invariant mass of the pair is consistent with the Z boson mass.

The parts of the analysis are:

- The data Pre-Selection, designed to acquire a high purity sample of large real missing energy data (this is the cleanup from the junk that end up in the sample: everything that goes wrong in the detector as well as cosmics, end up in this sample). The result of this cleaning of the sample is shown in Figure 2.
- The W and Z boson QCD associated production background estimate. As mentioned, the $Z(\rightarrow e^+e^-) + jets$ data sample is used as standardizable candle to normalize the theoretical rate predictions. The sample used is the events selected from a high energy electron trigger that have a second electron and the invariant mass of the the two is consistent with the Z mass as shown in Figure 4.1. Exactly the same selection rules and corrections are applied in the data and Monte Carlo samples.
- The multijet QCD production background estimate. The JET data samples are used to normalize the theoretical rate prediction. The source of missing energy in QCD jet production is the small fraction of $b\bar{b}$ and $c\bar{c}$ content (with the b and c quarks decaying semileptonically) and largely the jet mismeasurements and detector resolution. Analyses that require a measurement of the QCD multijet background use the jet data when there are extra requirements such as a b-tagged jet or a lepton. The data give a reliable estimate for the QCD background in such analyses. Comparisons between data and

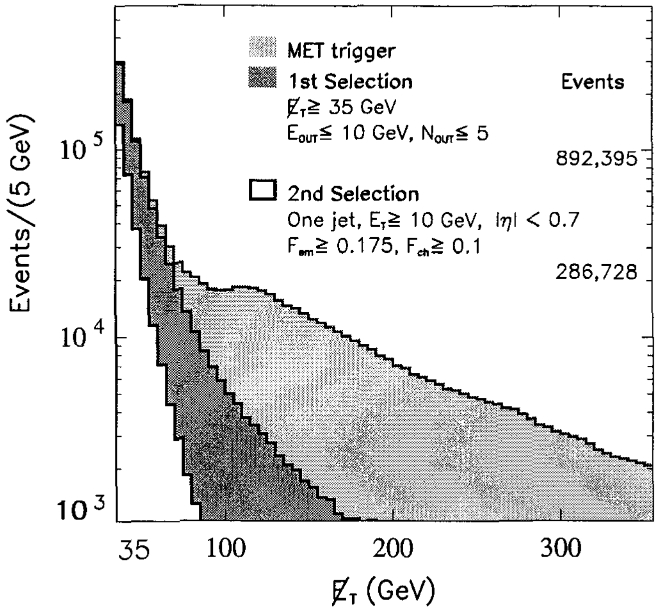


Figure 2. The $\not{E}_T$ spectrum after the online trigger and the two stages of the data preselection. The numbers of events surviving the first and second selections are 892,395 and 286,728, respectively. The variables E_{OUT}, N_{OUT} are energy and number of towers out of time.

HERWIG QCD Monte Carlo for the ΣE_T cross section measurement at CDF indicate agreement between the data and the Monte Carlo predictions. In the case of the large missing energy plus multijet ($\geq$ 3 jets) search the estimate of the QCD background is nontrivial. The missing energy trigger accepts QCD multijet events (45% of the online missing energy trigger are volunteers from jet triggers) and the trigger threshold is too high[f] to allow use of the low missing energy triggered data for extraction of the high missing energy spectrum. The high energy threshold jet triggered data with small

[f]For Run II the MET trigger now taking data, is designed with a lower threshold (25 GeV).

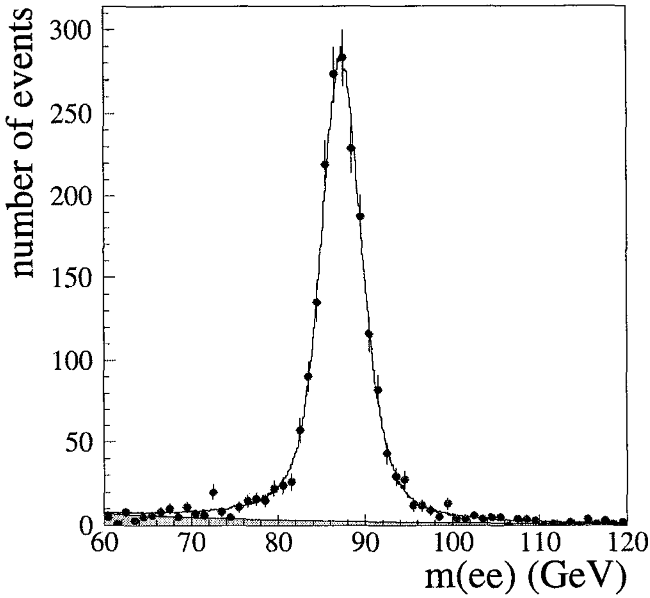


Figure 3. The Z^0 mass as reconstructed in the mode $Z^0 \to e^+e^-$ by the DØ detector. The shaded region at the bottom of the plot is the background contribution. The peak does not fall exactly on the true value of M_Z because not all of the energy corrections have been applied to the data. This is why it is referred to in the text as a *standardizable candle*; After the appropriate corrections are made the events in the Z boson mass peak are used for the normalization of the the vector boson+jets rates. The histogram is the prediction and the points the DØ data.

or no prescale (JET70, JET100) are not suitable to extract the QCD contribution to the high $\not\!\!E_T$ tails as they themselves constitute signal candidate samples. The lower energy threshold jet triggered data with large prescales (JET20, JET50) are used in this analysis to estimate the QCD jet production contribution to the high missing energy spectrum. Large statistics 3-jet QCD Monte Carlo samples are generated to simulate the JET20 and JET50 data samples and used to compare the shapes of the missing energy and the N-jet distribution with the data. The predictions are absolutely normalized to the data.

- The comparisons of the total background estimates with the data in the control regions around the Blind Box.

There are seven control regions around the blind box formed by inverting the requirements which define it. We compare the

Table 3. Comparison of the Standard Model prediction and the data in the control regions and the signal candidate region (blind box). After the contents of the control regions were compared in detail to standard model predictions, we opened the box and found 74 events. ($\not{E}_T$ and H_T in GeV.)

Region Definition	EWK	QCD	All	Data
$\not{E}_T \geq 70, H_T \geq 150, N_{trk}^{iso} > 0$	14	6.3	20±5	10
$\not{E}_T \geq 70, H_T < 150, N_{trk}^{iso} = 0$	2.3	6.3	8.6±4.5	12
$35 < \not{E}_T < 70, H_T > 150, N_{trk}^{iso} = 0$	1.95	135	137±28	134
$\not{E}_T > 70, H_T < 150, N_{trk}^{iso} > 0$	1.7	<0.1	1.7±0.3	2
$35 < \not{E}_T < 70, H_T > 150, N_{trk}^{iso} > 0$	14	9.4	23±6	24
$35 < \not{E}_T < 70, H_T < 150, N_{trk}^{iso} = 0$	5	413	418±69	410
$35 < \not{E}_T < 70, H_T < 150, N_{trk}^{iso} > 0$	3.3	28	31±10	35
Signal Candidate Region				
$\not{E}_T \geq 70, H_T \geq 150, N_{trk}^{iso} = 0$	35	41	76±13	74

Standard Model background predictions in the control regions with the data. The results are shown in Table 3.

Of the 76 events predicted in the blind box, 41 come from QCD and 35 from electroweak processes. Of the latter we estimate ~37% coming from $Z(\rightarrow \nu\bar{\nu})+ \geq 3$ jets, ~20% from $W(\rightarrow \tau\nu)+ \geq 2$ jets, ~20% from the combined $W(\rightarrow e(\mu)\nu_e(\nu_\mu))+ \geq 3$ jets, and ~20% from $t\bar{t}$ production and decays. We also compare the kinematic properties between Standard Model predictions and the data around the box and find them to be in agreement.[17]

As we mentioned in the introduction we can stop here and show histograms of how well the data match the Standard Model predictions both in the blind box and in the control regions 4.1.

We can combine the level of agreement in this channel with a number of searches in other final states and make a global analysis of how well all the data in all analyses match the Standard Model predictions. However we choose to take the analysis a step further. We pick two models and provide an answer to a model builder who wants to know how light a gluino is allowed to be based on this particular search. We study the hadroproduction of scalar quarks and gluinos and all their decays in minimal Supersymmetry and Supergravity frameworks. Important theoretical considerations that need to be underlined are a) the use of $\tan\beta = 3$ to generate datasets of squark and gluino events, b) the study of the production of only thefirst two heavy generations of squarks $(\tilde{u}, \tilde{d}, \tilde{c}, \tilde{s})$ in the general MSSM framework while in mSUGRA the production of the bottom squark $(\tilde{b})$ is also considered. Work which is now underway shows that the results are valid also for $\tan\beta = 30$.

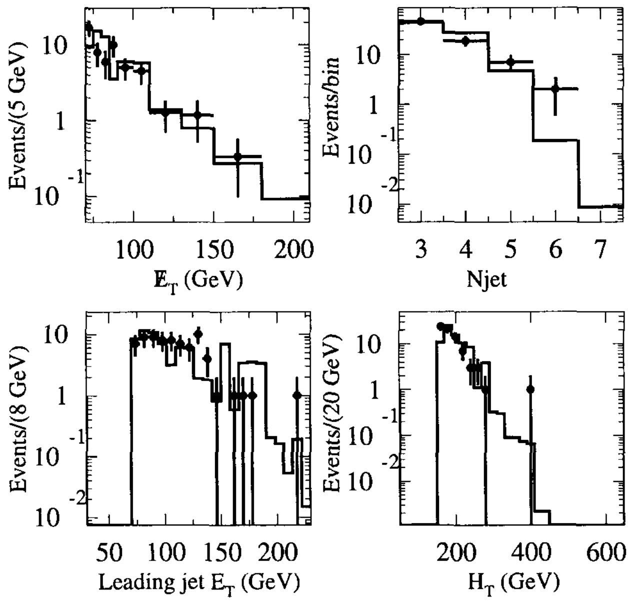


Figure 4. Comparison in the blind box between data (points) and Standard Model predictions (histogram) of $\not{E}_T$, N_{jet}, leading jet E_T and H_T distributions. There are 74 events in each of these plots, to be compared with 76±13 SM predicted events. Note that the $\not{E}_T$ distribution is plotted with a variable bin size; the bin contents are normalized as labelled.

Based on the observations, the Standard Model estimates and their uncertainties, and the relative total systematic uncertainty on the signal efficiency, we derive the 95% C.L. upper limit on the number of signal events. The bound is shown on the $m_{\tilde{q}} - m_{\tilde{g}}$ plane in Figure 3. For the signal points generated with mSUGRA, the limit is also interpreted in the $M_0 - M_{1/2}$ plane.[17]

4.1.1. *What did we learn?*

We learned that there is no excess of multijet events at the tails of the missing energy distribution in 84 pb^{-1} of data. We learned that we can simulate well the tails of the missing energy adequately for 84 pb^{-1} of data and predict the standard model backgrounds robustly. We learn that the

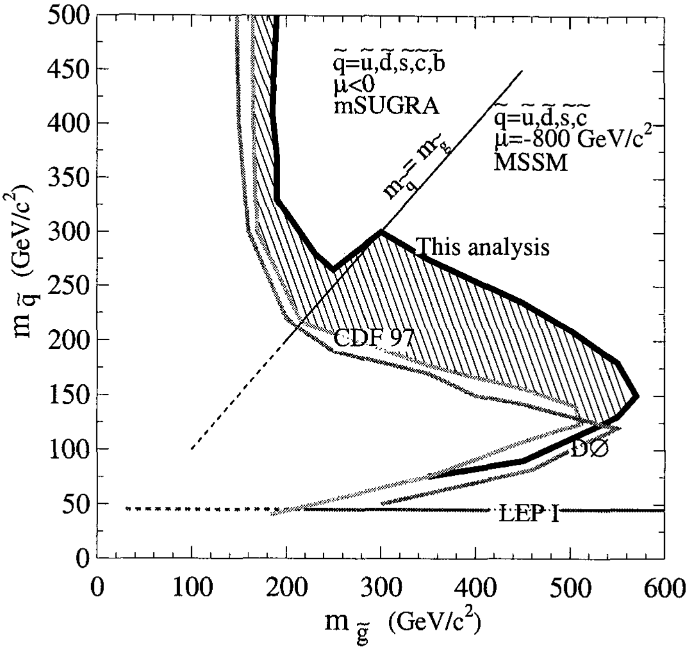


Figure 5. The 95% C.L. limit curve in the $m_{\tilde{q}}$ – $m_{\tilde{g}}$ plane for $\tan\beta = 3$; the hatched area is newly excluded by this analysis. Results from some previous searches are also shown.

LEP results are in accord with the Tevatron results when we both consider a minimal supersymmetric model. And we learn that the scale of SUSY can still be the electroweak scale. Lets discuss the latter a bit more.

The notion of naturalness and fine-tuning[18] is not extremely well defined, but it is used unavoidably since it is the fine-tuning in the Standard Model that motivates low-energy supersymmetry and supports the projection that superpartners should be found before or at the LHC. Many measures and studies of fine-tuning have appeared in the literature.[19–23]

In a model-independent analysis,[24] naturalness constraints are weak for some superpartners, *e.g.* the squarks and sleptons of the first two generations. In widely considered scenarios with scalar mass unification at a high scale, such as minimal supergravity, it is assumed that the squark and slepton masses must be $\lesssim$1 TeV/c^2. This bound places all scalar superpartners within the reach of present and near future colliders. This assumption is reexamined[23] in models with strong unification constraints, and the squarks and sleptons are found to be natural even with masses above 1 TeV/c^2. Furthermore by relaxing the universality constraints the naturalness upper limits on supersymmetric particles increase significantly[20] without extreme fine-tuning. This suppresses sparticle mediated rare processes and the problem of SUSY flavor violations is ameliorated. The fine-tuning due to the chargino mass is found to be model dependent.[22] With or without universality constraints the gluino remains below 400 GeV/c^2.[20] In fact, as it is pointed inRef. [22] the tightest constraints on fine-tuning come from the experimental limits on the lightest CP-even Higgs boson and the gluino for a number of supersymmetry models. It is then those two key particles that are within reach of the Tevatron collider. These results follow from the observation that fine-tuning is mainly dominated by M_3, the gluino mass parameter at the electroweak scale, and this dominant contribution can be partly canceled by negative contributions from other soft parameters, as can be seen from the expansion of the Z mass in terms of the input parameters (and for fixed $\tan\beta = 2.5$):[22]

$$M_Z^2 = -1.7\mu^2(0) \; + \; 7.2M_3^2(0) \; - 0.24M_2^2(0) \; + \; 0.014M_1^2(0) + ...$$

The required cancellation is easier if $M_3(0)(\sim m_{\tilde{g}})$ is not large (or alternatively if M_2 is increased for a given M_3). Using the results of this analysis on the gluino mass we get:

$$M_3 \gtrsim 300 \rightarrow \frac{7.2M_3^2}{M_Z^2} \gtrsim 80$$

A similar relation is derived using the LEP limits on the chargino $m_{\tilde{\chi}_1^+} \gtrsim$ 100 GeV/c^2[25,26,27] that points to the consideration of gaugino mass non-unification with a lighter gluino. The main effect of a relatively light gluino is the enhancement of the missing energy plus multijet signal with a lepton veto, since for a given chargino mass not yet excluded, the $\widetilde{g}\widetilde{g}$ cross section is enhanced and the gluino cascade decays through charginos are suppressed (fewer leptons are produced in the final state.[28,29])

4.1.2. *The* $\not\!\!E_T$ *trigger*

The $\not\!\!E_T$ trigger drives a number of analyses a few of which are the following:

- **Vector boson** production and leptonic decays. Although there is a dedicated $\not{E}_T$ plus lepton trigger for the study of the W boson, W QCD associated production remains a crucial background for a number of searches beyond the Standard Model and the $\not{E}_T$ plus jets trigger provides a good sample to study these processes. For Z production and decay, the $\not{E}_T$ sample provides a dataset to measure directly the $Z \to \nu\bar{\nu}$ + jets cross section. Furthermore again, the Z boson QCD associated production is a background to many searches.
- **top quark** production and decay to a $W + b$. The $\not{E}_T$ trigger provides an alternate dataset to measure the top cross section.
- **Associated Higgs-W and Higgs-Z production.** The $\not{E}_T$ combined with a b quark tagging or a tau lepton tagging trigger can provide a highly efficient triggering scheme for the discovery of the Higgs boson.
- **Beyond the Standard Model searches.** To mention a few, the $\not{E}_T$ trigger can be used to search for:
  - **Supersymmetric partners**: In R-Parity conserving supersymmetric scenarios the LSP (Lightest Supersymmetric Particle) escapes the detector and appears as energy imbalance. Examples are squark and gluino searches, scalar top and scalar bottom quark (utilizing an additional heavy flavor tag) searches. In gauge mediated supersymmetry breaking scenarios that incorporate gravity the LSP is the gravitino – the spin 3/2 partner of the graviton. The gravitino goes undetected and produces energy imbalance.
  - **Leptoquarks**: The leptoquark decays to a quark and a neutrino resulting in large $\not{E}_T$.
  - CHArged Massive Particles (**CHAMPS**): These are long-lived massive particles that if they are penetrating enough can go undetected and cause energy imbalance.
  - **Gravitons**: In braneworld theories of extra dimensions,[30,33] the graviton can be produced in high energy hadron collisions and escapes to the extra spatial dimensions resulting in energy imbalance.

The specifications of the $\not{E}_T$ trigger acquiring RUNII data was determined using the Run 1B data and the main feature of the trigger is the lower Level 2 $\not{E}_T$ threshold (25 GeV, compared to 35 GeV in Run 1B).

4.2. *Example 2: Monojets and missing energy: Looking for KK graviton emission*

By now you know exactly the steps of how to do this analysis: you need the clean $\not\!\!E_T$ triggered data sample and orthogonal data samples to normalize your background estimates, so that your uncertainties in the standard model predictions are dominated by the data statistics and not by systematics.

4.2.1. *On the theory*

It has been discussed since 1996[31] that weak scale superstrings have experimental consequences in collider physics. The missing energy as a signature was discussed already in Ref. 32 Consider the braneworld scenario where gravity propagates in the $4 + n$ dimensional bulk of spacetime, while the rest of the standard model fields are confined to the 3+1 dimensional brane. Assume compactification of the extra n dimensions on a torus with a common scale R, and identify the massive Kaluza-Klein (KK) states in the four-dimensional spacetime. In such a model, the Planck scale M_{Pl}, the compactification scale R, and the new effective Planck scale M_D, are related by the expression:

$$M_{Pl}^2 \sim R^n M_D^{2+n}$$

where n is the number of extra dimensions.

Searches for extra dimensions at colliders focus on the search for the emission of gravitons or the effects of the exchange of virtual gravitons (see J. Hewett, this school proceedings). Here we focus on a search for graviton emission.

There are three processes which can result in the emission of gravitons:

$$\begin{aligned} &q\bar{q} \to gG \\ &qg \to qG \\ &gg \to gG \end{aligned}$$

where G is the graviton. The Feynman diagrams for these processes are shown in Figure 6.

In all three cases the signature will be jets + $\not\!\!E_T$. We included in PYTHIA these processes using the large toroidal extra dimensions model of Arkani-Hamed, Dimopoulos, and Dvali,[33] and the calculations of Giudice, Rattazzi, and Wells.[34]

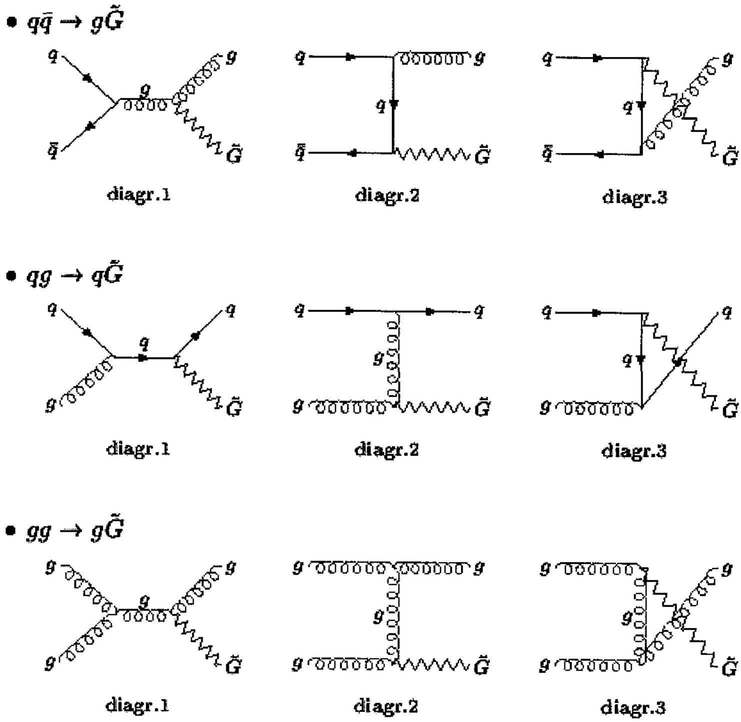


Figure 6. Feynman diagrams for the emission of real gravitons in $p\bar{p}$ collisions. Top: $q\bar{q} \to gG$. Middle: $qg \to qG$. Bottom: $gg \to gG$.

The differential cross-sections for the parton processes relevant to graviton plus jet production in hadron collisions are given in Equations 3, 4, and 5.

$$\frac{d^2\sigma}{dt\ dm}(q\bar{q} \to gG) = \frac{\alpha_s}{36}\ \frac{2\pi^{n/2}}{\Gamma(n/2)}\ \frac{1}{s{M_D}^{2+n}}\ m^{n-1}\ F_1(t/s, m^2/s), \quad (3)$$

$$\frac{d^2\sigma}{dt\ dm}(qg \to qG) = \frac{\alpha_s}{96}\ \frac{2\pi^{n/2}}{\Gamma(n/2)}\ \frac{1}{s{M_D}^{2+n}}\ m^{n-1}\ F_2(t/s, m^2/s), \quad (4)$$

$$\frac{d^2\sigma}{dt\ dm}(gg \to gG) = \frac{3\alpha_s}{16}\ \frac{2\pi^{n/2}}{\Gamma(n/2)}\ \frac{1}{s{M_D}^{2+n}}\ m^{n-1}\ F_3(t/s, m^2/s) \quad (5)$$

The Mandelstam variable t in Equations. 3-5 is defined as $t = (p_q - p_G)^2$.

The calculation of graviton emission is based on an effective low-energy theory that is valid at parton energies below the scale M_D.

The $F_i(x, y)$ functions in Equations 3-5 are

$$
\begin{aligned}
F_1(x,y) &= \frac{1}{x(y-1-x)}\left[-4x(1+x)(1+2x+2x^2)+\right.\\
&\quad \left. y(1+6x+18x^2+16x^3)-6y^2x(1+2x)+y^3(1+4x)\right], \quad (6)\\
F_2(x,y) &= -(y-1-x)\ F_1\left(\frac{x}{y-1-x}, \frac{y}{y-1-x}\right) =\\
&\quad \frac{1}{x(y-1-x)}\left[-4x(1+x^2)+y(1+x)(1+8x+x^2)\right.\\
&\quad \left.-3y^2(1+4x+x^2)+4y^3(1+x)-2y^4\right], \quad (7)\\
F_3(x,y) &= \quad \frac{1}{x(y-1-x)}\left[1+2x+3x^2+2x^3+x^4\right.\\
&\quad \left.-2y(1+x^3)+3y^2(1+x^2)-2y^3(1+x)+y^4\right]. \quad (8)
\end{aligned}
$$

The function $F_1(x, y)$ determines the cross-section for $f\bar{f} \to \gamma G$. $F_1(x, y)$ is invariant under exchange of the Mandelstam variables t and u; this is reflected in the property $F_1(x, y) = F_1(y - 1 - x, y)$. The same property holds also for the function $F_3(x, y)$, relevant to the QCD process.

Cross sections from PYTHIA for individual graviton production sub-processes are shown as a function of M_D for different values of n in Figures 7 and 8. Figure 7 compares the different sub-processes for the same value of n, while Figure 8 compares different values of n for the same sub-process.

Note that the cross-section for $q\bar{q} \to gG$ is larger for larger values of n, relative to the other sub-processes. This can be traced to the different dependences of F_1, F_2, and F_3 on m^2/s (labelled y in equations 6-8). F_2 and F_3 have a dependence on quartic dependence on y, whereas F_1 has only a cubic dependence. This results in larger splittings at high values of M_D between different values of n for $qg \to qG$ and $gg \to gG$ compared to $q\bar{q} \to gG$. For $M_D = 1$ TeV a significant number of heavy KK gravitons will be produced with masses averaging in the hundreds of GeV, as can be seen in Fig. 9. The peak of the mass distribution is higher for $n = 6$ extra dimensions, because the density of KK states is a more rapidly increasing function than for $n = 2$. This difference does not show up in the $\not{E}_T$ distribution (Fig. 9). This is due to two competing effects: (1) the heavier KK gravitons for $n = 6$ have larger transverse energy, but (2) the rapidly decreasing parton distribution functions cause the heavier gravitons to be

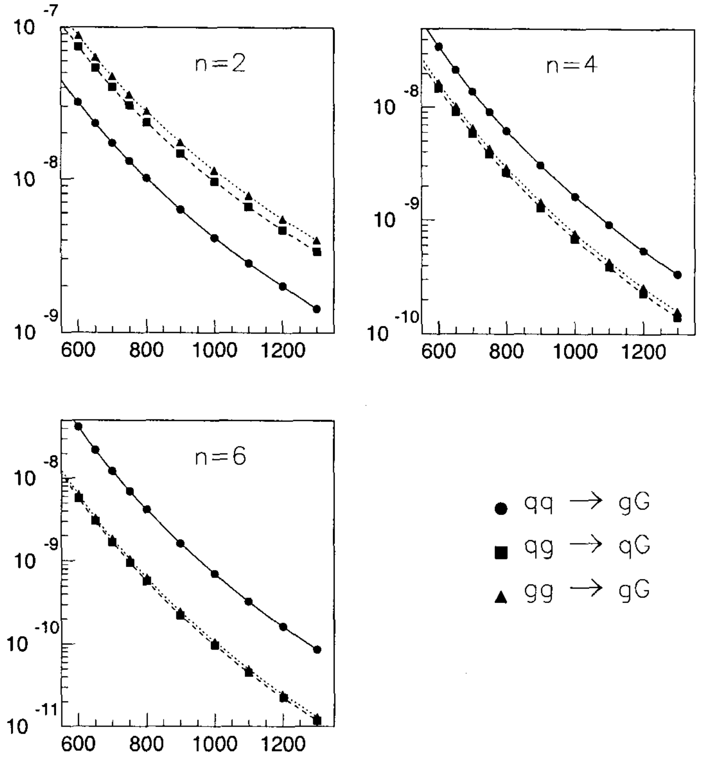


Figure 7. Cross sections for each subprocess for (a) n=2, (b) n=4, and (c) n=6.

produced near threshold. These two effects cancel, leaving nearly identical $\not\!\!E_T$ distributions for different values of n.

4.2.2. *The analysis*

The data preselection and the jet fiducial requirements are the same as in the case of the multijet plus $\not\!\!E_T$ analysis used in the search for gluinos and squarks and they are designed to retain a high purity in the real missing energy sample.

In the missing energy plus one jet search, the missing energy comes from the KK graviton tower and the one jet from the recoiling parton. To

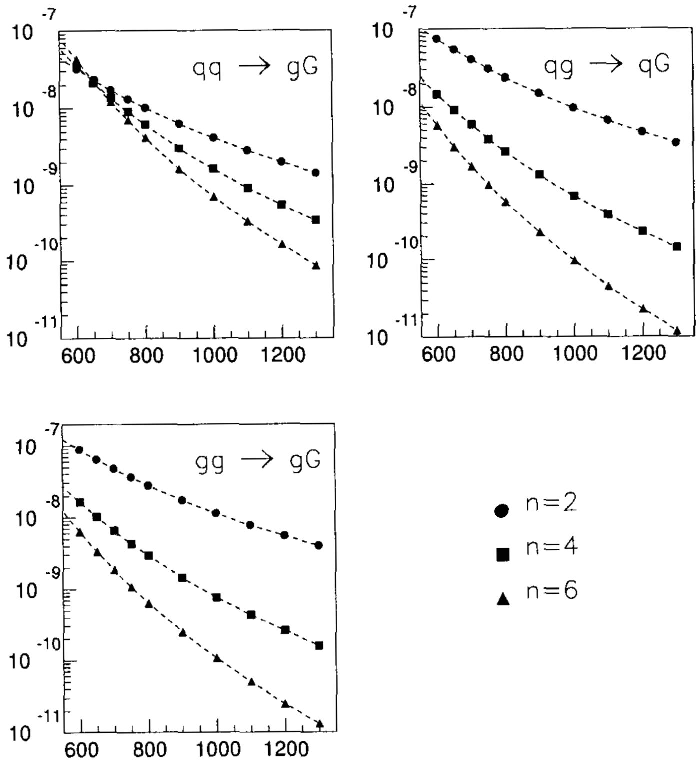


Figure 8. Cross sections for n=2, 4, and 6, for each subprocess: (a) $q\bar{q} \rightarrow gG$ (b) $qg \rightarrow qG$, and (c) $gg \rightarrow gG$.

reduce the background contribution from $W \rightarrow \ell\nu$+jets we apply the same *indirect lepton veto* as in the multijet analysis.

To define the signal region we use three variables: the $\not{E}_T$, the N_{jet} and the Isolated Track Multiplicity N_{trk}^{iso}. The requirements are, no isolated tracks, large missing energy and one or two jets (when two the second must not be grossly mismeasured).

The value of the missing energy requirement for the definition of the graviton signal region is driven by the missing energy Level 2 trigger efficiency and optimized for the graviton signal. The N_{jet} requirement is motivated by the graviton monojet signal characteristic final state. The

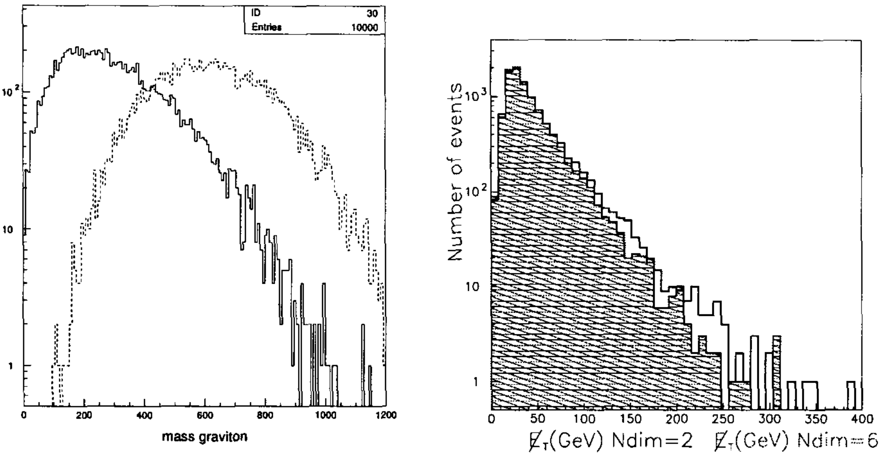


Figure 9. (l) The graviton mass distribution for $n = 2$ and $n = 6$ extra dimensions. (r) The $\not{E}_T$ distribution for n=2 and $n = 6$ extra dimensions(shaded) for $qg \to qG$ process simulation.

second jet is primarily allowed so that the QCD background can be calculated using the HERWIG Monte Carlo and the CDF detector simulation with a reliable normalization from the CDF QCD data, and additionally so that the systematic uncertainty of the signal due to ISR/FSR is kept as low as possible. Allowing a second jet also permits interpretation of the results with a k-factor inclusion in the signal cross section estimate.

The N_{trk}^{iso} requirement increases the sensitivity to the graviton signal by reducing the W+jet contribution, while at the same time retaining the signal. The rest of the analysis path is based on the kinematics and aims at high sensitivity for the signal.

4.2.3. *Summary of Standard Model processes with $\not{E}_T$ + jet in the final state*

The Standard Model processes with large missing energy and multijets in the final state that constitute backgrounds to the graviton search are

- Z+jets : QCD associated Z production with $Z \to \nu\nu$ is the most significant and irreducible background component. For the Z+jets backgrounds the PYTHIA Monte Carlo simulation is used and normalized to the $Z \to ee+ \geq 1$ jet data (our standardizable candle).

- W+jets : QCD associated W production with $W \rightarrow e\nu$, $W \rightarrow \mu\nu$, and $W \rightarrow \tau\nu$ has a large $\not{E}_T$ + jet contribution. For the W+jets and backgrounds the PYTHIA Monte Carlo simulation is used and normalized using the $Z \rightarrow ee+ \geq 1$ jet data sample.
- top, single top : In $t\bar{t}$ production a W from a top decay can decay semileptonicaly and contribute to the high $\not{E}_T$ tails. We use PYTHIA to simulate $t\bar{t}$ production with all inclusive top decays. For the normalization to the data luminosity the theoretical calculation $\sigma(t\bar{t}) = 5.1\ pb \pm 18\%$ (which is consistent with the CDF $t\bar{t}$[45] cross section) is used. HERWIG and PYTHIA Monte Carlo programs are used to simulate single top production via W-gluon fusion and W* production respectively. The theoretical cross sections $\sigma(tq') = 1.7$ pb $\pm 15\%$ and $\sigma(bt) = 0.73$ pb $\pm 9\%$[45] are used to normalize the samples.
- dibosons : For WW, WZ, ZZ production PYTHIA is used. For the normalization, the theoretical cross sections calculated for each diboson process are used: $\sigma(WW) = 9.5 \pm 0.7$ pb, $\sigma(WZ) = 2.6 \pm 0.3$ pb and $\sigma(ZZ) = 1 \pm 0.2$ pb[45] are used.
- QCD : The QCD background is generated with HERWIG and normalized to the CDF jet data using dijet events. Clearly one does not expect this to be a significant physics background.

Once the signal to background ratio is optimized as a function of the measured variables (kinematical,topological etc), the requirements are set and the data are compared to the standard model predictions. From there of course we can interprete the results as a limit in the allowable cross section of non-standard model processes, extra dimensions and what not. The result of this particular analysis has been reported in conferences and will appear soon in the literature but similar analyses from CDF and D0 have already reported results.[47]

5. What next?

It is a very interesting time for physics, indeed for fundamental physics. Recent experimental cosmology results are changing the ideas we had about how much of the universe we know and how well. Results in particle physics, such as the massiveness of the neutrinos, the tremendous precision at which the Standard Model holds in the up to now explored electroweak scale region (without the Higgs boson showing up), are all puzzling us. Putting together a picture of how spacetime happens and how it is filled, by means of observations and studies in a controlled experimental environment, is the primary and urgent work of both experimentalists and theorists. In these lectures I gave examples of how we go about this using colliders.

Many thanks to the school organizers Joe Lykken, Steve Gubser, and Kalyana T. Mahanthappa; the TASI 2001 students; Kevin Burkett and all my CDF collaborators always. The author is supported by NSF and the Pritzker Foundation.

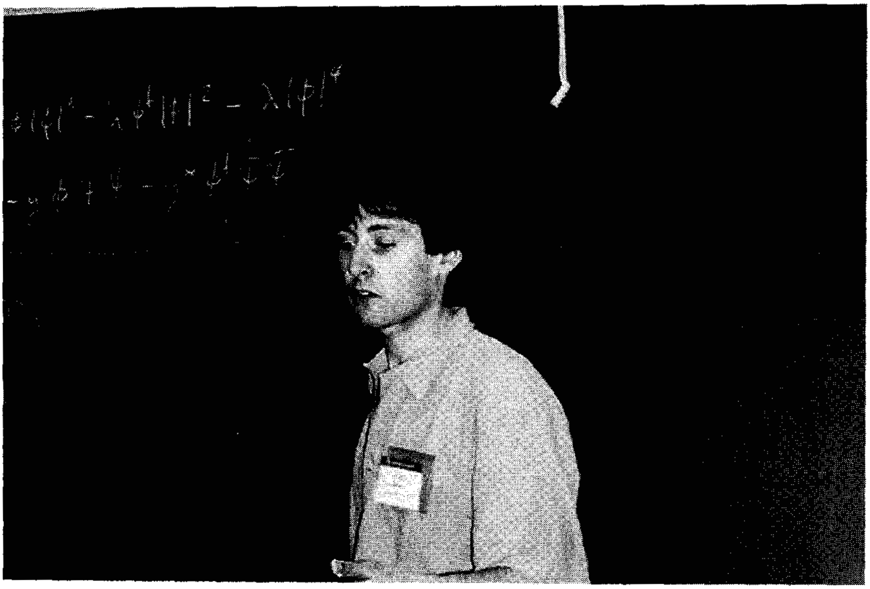
References

1. *Medical Accelerators in Radiotherapy: Past Present and Future*, Physics Medica, Vol. XII N.4, (1996).
2. U. Amaldi,
http://web.cern.ch/AccelConf/e96/PAPERS/ORALS/FRY01A.PDF]
3. APS News, Backpage, December 2002
http://www.aps.org/apsnews/1202/120213.htm
4. NLC Design and Physics Working Group, FERMILAB-PUB-96/112, 1996.
5. J. Polchinski, arXiv:hep-th/9607050 (1996).
6. http://hep.uchicago.edu/cdf/smaria/ms/aaas03.html
7. e.g. Pierre Ramond, Phys. Rev. **D3**, 2415, (1971).
8. G. Kane, *Supersymmetry*, Perseus Books, (2000)
9. P. Drell, hep-ex/9701001, (1997).
10. R. Markeloff, University of Winsconsin, Ph.D. thesis, (1994).
11. The D0 Collaboration, Phys. Rev. Lett. **79**, 1203, (1997).
12. T. Affolder *et al.*, The CDF Collaboration, to be submitted.
13. T. Affolder *et al.*, The CDF Collaboration, Phys. Rev. Lett. **85**, 5704, (2000).
14. T. Affolder *et al.*, The CDF Collaboration, Phys. Rev. Lett. **85**, 1378 (2000).
15. T. Affolder *et al.*, The CDF Collaboration, FERMILAB-PUB-00/073-E, submitted to Phys. Rev. Lett. March 31, (2000).
16. N_{trk}^{iso} is the number of high momentum isolated tracks in the event. Tracks qualify as such if they have transverse momentum $P_T \geq 10$ GeV/c, impact parameter $d_0 \leq 0.5$ cm, vertex difference $|z_{track} - z_{event}| < 5$ cm and the total transverse momentum ΣP_T of all tracks (with impact parameter $d_0^i \leq 1$ cm) around them in a cone of $\Delta R \equiv \sqrt{\Delta\eta^2 + \Delta\phi^2} = 0.4$ is $\Sigma P_T \leq 2$ GeV/c.
17. M. Spiropulu, Harvard University Ph.D thesis (2000); T. Affolder *et al.* The CDF Collaboration,Phys. Rev. Lett. **88**, 241802 (2002).
18. E. Gildener, Phys. Rev. **D14**, 1667, (1976);
E.Gildener and S. Weinberg, Phys. Rev. **D13**, 3333, (1976);
L. Maiani, Proceedings of the Summer School of Gif-Sur-Yvette (1980);
M. Veltman, Acta Phys. Polon. **B12**, 437, (1981);
S. Dimopoulos and S. Raby, Nucl. Phys. **B192**, 353, (1981);
E. Witten, Nucl. Phys. **B188**, 513, (1981);
S. Dimopoulos and H. Georgi Nucl. Phys. **B193**, 150, (1981).
19. R. Barbieri and G.F. Giudice, Nucl. Phys. **B306** 63, (1988).
20. S. Dimopoulos and G.F. Giudice, Phys. Lett. **B357**, 573, (1995).

21. G. Anderson and D. Castano, Phys.Rev. D52, 1693, (1995).
22. M. Bastero-Gil, G. L. Kane and S. F. King Phys. Lett. B **474**, 103 (2000); G. L. Kane and S .F. King, Phys. Lett. **B451** 113, (1999), and references therein.
23. J.L. Feng, K.T. Matchev and T. Moroi, Proceedings of PASCOS'99, Lake Tahoe, December 10-16, 1999;
J.L. Feng, K.T. Matchev and T. Moroi, Phys. Rev. Lett. **84** 2322, (2000);
J. Bagger, J. Feng, N. Polonsky, Nucl. Phys. **B563**, 3, (1999).
24. S. Coleman and D.J. Gross, Phys. Rev. Lett. **31**, 851, (1973).
25. S. Rosier-Lees, *LEP SUSY Working Group Status Report*, LEPC, http://lepsusy.web.cern.ch/lepsusy/, (2000).
26. S. Martin, DPF, http://www.dpf2000.org/BSM1.htm#ses1 (2000).
27. G. L. Kane, J. Lykken, B. D. Nelson and L. T. Wang, Phys. Lett. B **551**, 146 (2003).
28. V. Barger, C. E. M. Wagner, *et al.* hep-ph/0003154, (2000).
29. G. Anderson, H. Baer, C. Chen and X. Tata, Phys. Rev. **D61**, 095005, (2000)
30. G. Nordstrom, Z. Phys. **15**, 504 (1914); T. Kaluza, Preuss. Akad. Wiss, Berlin, Math. Phys. K **1**, 966 (1921); O. Klein Z. Phys. **37** 895; Nature **118**, 516 (1926).
31. J. D. Lykken, Phys. Rev. D **54**, 3693 (1996)
32. I. Antoniadis, N. Arkani-Hamed, S. Dimopoulos and G. R. Dvali, Phys. Lett. B **436**, 257 (1998)
33. N. Arkani-Hamed, S. Dimopoulos and G. R. Dvali, Phys. Lett. B **429**, 263 (1998)
34. G. F. Giudice, R. Rattazzi and J. D. Wells, Nucl. Phys. B **544**, 3 (1999) [arXiv:hep-ph/9811291].
35. T. Han, J. D. Lykken and R. J. Zhang, Phys. Rev. D **59**, 105006 (1999) [arXiv:hep-ph/9811350].
36. For a review, see J. Hewett and M. Spiropulu Annu. Rev. Nucl. Part. Sci., Vol. **52**: 397-424 (2002).
37. T. Sjöstrand, Comput. Phys. Commun. **82**, 74 (1994). PYTHIA V6.115 is used.
38. F. Abe *et al.*, Nucl. Inst. and Methods, **A271**, 387 (1988).
39. In the CDF coordinate system, ϕ and θ are the azimuthal and polar angles with respect to the proton beam direction. The pseudorapidity η is defined as $-ln[\tan(\theta/2)]$.
40. F. Abe *et al.*, The CDF Collaboration, Phys. Rev. **D45**, 1448 (1992).
41. If there are multiple vertices in the event we use the vertex with the largest $\sum P_T$ of associated tracks.
42. T. Affolder *et al.* The CDF Collaboration, Phys. Rev. Lett. **88**, 241802 (2002).
43. N_{trk}^{iso} is the number of high momentum isolated tracks in the event. Tracks qualify as such if they have transverse momentum $P_T \geq 10$ GeV/c, impact parameter $d_0 \leq 0.5$ cm, vertex difference $|z_{track} - z_{event}| < 5$ cm and the total transverse momentum ΣP_T of all tracks (with impact parameter $d_0^i \leq$ 1 cm) around them in a cone of $\Delta R \equiv \sqrt{\Delta\eta^2 + \text{it}\Delta\phi^2} = 0.4$ is $\Sigma P_T \leq$ 2 GeV/c.
44. G. Marchesini *et al.*, Comput. Phys. Commun. **67**, 465 (1992). HERWIG v5.6

is used. See hep-ph/9607393 (1996).

45. R. Bonciani *et al.*, Nucl. Phys. **B529**, 424 (1998); T. Stelzer, Z. Sullivan and S. Willenbrock, Phys. Rev. **D54**, 6696 (1996); T. Tait and C. P. Yuan, hep-ph/9710372 (1997); J. Ohnemus *et al.*, Phys. Rev. **D43**, 3626 (1991); Phys. Rev. **D44**, 1403 and 3477 (1991). The measured top cross section (T. Affolder *et al.*, Phys. Rev. D64 032002 (2001)) is in agreement with the theoretical prediction.
46. G. Zech, Nucl. Instrum. Methods **A277**, 608 (1989); T.Huber *et al* ., Phys Rev **D41**, 2709 (1990).
47. Recently the D0 collaboration (arXiv:hep-ex/0302014) reported their limit on direct graviton emission using a K-factor of 1.3 in the signal cross-section; D. Acosta *et al.*, The CDF Collaboration, Phys. Rev. Lett. **89** 281801 (2002); For a summary of LEP results on graviton emission, see G. Landsberg, arXiv:hep-ex/0105039 (2000).



MATTHEW J. STRASSLER

This page intentionally left blank

AN UNORTHODOX INTRODUCTION TO SUPERSYMMETRIC GAUGE THEORY

MATTHEW J. STRASSLER

Department of Physics
University of Washington
Box 351560
Seattle, WA 98195
E-mail: strassler@phys.washington.edu

Numerous topics in three and four dimensional supersymmetric gauge theories are covered. The organizing principle in this presentation is scaling (Wilsonian renormalization group flow.) A brief introduction to scaling and to supersymmetric field theory, with examples, is followed by discussions of nonrenormalization theorems and beta functions. Abelian gauge theories are discussed in some detail, with special focus on three-dimensional versions of supersymmetric QED, which exhibit solitons, dimensional antitransmutation, duality, and other interesting phenomena. Many of the same features are seen in four-dimensional non-abelian gauge theories, which are discussed in the final sections. These notes are based on lectures given at TASI 2001.

1. Introduction

These lectures are the briefest possible introduction to some important physical ideas in supersymmetric field theory.

I have specifically avoided restating what the textbooks already contain, and instead have sought to provide a unique view of the subject. The usual presentation on superfields is sidestepped; no introduction to the supersymmetric algebra is given; superconformal invariance is used but not explained carefully. Instead the focus is on the renormalization group, and the special and not-so-special qualitative features that it displays in supersymmetric theories. This is done by discussing the often-ignored classical renormalization group (the best way to introduce beta functions, in my opinion) which is then generalized to include quantum corrections. This approach is most effective using theories in both three and four dimensions. Initially, models with only scalars and fermions are studied, and the classic nonrenormalization theorem is presented. Then I turn to abelian gauge theories, and finally non-abelian gauge theories. Fixed points, unitarity theorems, duality, exactly marginal operators, and a few other amusing concepts surface along the way. Enormously important subjects are left out, meriting only a sad mention in my conclusions or a brief discussion in the appendix.

Clearly these lectures are an introduction to many things and a proper summary of none. I have tried to avoid being too cryptic, and in some sections I feel I have failed. I hope that the reader can still make something useful of the offending passages. I have also completely failed to compile a decent bibliography. My apologies.

My final advice before beginning: *Go with the flow.* If you aren't sure why, read the lectures.

2. Classical theory

2.1. *Free massless fields*

Consider classical free scalar and spinor fields in d space-time dimensions.

$$S_\phi = \int d^dx \; \partial_\mu \phi^\dagger \partial^\mu \phi \tag{1}$$

$$S_\psi = \int d^dx \; i\bar{\psi}\partial\!\!\!/\psi \tag{2}$$

By simple dimension counting, since space-time coordinates have mass dimension -1 and space-time derivatives have mass dimension $+1$, the dimensions of these free scalar and spinor fields are

$$\dim \phi = \frac{d-2}{2} \quad ; \quad \dim \psi = \frac{d-1}{2}$$

so in $d = 3$ scalars (fermions) have dimension $1/2$ (1) while in $d = 4$ they have dimension 1 ($\frac{3}{2}$). These free theories are scale-invariant — for any $s > 0$,

$$x \to sx \; ; \phi \to s^{-(d-2)/2}\phi \; ; \psi \to s^{-(d-1)/2}\psi \;\Rightarrow S_\phi \to S_\phi \; ; \; S_\psi \to S_\psi \; .$$

Less obviously, they are conformally invariant. (If you don't already know anything about conformal invariance, don't worry; for the purposes of these lectures you can just think about scale invariance, and can separately study conformal symmetry at another time.)

2.2. *Free massive fields*

Now let's consider adding some mass terms.

$$S = \int d^dx \ \left[\partial_\mu \phi^\dagger \partial^\mu \phi - m^2 \phi^\dagger \phi\right] \tag{3}$$

Scalar masses always have dimension dim $m = 1$ for all d. The propagator

$$\langle \phi(k)\phi(-k)\rangle = \frac{i}{k^2 - m^2}$$

is a power law $1/|x-y|^{d-2}$ (the Fourier transform of the propagator i/k^2) for $|x-y| \ll m^{-1}$, and acts like a delta function $\delta^4(x)$ (the Fourier transform of the propagator $-i/m^2$) at $|x-y| \gg m^{-1}$. The scale-dependent propagator interpolates between these two scale-invariant limits.

Exercise: Compute the propagator in position space and show that it does indeed interpolate between two scale-invariant limits.

What is the renormalization group flow associated with this theory? It can't be completely trivial even though the theory is free and therefore soluble. The number of degrees of freedom is two (one complex scalar equals two real scalars) in the ultraviolet and zero in the infrared, so obviously the theory is scale dependent. In particular, Green's functions are not pure power laws, as we just saw for the propagator. But how are we to properly discuss this given that the one coupling constant in the theory, m, does not itself change with scale? The approach we will use is to define a *dimensionless* coupling $\nu^2 \equiv m^2/\mu^2$ where μ is the renormalization-group scale – the scale at which we observe the theory. We can think of this theory as transitioning, as in Fig. 1, between two even simpler theories: the scale-invariant theory at $\mu \to \infty$ where the mass of ϕ is negligible and $\nu \to 0$, and the empty though scale-invariant $\nu \to \infty$ theory in the infrared, where the scalar does not propagate.

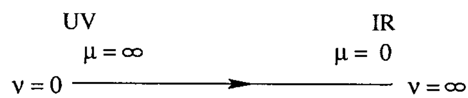


Figure 1. The effect of a mass term grows in the infrared.

Now consider a massive fermion.

$$S = \int d^dx \; \left[i\bar{\psi}\partial\!\!\!/ \psi - m\psi\psi \right] \tag{4}$$

Again, fermion masses always have dimension dim $m = 1$ for all d.[a] Better yet, consider a theory with two free fermions.

$$S = \int d^dx \; \sum_{n=1}^{2} \left[i\bar{\psi}_n \partial\!\!\!/ \psi_n - m_n \psi_n \psi_n \right] \tag{5}$$

Now we have a dimensionless coupling constant $\rho = m_1/m_2$.

Consider the classical scaling behavior (renormalization group flow) shown in Fig. 2 below. Note there are four scale-invariant field theories in this picture: one with two massless fermions, two with one massless and one infinitely massive fermion, and one with no matter content (indeed, no content at all!) We will call such scale-invariant theories "conformal fixed points", or simply "fixed points", to indicate that the dimensionless couplings of the theory, if placed exactly at such a point, do not change with scale.

Whatever are m_1 and m_2, or for our purposes the dimensionless couplings ν_1 and ν_2, scale transformations take the theory from the $\nu_1 = \nu_2 = 0$ conformal fixed point to one of the other fixed points. The parameter $\rho \equiv \nu_1/\nu_2$, which is scale-invariant, parameterizes the "flows" shown in the graph. The arrows indicate the change in the theory as one considers it at larger and larger length scales. In addition to flows which actually end at $\nu_1 = 0, \nu_2 = \infty$, note there are also interesting flows from $\nu_1 = \nu_2 = 0$ to $\nu_1 = \nu_2 = \infty$ which pass *arbitrarily* close to the fixed point $\nu_1 = 0, \nu_2 = \infty$. These can remain close to the intermediate fixed point for an *arbitrarily* large range of energy, (namely between the scales $\mu = m_2$

[a] Here and throughout I am using the notation of Wess and Bagger. In particular, ψ is a *two-component* complex left-handed Weyl fermion, with a spinor index $a = 1, 2$; $\bar{\psi}$, its conjugate, a right-handed Weyl anti-fermion, has a conjugate spinor index $\dot{a} = 1, 2$ which cannot be contracted with a! A kinetic term $\bar{\psi}^{\dot{a}} \sigma^{\mu}_{\dot{a}a} \partial_\mu \psi^a$ ($\sigma^0 = i\delta_{\dot{a}a}$, $\sigma^{1,2,3}$ are Pauli matrices) simply translates a left-handed Weyl fermion; it conserves a fermion number symmetry (the symmetry $\psi \to \psi e^{i\alpha}$). A so-called "Majorana" mass term $\psi^a \epsilon_{ab} \psi^b$ ($\epsilon_{11} = \epsilon_{22} = 0$, $\epsilon_{12} = 1 = -\epsilon_{21}$) breaks fermion number and mixes the particle with its antiparticle. It is very possible that left-handed neutrinos have such mass terms. Note Majorana mass terms can only be written for particles which are gauge-neutral; for instance, electrons cannot have them as long as QED is an unbroken gauge symmetry. Instead, an electron mass connects the two-component charge 1 left-handed electron ψ to the conjugate of the two-component charge -1 left-handed positron χ, via a coupling $\epsilon_{ab} \chi^a \psi^b$.

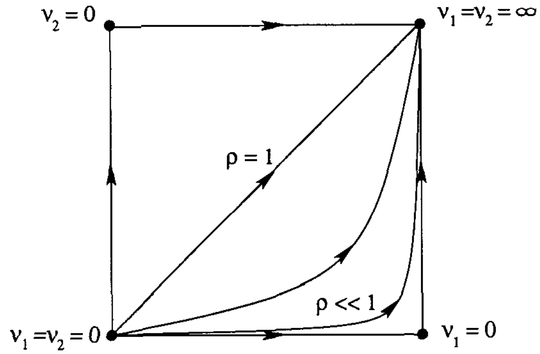


Figure 2. The scaling flow of two masses, with $\rho = m_1/m_2 = \nu_1/\nu_2$.

and $\mu = m_1 = \rho m_2$) although this is not obvious from the graph of the flow; it is something one must keep separately in mind.

A mass term is known as a "relevant" operator, where the relevance in question is *at long distances* (low energies.) Although the mass term has no effect in the ultraviolet — at short distance — it dominates the infrared (in this case by removing degrees of freedom.) We can see this from the fact that the dimensionless coupling ν grows as we scale from the ultraviolet toward the infrared. In fact we can define a beta function for $\nu = m/\mu$ as follows:

$$\beta_\nu \equiv \mu \frac{\partial \nu}{\partial \mu} = -\nu \tag{6}$$

That ν grows in the infrared is indicated by the negative beta function. More specifically, the fact that the coefficient is -1 indicates that ν scales like $1/\mu$. This tells us that the mass m has dimension 1. We will see examples of irrelevant operators shortly.

2.3. *Supersymmetry! The Wess-Zumino model*

Now let's add some "interactions" (more precisely, let's consider non-quadratic theories.)

$$S_{\text{Yukawa}} = \int d^d x \; \left[\partial_\mu \phi^\dagger \partial^\mu \phi + i\bar{\psi}\partial\!\!\!/\psi - y\phi\psi\psi - y^* \phi^\dagger \bar{\psi}\bar{\psi} - \lambda|\phi^\dagger \phi|^2\right] \tag{7}$$

The coefficients λ and y are dimensionful for $d < 4$ but dimensionless for $d = 4$. Appropriate dimensionless coefficients are $\lambda\mu^{d-4}$ and $y\mu^{(d-4)/2}$, with classical beta functions

$$\beta_{\lambda\mu^{d-4}} = (d-4)\lambda\mu^{d-4} \; ; \; \beta_{y\mu^{(d-4)/2}} = \frac{1}{2}(d-4)y\mu^{(d-4)/2}$$

showing the interactions are classically relevant for $d < 4$ (and irrelevant for $d > 4$!) but are "marginal" for $d = 4$ — they do not scale. The theories in question are thus classically scale invariant for $d = 4$ (and in fact conformally invariant.)

This Yukawa-type theory can be conveniently written in the following form

$$\begin{aligned} S_{\text{Yukawa}} &= S_{kin} + S_{int} + S_{int}^{\dagger} \\ S_{kin} &= \int d^d x \; \left[\partial_\mu \phi^\dagger \partial^\mu \phi + i\bar{\psi}\partial\!\!\!/\psi + F^\dagger F\right] \\ S_{int} &= \int d^d x \; \left[-y\phi\psi\psi + h\phi^2 F\right] \end{aligned} \tag{8}$$

Here F is just another complex scalar field, except for one thing – it has no ordinary kinetic term, just a wrong-sign mass term. This looks sick at first, but it isn't. F is what is known as an "auxiliary field", introduced simply to induce the $|\phi^\dagger\phi|^2$ interaction in (7). The classical equation of motion for F is simply $F^* \propto \phi^2$.

We can break the conformal invariance of the $d = 4$ theory by adding mass terms and cubic scalar terms:

$$S_{\text{Yukawa}} \to S_{\text{Yukawa}} - \int d^d x \; \left[M^2\phi^\dagger\phi + \frac{1}{2}m\psi\psi + h^*\phi|\phi^2| + h\phi^\dagger|\phi|^2\right] . \tag{9}$$

If M, h and λ are related so that the scalar potential is a perfect square, then we may write this in the form (8),

$$S_{int} = \int d^d x \; \left[-\frac{1}{2}m\psi\psi - y\phi\psi\psi + (M\phi + s\phi^2)F\right] \tag{10}$$

where s is a constant, and with S_{kin} as before.

When y and s are equal, and M and m are equal, the theory has supersymmetry. Given any spinor ζ, the transformations

$$\delta\phi = \sqrt{2}\zeta\psi \; ; \delta\psi_a = i\sqrt{2}\sigma^\mu_{a\dot{a}}\bar{\zeta}^{\dot{a}}\partial_\mu\phi + \sqrt{2}\zeta_a F ; \delta F = i\sqrt{2}\bar{\zeta}^{\dot{a}}\partial\!\!\!/_{\dot{a}a}\psi^a \tag{11}$$

change the Lagrangian only by a total derivative, which integrates to nothing in the action. I leave it to you to check this. This supersymmetric theory is called the "Wess-Zumino model"; it dates to 1974.

At this point there is a huge amount of supersymmetric superfield and superspace formalism one can introduce. There are lots of books on this subject and many good review articles. You don't need me to write another (and many of you have already read one or more of them) so I will not cover this at all. Instead I will state without proof how one may construct supersymmetric theories in a simple way. Those of you who haven't seen this before can take this on faith and learn it later. Those of you who have seen it will recognize it quickly.

Let us define a chiral multiplet, which we will represent using something we will call a chiral superfield Φ. Φ contains three "component" fields: ϕ, ψ, F. Note that Φ is complex: ϕ is a complex scalar (with two real components), ψ is a Weyl fermion (two complex components, reduced to one by the Dirac equation) and F is a complex auxiliary field (no propagating components.) Thus before accounting for the equations of motion there are four real bosonic and four real fermionic degrees of freedom; after using the equations of motion, there are two real bosonic propagating degrees of freedom and two fermionic ones. That the numbers of bosonic and fermionic degrees of freedom are equal is a requirement (outside two dimensions) for any supersymmetric theory.

Since Φ is complex we can distinguish holomorphic functions $W(\Phi)$, and antiholomorphic functions, from general functions $K(\Phi, \Phi^\dagger)$. As we will see, this fact is the essential feature which explains why we know so much more about supersymmetric theories than nonsupersymmetric ones — the difference is the power of complex analysis compared with real analysis.

Consider any holomorphic function (for now let it be polynomial) $W(\Phi)$. A well-behaved supersymmetric classical field theory can be written in the form of Eqs. (7)-(10) with

$$S_{int} = \int d^d x \left[-\frac{1}{2}\frac{\partial^2 W(\phi)}{\partial \phi^2}\psi\psi + \frac{\partial W(\phi)}{\partial \phi} F \right] \tag{12}$$

(In this expression the function $W(\Phi)$ is evaluated at $\Phi = \phi$.) The function W is called the "superpotential." (Note that dim $W = d - 1$.)

<u>Exercise:</u> Check that we recover the previous case by taking $W = \frac{1}{2}m\Phi^2 + \frac{1}{3}y\Phi^3$.

The potential for the scalar field ϕ is always

$$V(\phi) = |F|^2 = \left| \frac{\partial W(\phi)}{\partial \phi} \right|^2 \tag{13}$$

Notice this is positive or zero for all values of ϕ.

It is a theorem that supersymmetry is broken if $\langle F \rangle \neq 0$; you can see this from the transformation laws (11), in which F appears explicitly. (By contrast, $\langle \phi \rangle$ does not appear in the transformation laws, so it may be nonzero without breaking supersymmetry automatically.) From (13) we see that $V(\phi) = 0$ in a supersymmetry-preserving vacuum. This also follows from the fact that the square of one of the supersymmetry generators is the Hamiltonian; if the Hamiltonian, acting on the vacuum, does not vanish, than the supersymmetry transformation will not leave the vacuum invariant, and the vacuum thus will not preserve the supersymmetry generator.

To find the vacua of a supersymmetric theory is therefore much easier[b] than in a nonsupersymmetric theory. In the nonsupersymmetric case we must find all ϕ for which $\frac{\partial V}{\partial \phi} = 0$, and check which of these extrema is a local or global minimum. In the supersymmetric case, we need only find those Φ for which $\frac{\partial W}{\partial \Phi} = 0$; we are guaranteed that any solution of these equations has $V = 0$ and is therefore a supersymmetric, global minimum (although it may be one of many).

For example, if we take the theory $W = \frac{1}{3}y\Phi^3$, then $V(\phi) = |\phi^2|^2$; we can see it has only one supersymmetric minimum, at $\phi = 0$.

<u>Exercise:</u> Find the scalar potential and the supersymmetric minima for $W = \frac{1}{3}y\Phi^3 + \frac{1}{2}m\Phi^2$, $W = \frac{1}{3}y\Phi^3 + \xi\Phi$, $W = \frac{1}{3}y\Phi^3 + c$, where ξ and c are constants. In each case, note and carefully interpret what happens as y goes to zero.

2.4. *The XYZ model*

Let us consider a theory with three chiral superfields X, Y, Z with scalars x, y, z, and a superpotential $W = hXYZ$. Then the potential $V(x, y, z) = |h^2|(|xy|^2 + |xz|^2 + |yz|^2)$ has minima whenever two of the three fields are zero. In other words, there are three complex planes worth of vacua (any x if $y = z = 0$, any y if $z = x = 0$, or any z if $x = y = 0$) which intersect at the point $x = y = z = 0$. This rather elaborate space of degenerate vacua — noncompact, continuous, and consisting of three intersecting branches — is called a "moduli space". The massless complex fields whose expectation values parameterize the vacua in question (x on the X-branch, *etc.*) are

[b] Usually! we will return to a subtlety later.

called "moduli". It is useful to represent the three complex planes as three intersecting cones, as in Fig. 3. The point where the cones intersect, and all fields have vanishing expectation values, is called the "origin of moduli space."

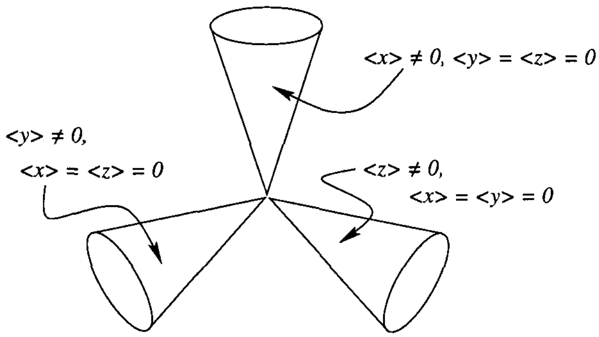


Figure 3. The moduli space of the XYZ model.

Exercise: Prove that these vacua are not all physically equivalent (easy — compute the masses of the various fields.) Note that there are extra massless fields at the singular point $x = y = z = 0$ where the three branches intersect. Then look at the symmetries — discrete and continuous — of the theory and determine which of the vacua are isomorphic to one another.

2.5. *XYZ with a mass*

Let's quickly consider what happens in the theory $W = hXYZ + \frac{1}{2}mX^2$. Then the Y and Z branches remain while the X branch is removed, as in Fig. 4. It is interesting to consider the classical renormalization group flow. Let's think of this in $d = 4$; then the theory with $m = 0$ is classically conformal, with h a marginal coupling, and m a relevant one.

What is happening in the far infrared? The theory satisfies the equation

$$\nabla^2 X^\dagger = m^\dagger(mX + hYZ)$$

which for momenta much lower than m (where $\nabla^2 X \ll |m^2|X)$ simply beomes $mX = -hYZ$. In this limit the kinetic term for X plays no role, and X acts like an auxiliary field. We may therefore substitute its equation

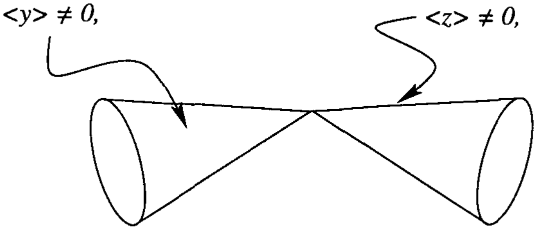


Figure 4. The moduli space once a mass for X is added.

of motion back into the Lagrangian, obtaining a low-energy effective theory for Y and Z with superpotential

$$W_L(Y,Z) = \kappa Y^2 Z^2 \quad ; \quad \kappa = -h^2/2m$$

At the origin of moduli space, where $\langle y \rangle = 0 = \langle z \rangle$, Y and Z are massless.

The superpotential Y^2Z^2 leads to interactions such as $\kappa^2|y^2z|^2$ and $\kappa y^2\psi_z\psi_z$ which have dimension higher than 4; equivalently, κ has negative mass dimension -1. Defining a dimensionless quantity $k = \kappa\mu = -h^2/2\nu$, we see it has a *positive* beta function ($\beta_k = 2\beta_h - \beta_\nu = +k$) so it becomes unimportant in the infrared. We therefore call κ, or k, an "irrelevant coupling", and Y^2Z^2 an "irrelevant operator." More precisely, at the origin of moduli space there is a fixed point in the infrared at which the massless chiral superfields Y, Z have $W(Y,Z) = 0$ (since the physical coupling $k \to 0$ in the infrared.) We say that κ (or k) is an irrelevant coupling, and Y^2Z^2 is an irrelevant operator[c] *with respect to this infrared fixed point.* The renormalization group flow for $\langle y \rangle = \langle z \rangle = 0$ is shown in Fig. 5.

Thus, we may think of this field theory as a flow from a (classically) conformal fixed point with $W(X,Y,Z) = hXYZ$ — one of a continuous class of fixed points with coupling h — to the isolated conformal field theory with $W(Y,Z) = 0$. In this flow, the mass term X^2 acts as a relevant operator on the ultraviolet fixed point, causing the flow to begin, and the flow into the infrared fixed point occurs along the direction given by the irrelevant operator $(YZ)^2$.

However, this description is only correct if $\langle y \rangle = \langle z \rangle = 0$, that is, exactly

[c]Still more precisely, the irrelevant operators are those which appear in the Lagrangian; it is not W but $|\partial W/\partial Y|^2$ which has dimension higher than 4. This shorthand is a convenient but sometimes confusing abuse of language. The assertion that κ is an irrelevant coupling is not subject to this ambiguity.

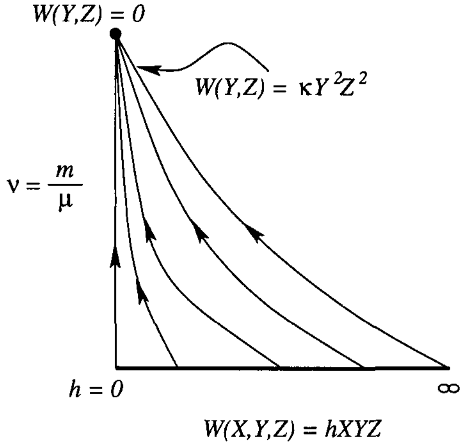


Figure 5. The flow at the origin of moduli space; the effect of the Y^2Z^2 superpotential renormalizes away in the far infrared.

at the origin of moduli space. Suppose $\langle y \rangle$ is not zero; then Z is massive at the scale hy^2 and we must re-analyze the flow below this scale. The far-low-energy theory in this case would have only Y in it. Here lies a key subtlety. The existence of the fixed point with *vanishing* superpotential $W(Y,Z)$ at the origin of moduli space might seem, naively, to imply that somehow Z would remain massless even when $\langle y \rangle \neq 0$. But this is inconsistent with the original classical analysis, using either the $XYZ + \frac{1}{2}mX^2$ superpotential or the kY^2Z^2 superpotential. How is this contradiction resolved? To understand this better, consider more carefully the effect of the nonvanishing superpotential for finite $\langle y \rangle$. At momenta $\mu \gg \kappa\langle y^2 \rangle$ the theory has light fields Y, Z, as does the conformal point, but at momenta $\mu \ll \kappa\langle y^2 \rangle$ only Y is light; see Fig. 6. *Thus the limits* $\mu \to 0$ *and* $\langle y \rangle \to 0$ *do not commute!* This is one way to see that the infrared conformal theory at the origin of moduli space is effectively disconnected from the infrared theory at $\langle y \rangle \neq 0$.

Exercise: Show that the theory with $W(X,Y,Z) = hXY^2 + mYZ + \xi X$ has no supersymmetric vacuum. Supersymmetry is spontaneously broken. Find the supersymmetry-breaking minimum and check that there is a massless fermion in the spectrum (the so-called Goldstino.)

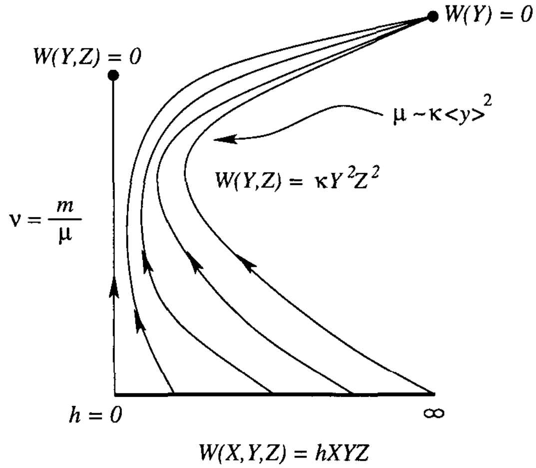


Figure 6. Away from the origin of moduli space, the combination of W and $\langle y \rangle$ drive the flow to a different endpoint.

2.6. *Solitons in XYZ*

What happens to the classical theory if we take the superpotential $W = hXYZ + \xi X$? The equations $\frac{\partial W}{\partial X} = hYZ + \xi X = 0$, $\frac{\partial W}{\partial Y} = hXZ = 0$, $\frac{\partial W}{\partial Z} = hXY = 0$, have solutions $X = 0$, $YZ = \xi/h$. (Henceforth we will usally not distinguish the chiral superfield's expectation value from that of its component scalar field; thus $X = 0$ means $\langle x \rangle = 0$.) Now the equation $YZ = c$, c a constant, is a hyperbola. To see this consider c real (without loss of generality) and now note that for Y real the equation $YZ = c$ gives a real hyperbola; rotating the phase of Y gives a complex hyperbola. What has happened to the three branches is this: the X-branch has been removed, while the two cones of the Y and Z-branches have been joined together as in the figure below. The moduli space $YZ = c$ (Fig. 7) can be parameterized by a single modulus $\frac{Y}{\sqrt{c}}$ which lives on the complex plane with the point at zero removed.

This seemingly innocent theory hides something highly nontrivial. Let us take the case of $d = 3$, and use polar coordinates r, θ on the two spatial directions. Now suppose that we can find a time-independent circularly-symmetric solution to the classical equations, of characteristic size r_0, in which

$$Y(r,\theta) = \sqrt{c} f(r) e^{i\theta} \ , \ Z(r,\theta) = \sqrt{c} f(r) e^{-i\theta}$$

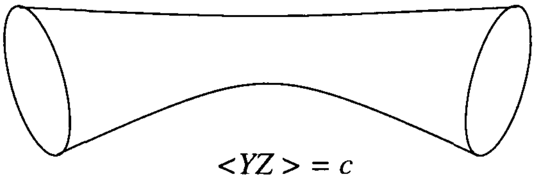


Figure 7. The moduli space when $W = XYZ + \xi X$.

where $f(r \to 0) = 0$ (to avoid multivalued fields at the origin) and $f(r \gg r_0) = 1$ (so that $YZ = c$ at large r, meaning the potential energy density in this solution is locally zero for $r \gg r_0$.) This object would be a candidate for a "vortex" soliton, a composite particle-like object, in which Y and Z wind (in opposite directions, maintaining $YZ = c$) around the circle at spatial infinity, and in which the energy density is large only in a "core" inside $r = r_0$ (Fig. 8).

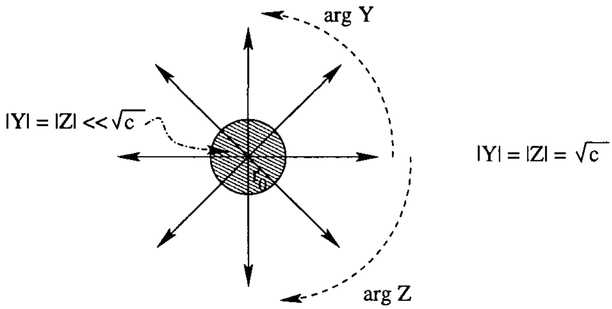


Figure 8. The vortex, which has logarithmically divergent energy.

However, although the energy density falls to zero rapidly for $r \gg r_0$, the total energy of this vortex diverges. The kinetic energy from the winding of Y and Z is logarithmically divergent — so any soliton of this type would have infinite energy. Why even consider it? Well, a vortex and antivortex *pair*, a distance Δ apart, will have finite energy, and this even if they are quite far apart, with $\Delta \gg r_0$, as in Fig. 9. For such a pair, Y and Z would wind locally but not at spatial infinity. Consequently there would be no logarithmic divergence in the energy; instead the energy would go as twice the core energy plus a term proportional to $\log \Delta$. Since we cannot pull the solitons infinitely far apart, we should think of these objects as logarithmically confined solitons.

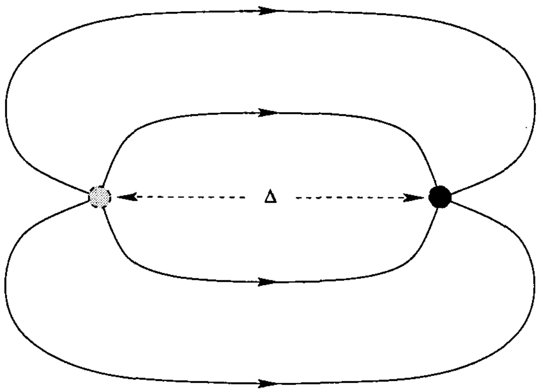


Figure 9. A finite-energy vortex/antivortex configuration; there is no winding of $\arg Y$ at infinity.

This is not as strange as it sounds. *In three dimensions, massive electrons exchanging photons are logarithmically confined!* In other words, these vortices are no worse than electrons! and indeed they behave as though they are exchanging some massless particle. We'll come back to this idea later.

In four dimensions, we may repeat the same analysis. Instead of particle-like vortices, we would find logarithmically-confined vortex *strings*, extended in the third spatial dimension. Again, these really can be pair-produced in the theory (or rather, as in Fig. 10, a closed loop of this string can be created at finite energy cost.)

2.7. *A more general form*

Finally, I should point out that I have by no means written the most general supersymmetric theory. Taking any real function $K(\Phi^i, \Phi^{i\dagger})$ of chiral superfields Φ^i and their conjugates, called the "Kähler" potential, and a superpotential $W(\Phi)$, one can construct a supersymmetric theory by writing

$$S_{\text{general}} \;=\; S_{kin} + S_{int} + S_{int}^{\dagger}$$

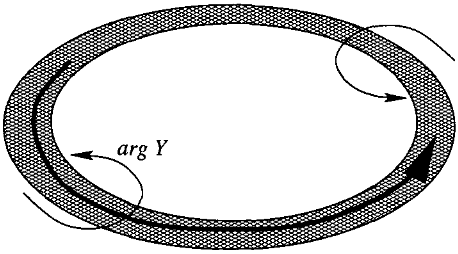


Figure 10. In $d = 4$, a closed vortex plays the role of the vortex/antivortex configuration of $d = 3$.

$$
\begin{aligned}
S_{kin} &= \int d^dx \left[\frac{\partial^2 K(\phi, \phi^\dagger)}{\partial \phi^{j\dagger}\, \partial \phi^i} \left(\partial_\mu \phi^{j\dagger} \partial^\mu \phi^i + i \bar{\psi}^j \not{\partial} \psi^i + F^{j\dagger} F^i \right) \right. \\
&\qquad \left. + \text{ higher order terms} \right] \\
S_{int} &= \int d^dx \left[-\frac{1}{2} \frac{\partial^2 W(\phi)}{\partial \phi^i\, \partial \phi^j} \psi^i \psi^j + \frac{\partial W(\phi)}{\partial \phi^i} F^i \right]
\end{aligned}
\tag{14}
$$

where repeated indices are to be summed over. (The higher order terms in S_{kin} are given in Wess and Bagger; we will not need them but they are quite interesting.) More compactly, through the definition of the "Kähler metric"

$$
K_{i\bar{j}} \equiv \frac{\partial^2 K(\phi, \phi^\dagger)}{\partial \phi^{\bar{j}\dagger}\, \partial \phi^i}
$$

and

$$
W_i \equiv \frac{\partial W(\phi)}{\partial \phi^i} \; ; \; W_{ij} \equiv \frac{\partial^2 W(\phi)}{\partial \phi^i\, \partial \phi^j} \, ,
$$

we may write

$$
\begin{aligned}
S_{kin} &= \int d^dx \left[K_{i\bar{j}} \left(\partial_\mu \phi^{\bar{j}\dagger} \partial^\mu \phi^i + i \bar{\psi}^{\bar{j}} \not{\partial} \psi^i + F^{\bar{j}\dagger} F^i \right) + \cdots \right] \\
S_{int} &= \int d^dx \left[-\frac{1}{2} W_{ij} \psi^i \psi^j + W_i F^i \right] .
\end{aligned}
\tag{15}
$$

Note that the scalar potential of the theory is now modified! It is now

$$
V(\phi_i) = W_{\bar{j}}^\dagger \left(K^{-1} \right)^{i\bar{j}} W_i
$$

(Previously we considered only a "canonical" Kähler potential $K = \sum_i \Phi_i^\dagger \Phi_i$, with a metric $K_{i\bar{j}} = \delta_{i\bar{j}}$.) The above potential is still positive or zero, however, so the condition for a supersymmetric vacuum, which was

$$W_i = 0 \quad \text{for all } i \ ,$$

remains true *unless* the metric $K_{i\bar{j}}$ is singular! Such singularities do occur (and have definite physical origins) so one must not neglect this possibility.

Exercise: Take the theory of a single chiral superfield Φ, with $W = y\Phi^3$, and rewrite it by defining a new chiral superfield $\Sigma \equiv \Phi^3$. The superpotential is now $W = y\Sigma$, for which $dW/d\Sigma \neq 0$, even for $\Sigma = 0$. Compute the Kähler potential and the Kähler metric, and show that the theory does have a supersymmetic vacuum at $\Sigma = 0$. The moral: one cannot determine from the superpotential alone whether a theory breaks supersymmetry! At a minimum, additional qualitative information about the Kähler metric is required.

Even this set of supersymmetric theories is a small subset of the whole. For example, we have considered only theories with two-derivative terms. However, there is no reason to restrict ourselves in this way. For example, in the theory $W(X,Y,Z) = hXYZ + \frac{1}{2}mX^2$ and a canonical Kähler potential, the classical effective theory at scales $\mu \ll m$ for the fields Y and Z most certainly has terms in its Lagrangian with four or more derivatives of y and/or z, suppressed by inverse powers of m. For these more general cases, the Lagrangian is not fully specified by the Kähler and superpotential alone.

Exercise: Check this last claim by substituting the equation of motion for X into the action of the original theory and expanding in $1/|m^2|$.

3. Perturbation theory

3.1. *The quantum Wess-Zumino model*

Now let's return to the theory with a single field Φ and a superpotential $W = \frac{1}{3}y\Phi^3 + \frac{1}{2}m\Phi^2$, with Lagrangian obtained using (12). Classically m is a relevant coupling; when m is zero, y is scale-invariant and the theory is conformal. What happens quantum mechanically?

Since most such theories are divergent, we must regulate them. We can do this by putting in a cutoff at a scale Λ_{UV} (though this is difficult in supersymmetric theories since most cutoffs violate supersymmetry) or by introducing ghost fields (called Pauli-Villars regulators) of mass Λ_{UV} which cancel the degrees of freedom at very high momentum while leaving those at low momentum.

Having done this, we can guess the form of perturbative corrections to any coupling constant using dimensional analysis. If the theory has one or more dimensionless coupling constants, then we expect any coupling of dimension p to get a correction of order $(\Lambda_{UV})^p$. (For $p = 0$ we expect a $\log \Lambda_{UV}$ correction.) In particular, in four dimensions a $\lambda\phi^4$ interaction gives a quadratic divergence (Λ_{UV}^2 times a function of λ) to $M^2|\phi|^2$. This is not so for ϕ^4 theory in $d = 3$; its coupling λ has dimension 1, and this reduces the degree of divergence possible in any diagram. In fact the divergence in M^2 is now proportional to $\lambda\Lambda_{UV}$. Furthermore, while λ itself gets a logarithmic divergence in $d = 4$, in $d = 3$ it gets only finite corrections.

Exercise: Check these statements about ϕ^4 theory.

However, even correct dimensional analysis can overestimate the degree of divergence if there are symmetries around. A coupling constant which *breaks* a symmetry cannot get an additive divergence, but only a multiplicative one. For example, in the Yukawa theory in Eq. (7), the fermion mass term might be expected to get a linear divergence of order $|y^2|\Lambda_{UV}$. However, the theory (7) has an explicitly broken chiral symmetry $\psi \to \psi e^{i\alpha}$; it is broken by both m and the coupling y. It also has a symmetry $\phi \to \phi e^{i\beta}$ broken by y and h. But what can you do with broken symmetries? Just ask our teachers, who understood the chiral Lagrangian of QCD! Replace the broken symmetries with "spurious" ones, under which the symmetry-breaking couplings m, y, h — thought of as though they were background scalar fields, called "spurions" — transform with definite charges. Here's a table of dimensions and of spurious charges under two spurious symmetries:

	ϕ	ψ	M^2	h	λ	m	y
$d = 4$ dimension	1	$\frac{3}{2}$	2	1	0	1	0
$d = 3$ dimension	$\frac{1}{2}$	1	2	$\frac{1}{2}$	1	1	$\frac{1}{2}$
$U(1)_\alpha$	0	1	0	0	0	-2	-2
$U(1)_\beta$	1	0	0	1	0	0	-1

(16)

For simplicity suppose for the moment that $h = 0$; now let us see why m cannot have a linear divergence in four dimensions and is finite in three. We want to know Δm, the quantum mechanical corrections to the fermion mass. These must be in the form of polynomials in the coupling constants times a possible power of Λ_{UV}. But by the above spurious symmetries, Δm *must* be proportional to m; otherwise m and Δm could not possibly have the same charge under the spurious symmetries. (Equivalently, this is required by the fact that when $m = 0$ the spurious symmetry $U(1)_{2\alpha-\beta}$ becomes a real one, and this true symmetry then forbids a non-zero Δm.) Therefore the linear dimension of Δm has already been soaked up by the factor of m, and we can therefore have only a logarithm of Λ_{UV} appearing in Δm. In $d = 3$ in this theory, it is even better; Δm comes from interactions, but all of the couplings have positive mass dimension, making even a logarithmic divergence impossible.

<u>**Exercise:**</u> Show that in this theory M^2 has a divergence Λ_{UV}^{d-2} while *all* other couplings, as well as the wave-function renormalization factors for ϕ and ψ, are log divergent in $d = 4$ and *finite* in $d = 3$ (once M^2 has been renormalized.)

Once you've gone through this exercise, you're ready to see why supersymmetry is so powerful. Supersymmetry requires $M^2 = |m^2|$; but this must hold even quantum mechanically (assuming supersymmetry is preserved) so the divergences in M^2 must be reduced down to those for m! Thus in $d = 4$ the above Wess-Zumino model has at most log divergences; and in $d = 3$ it is completely finite!

How can this happen? Let's look at the diagrams in Fig. 11.

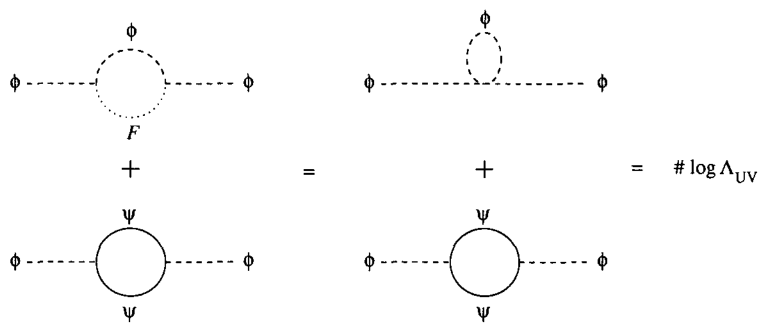


Figure 11. The quadratic divergences to the scalar mass cancel.

Now, spurious symmetries can do more; they can constrain finite as well as infinite quantum corrections. For example, if $h = 0$ and $y = 0$ then the $U(1)_\beta$ symmetry is genuine; therefore the effective potential for ϕ can only be a function of $\phi^\dagger\phi$, and a quintic ϕ^5 is clearly forbidden. Once $h \neq 0$ (still with $y = 0$ for simplicity) then that symmetry is lost. But it is easy to see that the coefficient of $\phi|\phi^2|^2$ *must* be proportional to h^* (times a polynomial in λ) since otherwise there is no way for the quantum effective Lagrangian to respect the spurious symmetries!

Now comes the astounding part. The spurious symmetries of supersymmetric theories profoundly constrain the finite as well as infinite corrections to supersymmetric theories. In particular, *the superpotential cannot be renormalized at any order in perturbation theory!* All quantum corrections must appear in the Kähler potential or in higher-derivative operators.[d]

Let us prove this, using modern methods, in the model with $W = \frac{1}{3}y\Phi^3 + \frac{1}{2}m\Phi^2$; the generalization is straightforward though tedious. The key is that the renormalized effective superpotential (which must be carefully defined, in a way that I am avoiding getting into here) is itself a homomorphic function of the chiral fields Φ (and not their conjugates) *and* of the coupling constants y and m (and not *their* conjugates). In fact, one should think of y and m as additional "background" chiral fields! That is, they act as though they are the expectation values of scalar components of other, nonpropagating, chiral fields. The mathematics of supersymmetry, and in particular the Feynman graph expansion, automatically treats them this way.

We start with two special cases. First, suppose $y = 0$ and $m \neq 0$. In this case there is obviously no renormalization since there are no quantum effects. Next, suppose $m = 0$ and $y \neq 0$. This is more interesting. The theory has one real symmetry and one spurious symmetry. The real symmetry is especially curious, in that it takes

$$\phi \to \phi e^{2i\alpha/3} \; ; \; \psi \to \psi e^{-i\alpha/3} \; ; \; F \to F e^{-4i\alpha/3} \; ; \; y \to y \; ; W \to e^{2i\alpha} W$$

which means that different parts of the Φ superfield transform differently. In fact the charge of ϕ is one unit greater than that of ψ and two units greater than that of F. Such a symmetry is called an R-symmetry, and it plays a special role since it does not commute with supersymmetry transformations.

[d]The original proofs of this, made in the 1980s, did not use spurions, so this language is somewhat ahistorical. It was Seiberg's great insights in 1993 which led him to this proof.

Note that the superpotential, thought of as a function of ϕ, transforms with charge 2; this is a requirement of any R-symmetry, as can be seen from the action (14). The spurious symmetry is more ordinary, except that y transforms:

$$\phi \to \phi e^{i\beta} \;;\; \psi \to \psi e^{i\beta} \;;\; F \to Fe^{i\beta};\; y \to ye^{-3i\beta} \;;\; W \to W$$

leaves the action invariant. Now, what terms can we write in the effective superpotential? We can only write objects which carry charge 2 under the R-symmetry and are neutral under the spurious symmetry. In perturbation theory, every term in the superpotential must be of the form $y^p\Phi^q$, p,q integers; since W is holomorphic we cannot write any powers of y^* or $\Phi^\dagger$. This is very important. We cannot have *any* functions of $|y|^2$, in contrast to what would have occurred in a nonsupersymmetric theory where holomorphy is not an issue. Clearly $y\Phi^3$ is the unique choice, and therefore the superpotential remains, even quantum mechanically, $W = \frac{1}{3}y\Phi^3$. This is remarkable; no mass term can be generated in the effective superpotential. Moreover, no Φ^4 term can be generated either. This can be understood by thinking about the component fields; the resulting term $\phi^2\psi^2$ in the Lagrangian is simply forbidden by the chiral symmetries. Clearly this is also true for any Φ^k, $k > 3$.

Now finally let us consider the more complicated case $W = \frac{1}{3}y\Phi^3 + \frac{1}{2}m\Phi^2$. In this case there are no real symmetries. However there are two spurious symmetry, one of them an R-symmetry. Under the ordinary symmetry, Φ, y, m have charge $1, -3, -2$ respectively. Invariance under this symmetry requires the superpotential depend only on $m\Phi^2$ and $u \equiv y\Phi/m$. The choice of R-symmetry is a bit arbitrary (since we may take linear combinations of any R-symmetry and the spurious but ordinary symmetry to get a new R-symmetry) but a simple choice is to assign, as before, charges $2/3, 0, 2/3$ to Φ, y, m. This means ϕ, ψ, F have charge $2/3, -1/3, -4/3$ as before.

	$\phi(\Phi)$	ψ	F	W	m	y	$u = y\Phi/m$
$d=4$ dimension	1	$\frac{3}{2}$	2	3	1	0	0
$d=3$ dimension	$\frac{1}{2}$	1	$\frac{3}{2}$	2	1	$\frac{1}{2}$	0
$U(1)_R$	$\frac{2}{3}$	$-\frac{1}{3}$	$-\frac{4}{3}$	2	$\frac{2}{3}$	0	0
$U(1)_\Phi$	1	1	1	0	-2	-3	0

(17)

Notice that u is invariant under this as well, while $m\phi^2$ has R-charge 2, so the superpotential must take the form

$$W_{eff} = \frac{1}{2}m\Phi^2 f\left(\frac{y\Phi}{m}\right)$$

where f is a *holomorphic* function of its argument. This fact is crucial. We know that when $y = 0$, $W_{eff} = \frac{1}{2}m\Phi^2$, so $f(0) = 1$. Therefore the mass term in the superpotential (we will soon see how important it is to say "in the superpotential") cannot be renormalized even when it is nonzero; any attempt to correct m with a factor of y is always accompanied by a field Φ, which means this correction does not give a contribution to Φ^2.

We also determined just a moment ago that when $m \to 0$, $W = \frac{1}{3}y\Phi^3$, so $f(|u| \to \infty) = 2u/3$. This in turn guarantees there can be no correction to the coefficient of Φ^3.

What about Φ^4? Its coefficient is y^2/m times a number, which might be zero. Let's consider first what graphs might lead to a corresponding $\phi^2\psi^2$ term in the Lagrangian. The mass term means that we can draw a non-vanishing tree graph, the first in Fig. 12, which is proportional to m, with two $\phi\psi\psi$ vertices and one chirality-flipping mass insertion, with a factor of $1/m^2$ coming from the propagator of the virtual ψ. However, we are interested in quantum effective actions, to which tree graphs do not contribute. To get a quantum contribution to a $\phi^2\psi^2$ term, we see that we need *more* than two $\phi\psi\psi$ vertices. But there's the rub; this means that the coefficient of this term in the quantum superpotential is proportional at least to y^4, or more precisely (if you look at the second diagram in Fig. 12 carefully) $y^2|y|^2$. This is in contradiction to the general form of the superpotential; therefore this term must vanish. And so on, for all perturbative contributions to terms in the superpotential. To all orders in perturbation theory, $f(u) = f_{classical}(u) = 1 + \frac{2}{3}u$, and $W_{eff} = W_{classical}$.

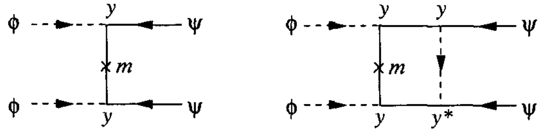


Figure 12. The first diagram has the right form, but is classical; the second diagram contributes to the quantum effective action, but has the wrong form.

But how far can we carry this argument? What about non-perturbative corrections? We know these corrections must be very small in the limit that

y is very small and m is finite; therefore $f(u) \approx 1 + \frac{2}{3}u$, to all orders in u, even nonperturbatively near $u = 0$. And we still know that $f(u) \to \frac{2}{3}u$ as $u \to \infty$, because our arguments using the real symmetry of the $m = 0$ case left no room, in that case, for an unknown function even non-perturbatively. Now, the claim is that there are no holomorphic functions *except* $f_{classical}$ which have these properties — and therefore $f = f_{classical}$ *exactly.*

Note holomorphy is essential here, as is the fact that we know the superpotential when $|u|$ is large but has arbitrary phase. For instance, holomorphy rules out functions such as

$$e^{-1/|u|^2} + \frac{2}{3}u$$

whereas our constraint on $u \to \infty$ rules out functions such as

$$e^{-1/u^2} + \frac{2}{3}u$$

which has the wrong behavior for small imaginary u.

Fantastic. The superpotential for this theory gets no quantum corrections; the coupling constants appearing there are unaltered. It would seem, then, naively, that the coupling constants of this theory do not run. But this sounds wrong. We have already argued that all the coupling constants of the theory should have logarithmic divergences in $d = 4$; has supersymmetry eliminated them? And should there be no finite renormalizations whatsoever? Indeed, this is far too facile. The effective superpotential is well under control, but the effective Kähler potential is not. The latter potential is *real*, so it can contain real functions of y^*y and m^*m appearing all over the place. Consequently we cannot make any strong statements about its renormalization. But how does this affect the coupling constants?

The resolution of this puzzle is that the coupling y appearing in $W(\Phi)$ is not a physical quantity. Let us rename it $\hat{y}$. By construction it is a holomorphic quantity. Note that we can change it by redefining our fields by $\Phi \to a\Phi$, where a is any complex constant. This changes $\hat{y}$, and it also changes the Kähler potential, making the kinetic terms noncanonical. Physical quantities (such as the running coupling constants, as measured, say, in scattering amplitudes at particular scales) must be independent of such field redefinitions. To define physical quantities, we should be more careful. Let us take the Kähler potential to have the form

$$K(\Phi^\dagger, \Phi) = Z\Phi^\dagger\Phi \ .$$

Z gives the normalization of the wave-function of Φ. Note that a propagator representing an incoming or outgoing particle state should be $i/(k^2 - m^2)$.

Thus the presence of the factor Z, giving $i/Z(k^2 - m^2)$ for the propagator, implies that we will have to take care in normalizing Green functions, a point we will return to shortly.

The rescaling of Φ by a factor a changes $\hat{y}$ by a^{-3} and Z by $|a|^{-2}$. A natural definition of an invariant coupling is then the quantity $|y|^2 \equiv \hat{y}^\dagger Z^{-3}\hat{y}$, which is clearly invariant under field redefinitions of the sort we were just considering. The coupling $|y|$ is physical but non-holomorphic, in contrast to the unphysical but holomorphic $\hat{y}$.

But now we see how renormalization of the *wave function* — divergences or even finite renormalizations which affect the kinetic terms in the Kähler potential — can in turn renormalize *physical* coupling constants. While $\hat{y}$ cannot be renormalized and become a scale-dependent function, Z can indeed become a scale-dependent function of μ. In fact, we may expect that the graph in Fig. 13 will renormalize the wave function of Φ by

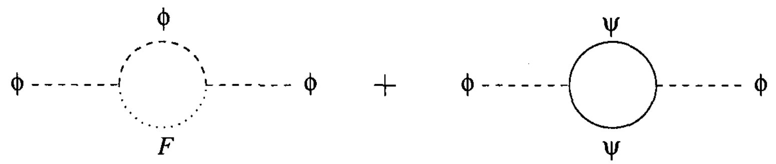


Figure 13. One-loop diagrams contributing to $Z(\mu)$.

$$Z(\mu) = 1 + \tilde{c}_0 \frac{|y|^2}{16\pi^2} \log(\mu/\Lambda_{UV}) = Z(\mu_0) + \tilde{c}_0 \frac{|y|^2}{16\pi^2} \log(\mu/\mu_0) \tag{18}$$

where $\tilde{c}_0$ is a constant of order 1 and μ_0 is an arbitrary scale. Shortly we will determine the sign of $\tilde{c}_0$.

We may now determine the scaling behavior of the physical coupling y (again, to be distinguished from the holomorphic coupling $\hat{y}$ appearing in the superpotential.)

$$\beta_{|y|^2} = y^*\beta_y + y\beta_{y^*} = -3|y|^2 \frac{\partial \ln Z}{\partial \ln \mu}.$$

Let us define the *anomalous mass dimension* $\gamma(y)$ of the field Φ by

$$\gamma = -\frac{\partial \ln Z}{\partial \ln \mu}$$

which tells us how Z renormalizes with energy scale.

Exercise: Why should we think of this as an "anomalous dimension"? What is the relation between $Z(\mu)$ and the dimension of a field? Show $\dim \Phi = 1 + \frac{1}{2}\gamma$.

Then we have an *exact* relation

$$\beta_y = \frac{3}{2} y \gamma(y) \tag{19}$$

We can use Eq. (18) at small y to find the approximate result

$$\gamma = -\tilde{c}_0 \frac{|y|^2}{16\pi^2} + \text{order } (|y|^4) \Rightarrow \beta_y = -\frac{3}{2}\tilde{c}_0 y \frac{|y|^2}{16\pi^2} + \text{order } (|y|^4) \ ,$$

but when y is larger we have no hope of computing $\gamma(y)$, and therefore none of computing β_y. Fortunately, even though $\gamma(y)$ itself is an unknown function, relations such as (19) can be extremely powerful in and of themselves, as we will see shortly.

Let us understand where this relation (19) came from by looking at diagrams. Fig. 14 shows the full propagator, which is proportional to Z^{-1}; therefore we must normalize the external fields in any physical process by a factor of $1/\sqrt{Z}$. The graphs contributing to the physical $|y|^2$ take the form

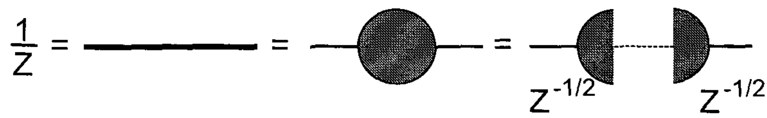


Figure 14. The full propagator is proportional to Z^{-1}; external fields must then be normalized with a factor of $Z^{-1/2}$.

of Fig. 15; but supersymmetry eliminates all corrections to the holomorphic Φ^3 vertex, making the graphs much simpler — and (remembering that we must normalize the fields) proportional to Z^{-3}.

As an important aside, let me note that chiral superfields have a very special property, namely that products of chiral superfields have *no short-distance singularities!* In contrast to expectations from non-supersymmetric field theories, composite operators built from chiral fields (but no antichiral or real fields) have the property that they may defined without a short-distance subtraction. The dimension of such a composite is the sum of the dimensions of its components.

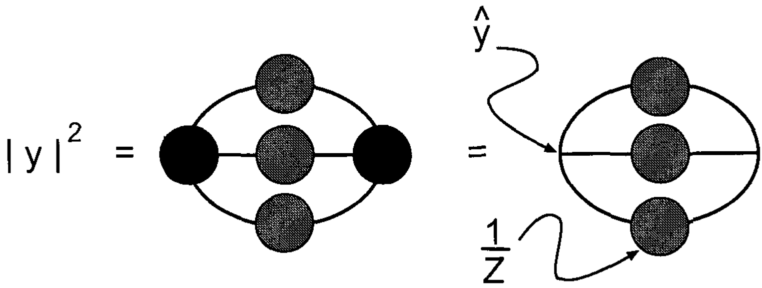


Figure 15. The physical coupling $|y|^2$ gets quantum corrections only from Z (grey circles); the vertex factors (black circles) get no quantum contributions, since the holomorphic coupling $\hat{y}$ is unrenormalized.

Exercise: Show that $\beta_y = y[\dim(\Phi^3) - \dim W] = y[3 \dim(\Phi) - \dim W]$.

Now, there is a very important theorem which we may put to use. Near any conformal fixed point (including a free field theory) all gauge invariant operators $\mathcal{O}$ must have dimension greater than or equal to 1 (or more generally, $(d-2)/2$). If its dimension is 1 (or more generally, $(d-2)/2$), then $\nabla^2\mathcal{O} = 0$ (i.e., the operator satisfies the Klein-Gordon equation.) This is true without any appeal to supersymmetry!

The theorem applies to the scalar field ϕ which is the lowest component of Φ. *Therefore,* $\gamma \geq 0$*; and* $\gamma = 0$ *if and only if* $y = 0$. From this we may conclude that

$$\gamma(y) = c_0 \frac{|y|^2}{16\pi^2} + \text{ order } (|y|^4)$$

where c_0 is a *positive* constant. This in turn implies $\beta_y > 0$, and so y flows to zero in the far infrared.

Exercise: Calculate c_0.

Thus, rather than being a conformal field theory, as it was classically, with an *exactly* marginal coupling y, the quantum $W = y\Phi^3$ theory flows logarithmically to a free conformal field theory with $y = 0$. We refer to y as a *marginally irrelevant* operator; it is marginal to zeroth order in y, but when y is nonzero then $\beta_y > 0$. The quantum renormalization group flow of the theory with nonzero y and m is shown in Fig. 16.

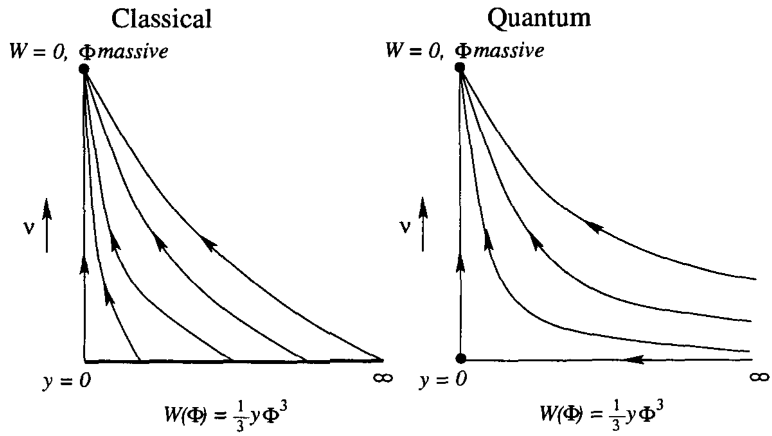


Figure 16. The scale-invariance of y is lost through quantum effects; even if $m = 0$ the coupling y flows.

3.2. *Wess-Zumino model in* $d = 3$

Now, what do we expect to happen in three dimensions? Here the formula (19) is not really appropriate, because it leaves out the *classical* dimension of y. Perturbation theory can only be done in dimensionless quantities, so we should study not y but $\omega = y/\sqrt{\mu}$. Already we notice a problem. The coupling ω is large, classically, when $\mu \ll 1/y^2$, so perturbation theory can't possibly work in the infrared! At long distances this theory will automatically be strongly coupled, unless large quantum effects change the scaling of ω drastically. But quantum effects will generally be small unless ω is large — so this can't happen self-consistently.

Let's be more explicit. The beta function for ω is

$$\beta_\omega = \omega \left[-\frac{1}{2} + \frac{3}{2}\gamma(\omega) \right]$$

Again γ must be positive (by the above theorem) and a perturbative power series in ω, beginning at order $\omega^2/16\pi^2$. It has a large negative beta function (meaning it grows toward the infrared) and will only stop growing if $\gamma(\omega) = \frac{1}{3}$. However, this can only occur if $\omega/4\pi \sim 1$, so a one-loop analysis will be insufficient by the time this occurs. Consequently, the most important behavior of the theory will occur in regimes where the perturbative expansion is breaking down, and nonperturbative effects might be important. We cannot expect perturbation theory to tell us everything about this theory, and specifically we cannot reliably calculate $\gamma(\omega)$.

However, suppose that there is *some* coupling ω_* for which $\gamma(\omega_*) = \frac{1}{3}$. This need not be the case; it could be that $\gamma < \frac{1}{3}$ for all values of ω. But if it is the case, then at ω_* the beta function of the dimensionless coupling ω vanishes, and the theory becomes truly scale invariant. (Notice that the beta function for y is nonzero there; but scale invariance requires that *dimensionless* couplings not run.) In fact, since $\gamma < \frac{1}{3}$ for $\omega < \omega_*$, and since (barring a special cancellation) we may therefore expect $\gamma > \frac{1}{3}$ for $\omega > \omega_*$, the beta function for ω is negative below ω_* and positive above it. The renormalization group flow for ω then is illustrated in Fig. 17. The point $\omega = \omega_*$ is a stable infrared fixed point; if at some scale μ the physical coupling ω takes a value near ω_*, then, at smaller μ, ω will approach ω_*.[e]

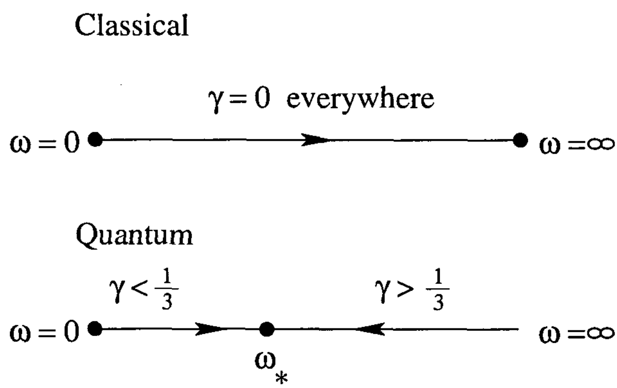


Figure 17. The scale-dependence of the coupling $\omega = y/\sqrt{\mu}$ is lost, due to large quantum corrections, at the value ω_*.

<u>**Exercise:**</u> What is the behavior of the wave function $Z(\mu)$ once this fixed point is reached?

At this conformal fixed point, the field Φ will have dimension $\frac{2}{3}$. Thus

[e]We know a scale can be generated in a classically scale-invariant theory; this is called "dimensional transmutation" and is familiar from QCD. Here we have a theory with a classical scale $\hat{y}^2$; but this scale is *lost* quantum mechanically. This equally common phenomenon is often called "dimensional antitransmutation."

the operator Φ^4, when appearing in the superpotential, represents not a marginal operator of dimension 2 (as in the free theory) but an irrelevant operator of dimension $\frac{8}{3} > 2$.

<u>**Exercise:**</u> For the mass term $m\Phi^2$, what is the dimension of m at the fixed point?

Should we skeptical of the existence of this fixed point? No; massless nonsupersymmetric $\lambda\phi^4$ theory has a fixed point for $d = 4 - \epsilon$ dimensions, with $\lambda \sim \epsilon$. This "Wilson-Fisher" fixed point is easy to find in perturbation theory. It has been verified that it continues all the way to $d = 3$ dimensions. A similar analysis can be done for a complex scalar and two-component complex fermion coupled as in (7); this too shows a similar fixed point.

As another example, a theory (the $O(N)$ model) with N massless scalar fields ϕ_i and a potential $V = \lambda(\sum_i \phi_i^2)^2$ can be shown, directly in $d = 3$ and at leading order in a $1/N$ expansion, to have a nontrivial fixed point $\lambda \sim 1/N$. A corresponding analysis can be done for a supersymmetric theory with N chiral fields Φ_i and a single chiral field X, with superpotential $hX(\sum_i \Phi_i^2)$.

<u>**Exercise:**</u> Verify the claims of the previous paragraph for the $O(N)$ model. Then try the supersymmetric theory with $W(X, \Phi_i) = hX(\sum_i \Phi_i^2)$; show it has a fixed point. Compute the anomalous dimensions of X and Φ at the fixed point (hint — do the easiest one, then use conformal invariance to determine the other.) Work only to leading nonvanishing order in $1/N$. You may want to see Coleman's Erice lectures on the $O(N)$ model.

Thus the putative fixed point at $\omega = \omega_*$ is very plausible, and we will assume henceforth that it exists. Note that at this fixed point the effective theory is highly nonlocal. It has no particle states (the propagator $\langle\phi^\dagger(x)\phi(0)\rangle = x^{-4/3}$ does not look like a propagating particle or set of particles, which would have to have integer or half-integer dimension.) In this and all similar cases the superpotential and Kähler potential together are insufficient. Indeed it is very difficult to imagine writing down an explicit Lagrangian for this fixed point theory. Even in two dimensions it is rarely known how to write a Lagrangian for a nontrivial fixed point, although in two dimensions there are direct constructive techniques for determining the properties of many conformal field theories in detail. In $d = 3$ there are no such techniques known; we have only a few pieces of information, including the dimension of Φ.

3.3. *Dimensions and R-charge*

In fact there is another way (which at first will appear trivial) to determine the dimension of Φ at the fixed point. This requires looking a bit more closely at the symmetries of the Lagrangian (7). In particular, for $W = \frac{1}{3}\hat{y}\Phi^3$, we saw earlier that there is an R-symmetry

$$\phi \to \phi e^{2i\alpha/3}\ ;\ \psi \to \psi e^{-i\alpha/3}\ ;\ F \to F e^{-4i\alpha/3}$$

which is the unique R-symmetry of the theory. Under this symmety Φ has charge 2/3 and the superpotential has charge 2.

At a conformal fixed point, there is a close relation between the dimensions of many chiral operators and the R-charges that they carry. The energy-momentum tensor (of which the scale-changing operator, the "dilation" or "dilatation generator", is a moment) and the current of the R-charge are actually part of a single supermultiplet of curents. At a conformal fixed point, the dilation current and the R current are both conserved quantities, and the superconformal algebra can then be used to show that the dimension of a chiral operator is simply $(d-1)/2$ times its R charge. This implies that at any fixed point with $W \propto \Phi^3$, the dimension of Φ is 1 in $d=4$, 2/3 in $d=3$, and 1/3 in $d=2$. (In all cases, dim Φ = dim ϕ = dim $\psi - \frac{1}{2}$ = dim $F - 1$.) Since the dimension of Φ must be strictly greater than 1 (1/2) [0] at any nontrivial fixed point in $d = 4(3)[2]$, it follows that there can be nontrivial fixed points in two or three dimensions, but not for $d=4$.

3.4. *The quantum XYZ model in* $d = 3$

Now let us turn to some other fixed points. If we have three fields X, Y, Z and superpotential $W = \hat{h}XYZ$, then by symmetry $\gamma_X = \gamma_Y = \gamma_Z \equiv \gamma_0$. In four dimensions there can again be no nontrivial fixed point; the coupling h is marginally irrelevant. But in three dimensions, it is relevant and interesting dynamics can be expected. In particular, the coupling $\eta = h/\sqrt{\mu}$ has

$$\beta_\eta = \frac{1}{2}\eta[-1 + \gamma_X(\eta) + \gamma_Y(\eta) + \gamma_Z(\eta)] = \frac{1}{2}\eta[-1 + 3\gamma_0(\eta)]\ .$$

As in the previous theory we see that if there exists some η_* for which $\gamma_0(\eta_*) = \frac{1}{3}$, then the theory is conformal there. Let us assume η_* exists and the flow is as given in Fig. 18.

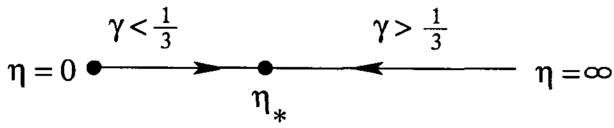


Figure 18. A fixed point for η.

<u>Exercise:</u> List the marginal and relevant operators at this fixed point. Note that some apparently relevant operators are actually just field redefinitions and are not actually present; for example, the addition of mXY to the superpotential $W = hXYZ$ has no effect, since we may simply redefine $Z \to Z - m$ and eliminate the coupling altogether. Such eliminable operators are called "redundant."

Now we can consider the effect of adding other operators to the theory. For example what happens if we add $\omega_x \sqrt{\mu} X^3$ to the theory?

<u>Exercise:</u> Prove that ω_x is a marginally irrelevant coupling. To do this use the facts that (1) at $\eta = 0$, $\omega_x \neq 0$, we know $\gamma_X > 0 = \gamma_Y = \gamma_Z$, and (2) $\gamma_X - \gamma_Y$, a continuous real function of the couplings, is known to be zero when $\omega_x = 0$.

Since ω_x (and similarly ω_y and ω_z, when we add them in as well) are marginally irrelevant couplings, we may wonder if this fixed point, located at $(\eta, \omega_x, \omega_y, \omega_z) = (\eta_*, 0, 0, 0)$, is an isolated point in the space of the four coupling constants. In fact, the answer is no. Examine the four beta functions

$$\beta_\eta = \frac{1}{2}\eta(-1 + \gamma_X + \gamma_Y + \gamma_Z)$$

$$\beta_{\omega_x} = \frac{1}{2}\omega_x(-1 + 3\gamma_X)$$

$$\beta_{\omega_y} = \frac{1}{2}\omega_y(-1 + 3\gamma_Y)$$

$$\beta_{\omega_z} = \frac{1}{2}\omega_z(-1 + 3\gamma_Z) \tag{20}$$

where $\gamma_X, \gamma_Y, \gamma_Z$ are functions of the four couplings. We see that a condition for all four beta functions to vanish simultaneously puts only *three* conditions on the anomalous dimensions $\gamma_X, \gamma_Y, \gamma_Z$. Specifically, the conditions are $\gamma_X(\eta, \omega_x, \omega_y, \omega_z) = \frac{1}{3}$ and similarly for γ_Y and γ_Z. Three conditions on four couplings imply that any solutions occur generally on one-dimensional subspaces (and since these couplings are complex, the subspace is one-complex-dimensional in extent.) Since the three anomalous dimensions must be equal on this subspace, the symmetry permuting the three fields is presumably unbroken on it. Let us therefore take $\omega_x = \omega_y = \omega_z = \omega_0$ and examine the anomalous dimension $\gamma_0(\eta, \omega_0)$. Fig. 19 indicates the renormalization group flow of the couplings. Notice that there is a *line* of conformal field theories ending at $\eta = \eta_*, \omega_0 = 0$ and extending into the η, ω_0 plane. The line ends at $\eta = 0, \omega_0 = \omega_*$, clearly the same ω_* as in the $W = \frac{1}{3}\hat{y}\Phi^3$ model (since for $\eta = 0$ we have three noninteracting copies of the latter model) The precise location of this line is totally unknown, since we do not know $\gamma_0(\eta, \omega_0)$; but if ω_* and/or η_* exists, then the line must exist also. We can define a new coupling $\rho(\eta, \omega_0)$ which tells us where we are along this line. This coupling is called an "exactly marginal coupling," and the operator to which it couples is called an "exactly marginal operator."

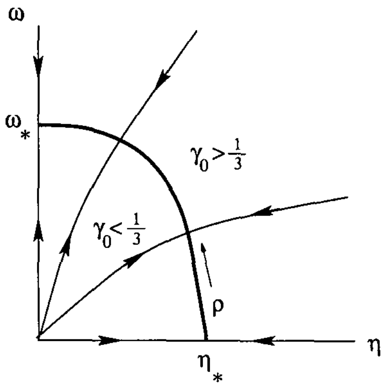


Figure 19. The (complex) line of fixed points lies at $\gamma_0(\omega, \eta) = \frac{1}{3}$ and may be parameterized by a single (complex) variable ρ.

<u>**Exercise:**</u> Argue there are no nontrivial fixed points in $d = 4$ in supersymmetric theories with only chiral superfields (and no gauge interactions.) Use the fact that since all chiral superfields are gauge invariant operators, their dimensions are greater than one. Then consider all possible interactions which are relevant at the free fixed point and might drive the theory to a nontrivial fixed point. Note loopholes in your proof.

4. Abelian gauge theories

4.1. *The classical theory*

Time to turn to gauge theories. Gauge bosons are contained in vector supermultiplets, given by superfields V, which are real and contain, in $d = 4$, a real vector potential A_μ, a Majorana fermion λ_a called a "gaugino", and a real auxiliary field D. In three dimensions, the only change is that the vector potential has one less component, which is made up by the presence of a single real scalar field φ. All of these are in the adjoint representation of the gauge group. In the case of $U(1)$, which we now turn to, they are all neutral.[f]

The kinetic terms of the pure $U(1)$ theory are

$$S_{gauge} = \frac{1}{e^2}\int d^dx \left[-\frac{1}{4}F_{\mu\nu}^2 - i\bar{\lambda}\partial\!\!\!/\lambda + \frac{1}{2}D^2\left\{+\frac{1}{2}(\partial_\mu\varphi)^2\right\}\right]$$

The last term is absent in four dimensions. Notice I have normalized all of the fields with a $1/e^2$ out front; this is convenient for many purposes. However, since e^2 has mass dimension $4 - d$, this means that I have made the dimension of the gauge field somewhat unusual. In four dimensions, it has dimension 1, like any free bosonic field, but in $d = 3$, the gauge field and scalar *also* have dimension 1, in contrast to the scalars in the chiral multiplet which were normalized with dimension $\frac{1}{2}$. This choice is arbitrary; but we will see soon why this is physically convenient.

It is instructive to count degrees of freedom. Accounting for gauge invariance but not the equations of motion, the gauge boson has $d - 1$ degrees of freedom, the Majorana fermion 4 real degrees of freedom (we will write them as 2 complex), the auxiliary field has 1 and the scalar field has $4 - d$; thus there are four bosonic and four fermionic degrees of freedom.

---

[f]Instead of V, it is often convenient to use W_α, a superfield containing the gaugino λ, the field strength $F^{\mu\nu}$, the auxiliary field D, and (in $d = 3$) φ. This object transforms homogeneously under gauge transformations (and is gauge invariant in the abelian case.) There is yet another useful superfield in $d = 3$ but I'll skip that here.

After the equations of motion, the gauge boson has $d-2$, the fermion 2, and the scalar $4-d$, so there are two bosonic and two fermionic *propagating* degrees of freedom.

What we have defined above is the vector supermultiplet of a theory with four supersymmetry generators. The chiral multiplet also comes from such a theory. The fact that the Majorana fermion has four real degrees of freedom is related to the number of supersymmetry generators. Confusingly, this much supersymmetry is known as $\mathcal{N}=1$ in $d=4$ and $\mathcal{N}=2$ in $d=3$. This is because in four dimensions there is *one* gaugino, while in $d=3$ the gaugino defined above is actually a reducible spinor, representing *two* copies of the smallest possible spinor.

4.2. *Extended supersymmetry*

There are other supersymmetries, each with their own vector and matter multiplets. A theory with *eight* supersymmetry generators has two Majorana spinors in $d=4$ and four of the smallest spinors in $d=3$; it is therefore called $\mathcal{N}=2$ in $d=4$ and $\mathcal{N}=4$ in $d=3$. Its vector multiplet contains one vector multiplet plus one chiral multiplet (both in the adjoint) from the case of four generators. Altogether it contains a gauge boson A_μ, *two* Majorana fermions λ, ψ, a complex scalar Φ, and three real auxiliary fields $D, \mathrm{Re}\ F, \mathrm{Im}\ F$, as well as (in three dimensions only) a real scalar φ. (With this much supersymmetry there is another multiplet, called a hypermultiplet, consisting of two chiral multiplets of opposite charge under the gauge symmetry; more on this below.) A theory with 16 supersymmetry generators — the maximum allowed without introducing gravity — has only vector multiplets, each of which contains one vector multiplet and three chiral multiplets of the 4-generator case (*i.e.*, one vector multiplet and one hypermultiplet of the 8-generator case,) all in the adjoint representation. It is called $\mathcal{N}=4$ in $d=4$ and $\mathcal{N}=8$ in $d=3$. Finally, in $d=3$ there are some cases with 2, 6 and 12 supersymmetry generators, which we will not have time to discuss.

In these lectures we will use the language of $d=4$ $\mathcal{N}=1$ (which is almost the same as $d=3$ $\mathcal{N}=2$) even to describe the other cases. This is common practise, since there is little convenient superfield notation for more than four supersymmetry generators.

Number of SUSY generators	$d=3$	$d=4$
4	$\mathcal{N}=2$	$\mathcal{N}=1$
8	$\mathcal{N}=4$	$\mathcal{N}=2$
16	$\mathcal{N}=8$	$\mathcal{N}=4$

(21)

4.3. *The gauge kinetic function*

Before proceeding further it is important to mention the θ angle in $d=4$ gauge theories. In the $d=4$ action, we should also include a term[g]

$$\int d^4x\ \frac{\theta}{32\pi^2} F^{\mu\nu}\tilde{F}_{\mu\nu} = \int d^4x\ \frac{\theta}{32\pi^2} F^{\mu\nu}F^{\rho\sigma}\epsilon_{\mu\nu\rho\sigma}$$

a term to which electrically charged objects are insensitive but which strongly affects magnetically charged objects. In fact we should define a generalized holomorphic gauge coupling

$$\tau \equiv \frac{1}{2\pi}\left[\theta + i\frac{8\pi^2}{e^2}\right] . \tag{22}$$

Then we may write the action as

$$S_{gauge} = \frac{i\tau}{8\pi}\int d^4x\ \left[\frac{1}{4}(F^2 + iF\tilde{F}) + i\bar{\lambda}\partial\!\!\!/\lambda - \frac{1}{2}D^2\right] + \text{hermitean conjugate.}$$

Even this is not sufficiently general. Consider, for example, adding a neutral chiral multiplet Φ to a theory with a $U(1)$ vector multiplet V. The complex scalar ϕ can have an expectation value. In principle, just as the low-energy QED coupling in nature depends on the Higgs expectation value through radiative effects, the gauge coupling for V could depend functionally on $\langle\phi\rangle$. In other words, we could write a theory of the form

$$\frac{i}{8\pi}\int d^4x\ \tau(\phi)\left[\frac{1}{4}(F^2 + iF\tilde{F}) + i\bar{\lambda}\partial\!\!\!/\lambda - \frac{1}{2}D^2\right] + \text{hermitean conjugate} + \cdots \tag{23}$$

where the dots indicate the presence of many other terms required by supersymmetry, which I will neglect here. Since τ is a holomorphic quantity, it must be a holomorphic function of the chiral superfield Φ. We will refer to this new holomorphic function as the "gauge kinetic function." Thus, to define our gauge theory, we need to specify at least a superpotential, a gauge

[g]We will not discuss the dimensional reduction of this object to three dimensions.

kinetic function, and a Kähler potential; the first two are holomorphic, and the latter is real.

We can modify the theory of V and Φ to have $d=4$ $\mathcal{N}=2$ invariance. To do this, we must make sure that the Kähler potential $K(\Phi,\Phi^\dagger)$ and the gauge kinetic function $\tau(\Phi)$ are related such that the two gauginos of the $\mathcal{N}=2$ vector multiplet (one of which, in the above notation, is in the $\mathcal{N}=1$ vector multiplet, while the other is in the multiplet Φ) have the same kinetic term. The simplest theory with $\mathcal{N}=2$ has τ a constant and $K=(1/g^2)\Phi^\dagger\Phi$. There is no superpotential in this theory; the moduli space is simply the complex ϕ plane.

Exercise: Derive the above-mentioned condition!

The $\mathcal{N}=4$ $U(1)$ gauge theory in four dimensions has three complex scalars ϕ_i, $i=1,2,3$, from its three $\mathcal{N}=1$ chiral multiplets, and no superpotential. This means it has a moduli space which is simply six dimensional unconstrained flat space, with an $SO(6)$ symmetry rotating the six real scalars into each other. This is an R-symmetry, since the four fermions of $\mathcal{N}=4$ are spinors of $SO(6)$, while the vector boson is a singlet and the scalars are in the **6** representation.

What about in three dimensions? The $\mathcal{N}=2$ vector multiplet has a single real scalar, so the classical moduli space is simply the real line. Similarly, the $\mathcal{N}=4$ vector multiplet has three real scalars, and the $\mathcal{N}=8$ vector multiplet has seven. But quantum mechanically this will not be the whole story. To see why, we must discuss duality.

4.4. *Dualities in three and four dimensions*

The pure Maxwell theory in $d=4$ has a famous symmetry between its electric and magnetic fields. One may phrase this as follows: given physical electric and magnetic fields E and B, one may find a gauge potential A_μ with $F_{\mu\nu}=\partial_\mu A_\nu-\partial_\nu A_\mu$, and one may *also* find a potential C_μ with $F_{\mu\nu}=\epsilon_{\mu\nu}{}^{\rho\sigma}(\partial_\rho C_\sigma-\partial_\sigma C_\rho)$. The electric fields of one potential are the magnetic fields of the other. Notice *both* gauge potentials have separate $U(1)$ gauge invariances, $A_\mu\to A_\mu+\partial_\mu\varphi$ and $C_\mu\to C_\mu+\partial_\mu\chi$; both invariances are unphysical, since the physical fields E and B are unaffected by them. Always remember that gauge symmetries are *not* physical symmetries; they are *redundancies* introduced only when we simply our calculations by replacing the physical E and B by the partly unphysical potential A_μ (or C_μ).

The two gauge invariances simply remove the unphysical longitudinal parts of A and C. "Electric-magnetic" duality exchanges the electric currents (to which A couples simply) with magnetic currents (to which C couples simply) and as such exchanges electrically charged particles with magnetic monopoles. We know from Dirac that the charge of a monopole is $2\pi/e$, so this transformation must exchange e^2 with $4\pi^2/e^2$ — weak coupling with strong coupling — and more generally $\tau \to -\frac{1}{\tau}$. (In modern parlance, this kind of duality, whose quantum version is discussed in more detail in my Trieste 2001 lectures, is often called an S-duality.)

In $d = 3$ there is also an electric-magnetic duality, but it does not exchange a gauge potential with another gauge potential. E and B are both three-vectors in $d = 4$, but in $d = 3$ the electric field is a two-vector and B is a spatial scalar. We may exchange the two-vector with the gradient of a scalar σ, and B with its time derivative; thus $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu = \epsilon_{\mu\nu\rho}\partial^\rho\sigma$ in three dimensions. Now $\sigma \to \sigma + c$ is a global symmetry. But the expectation value of the scalar σ represents a new degree of of freedom, one which spontaneously breaks this symmetry. The Goldstone boson of this breaking, $\partial\sigma$, is just the photon that we started with. It can be shown that σ is periodic and takes values only between 0 and $2\pi e^2$. Thus, when determining the moduli space of the pure $d = 3$ $\mathcal{N} = 2$ $U(1)$ gauge theory, we need to include not only the scalar φ but also σ, and in particular these two combine as $\varphi + i\sigma$ into a complex field Σ. Since σ is periodic, the moduli space of the theory — the allowed values for $\langle\Sigma\rangle$ — is a *cylinder*, shown in Fig. 20. (Similarly, the $\mathcal{N} = 4$ theory in $d = 3$ actually has a four-dimensional moduli space, while the $\mathcal{N} = 8$ theory has an eight-dimensional moduli space.) Note the cylinder becomes the entire complex plane in the limit $e \to \infty$, which in $d = 3$ (since e^2 has mass dimension 1) is equivalent to the far infrared limit; thus the theory acquires an *accidental* $SO(2)$ symmetry, as in figure Fig. 20. Similarly, the $\mathcal{N} = 4$ and $\mathcal{N} = 8$ cases have $SO(3)$ enhanced to $SO(4)$ and $SO(7)$ enhanced to $SO(8)$ in the infrared.[h]

[h]There is one more duality transformation in $d = 3$ that *does* exchange one gauge potential with another. Identify the gauge field strength as the dual of another gauge field $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu = \epsilon_{\mu\nu\rho}V^\rho$ (which looks gauge non-invariant, but read on.) The equation of motion $\partial_\mu F^{\mu\nu} = J_e^\nu$, where J_e is the conserved electric current, tells us that $J_e^\rho = \epsilon^{\mu\nu\rho}F^V_{\mu\nu}$ (and conservation of J_e is the Bianchi identity for F^V.) An electric charge (nonzero J_e^0) corresponds to localized nonzero magnetic field F^V_{12} for V. An object carrying nonzero magnetic field is a magnetic vortex. Thus unlike electric-magnetic duality, which exchanges electrically charged particles and magnetic monopoles, which are particles in 3+1 dimensions, this duality transformation exchanges electrically

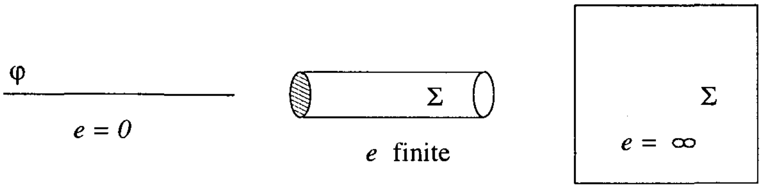


Figure 20. The moduli space is a cylinder of radius $\propto g$.

I have discussed these duality transformations in the context of pure $U(1)$ gauge theories. However, it is easy to extend them to the supersymmetric case, since all the superpartners in an abelian vector multiplet are gauge-neutral. I will do so as we need them.

4.5. *Classical $d = 4$ $\mathcal{N} = 1$ SQED*

Now we are ready to add matter to the theory. Let us add N_f chiral multiplets Q_r of charge k_r and N_f $\tilde{Q}^s$ of charge $-\tilde{k}_s$. The superpotential $W(Q_r, \tilde{Q}^s)$ and the gauge kinetic function $\tau(Q_r, \tilde{Q}^s)$ must be holomorphic functions of gauge-invariant combinations of the chiral multiplets. For now, classically, we will take $W = 0$ and $\tau = 4\pi i/e^2$. The kinetic terms for the charged fields are modified by the gauge interactions. Taking a canonical Kähler potential for simplicity, we have

$$\begin{aligned} S_{kin} = \textstyle\sum_r \int d^dx \Big[& D_\mu q_r^\dagger D^\mu q_r + i\bar{\psi}_r \not{D}\, \psi_r + F_r^\dagger F_r + \\ & k_r(\lambda \psi_r q_r^\dagger + \bar{\lambda}\bar{\psi}_r q_r + q_r^\dagger D q_r)\Big] \\ + \textstyle\sum_s \int d^dx \Big[& D_\mu \tilde{q}^{s\dagger} D^\mu \tilde{q}^s + i\overline{\tilde{\psi}^s} \not{D}\, \tilde{\psi}^s + \tilde{F}^{s\dagger} \tilde{F}^s \\ & -\tilde{k}_s(\bar{\lambda}\overline{\tilde{\psi}^s}\tilde{q}^s + \lambda\tilde{\psi}^s \tilde{q}^{s\dagger} + \tilde{q}^{s\dagger} D \tilde{q}^s)\Big] \end{aligned} \tag{24}$$

Here Q_r contains the superfields q_r, ψ_r, F_r and similarly for $\tilde{Q}^s$; the covariant derivative is $D_\mu = \partial_\mu + ikA_\mu$ acting on a particle of charge k (remember that the coupling e appears not here but in the kinetic term of the gauge boson); and D with no index is the auxiliary field in the vector multiplet.

charged particles and magetic vortices, which are particles in $2+1$ dimensions. For this reason we will call it "particle-vortex" duality; see Intriligator and Seiberg (1996).

In the special case where all the $k_r = k_s = 1$, then the Q's are rotated by an $U(N_f)$ global symmetry and the $\tilde{Q}$'s are rotated by a different $U(N_f)$ symmetry; let us call then $U(N_f)_L$ and $U(N_f)_R$ in analogy with terminology in QCD. The diagonal $U(1)_V$ which rotates Q and $\tilde{Q}$ oppositely is the symmetry which is gauged (and so is a redundancy, not a true symmetry.) The gauge-invariant chiral operators take the form $M_r^s \equiv Q_r \tilde{Q}^s$; all other gauge-invariant combinations of chiral superfields reduce to products of the M_r^s fields.

Since there is a term D^2 in the vector multiplet kinetic terms (23), we see that there will be a new contribution to the potential energy of the theory. In particular, for a canonical Kähler potential and τ a constant the potential will be

$$V(q_r, \tilde{q}_s) = \frac{1}{2g^2} D^2 + \sum_r |F_r|^2 + \sum_s |\tilde{F}_s|^2$$

where the equation of motion for D reads

$$D = \sum_r k_r |q_r|^2 - \sum_s k_s |\tilde{q}^s|^2$$

Again, a supersymmetric vacuum must have $V = 0$, and therefore all the auxiliary fields must *separately vanish.*

The condition $D = 0$ is very special. Let us consider first the simplest possible case, namely $N_f = 1$, with Q and $\tilde{Q}$ having charge 1 and -1. Although there are two complex fields, with four degrees of freedom, the gauge symmetry removes one of these, since we may use it to give q and $\tilde{q}$ the same phase. The real condition $D = |q|^2 - |\tilde{q}|^2 = 0$ removes one more degree of freedom and ensures that both q and $\tilde{q}$ have the same *magnitude.* In fact, it acts as though the gauge invariance of the theory were complexified! at least as far as the moduli space of the theory is concerned. The moduli space is then given by one complex parameter $v = \langle q \rangle = \langle \tilde{q} \rangle$, which we may also write in gauge invariant form as $v^2 = \langle M \rangle = \langle Q\tilde{Q} \rangle$. In short, the moduli space is simply the complex M plane. We began with two chiral multiplets; only one is needed to describe the moduli space.[i]

Why did one chiral multiplet of freedom have to disappear? Well, when M is nonzero, the gauge group is broken, and as we know very well, the

[i]We have assumed so far that the superpotential is zero. A superpotential $W = mQ\tilde{Q}$ simply gives the chiral multiplets masses, leaving only the massless vector multiplet and its unique vacuum at $q = \tilde{q} = 0$. If we add a superpotential $W = y(Q\tilde{Q})^2$, the resulting potential again has a vacuum only at $q = \tilde{q} = 0$, but Q and $\tilde{Q}$ are massless there.

photon is massive. But a massive gauge boson has to absorb a scalar field to generate its third polarization state, as we know from the electroweak Higgs mechanism. However, in supersymmetry it must be that a vector multiplet must absorb an entire chiral multiplet; otherwise there would be partial multiplets left over, which would violate supersymmetry. One massive photon means that one of the two massless chiral multiplets has paired up with the massless vector multiplet; the remaining fields form the massless and neutral chiral multiplet M.

How about $N_f = 2$, with Q_1, Q_2 of charge 1 and $\tilde{Q}_1, \tilde{Q}_2$ of charge -1, and no superpotential? In this case the condition $D = |q_1|^2 + |q_2|^2 - |\tilde{q}_1|^2 - |\tilde{q}_2|^2 = 0$ (combined with gauge invariance) leaves three massless chiral multiplets. It turns out that the solution to this equation is $M_1^1 M_2^2 = M_2^1 M_1^2$. Thus the *four* gauge invariant operators M_r^s, subject to the constraint $\det M = 0$, give us the three-complex-dimensional moduli space.

<u>**Exercise:**</u> Verify that $\det M = 0$ is the solution to the above equation. Hint; use the $SU(2) \times SU(2)$ flavor symmetry to rotate the vevs into a convenient form.

<u>**Exercise:**</u> Verify that for $N_f > 2$ the D-term constraints imply that the gauge-invariant operators M_r^s, subject to the constraint that M be a matrix of rank zero or one, parameterize the moduli space.

4.6. $\mathcal{N} = 2$ $d = 4$ *SQED*

Now let us slightly complicate the story by considering the $\mathcal{N} = 2$ $d = 4$ gauge theory. The $\mathcal{N} = 2$ vector multiplet has an extra chiral multiplet Φ. The fields Q_r and $\tilde{Q}^s$ can be organized into N_f hypermultiplets in which the indices r and s should now be identified. The global $U(N_f) \times U(N_f)$ symmetry will now be reduced to a single $U(N_f)$, because of the superpotential

$$W(\Phi, Q_r, \tilde{Q}^r) = \sqrt{2}\Phi \sum_r Q_r \tilde{Q}^r$$

required by the $\mathcal{N} = 2$ invariance. In normalizing the superpotential this way, I have assumed that the kinetic terms for Φ are normalized

$$K = \frac{1}{e^2}\Phi^\dagger \Phi$$

to agree with the normalization of the kinetic terms of the $\mathcal{N} = 1$ vector multiplet V. Sometimes it is more convenient to normalize Φ canonically; then a factor of the gauge coupling e appears in front of the superpotential.

Let us begin with the case $N_f = 1$. Now we have several conditions[j]

$$D = |Q|^2 - |\tilde{Q}|^2 = 0 \; ; \; F_\Phi^\dagger = Q\tilde{Q} = 0 \; ; \; F^\dagger = \Phi Q = 0; \; \tilde{F}^\dagger = \Phi\tilde{Q} = 0 \; ,$$

which clearly have no solution with nonzero Q and/or $\tilde{Q}$. In fact, in the language of the operator $M = Q\tilde{Q}$, which we know satisfies the D-term conditions, the F_Φ equation explicitly says $M = 0$, while $Q\tilde{F}^\dagger$ gives $\Phi M = 0$. This allows any nonzero Φ. When only Φ is nonzero, the gauge group is unbroken, so the photon is massless and the electric potential of a point charge is $1/r$ at large r. For this reason, this branch of moduli space is called the "Coulomb branch."

Does this structure make sense? Suppose Q and $\tilde{Q}$ were nonzero, so that the Higgs mechanism were operative; would this be consistent? As before, were the vector multiplet to become massive it would have to absorb an entire charged multiplet, which in this case would have to be the entire hypermultiplet. This would leave no massless fields to serve as moduli. Therefore this theory cannot have a branch of moduli space on which the photon is massive. By constrast, the vector multiplet scalars Φ can have expectation values without breaking the gauge symmetry at all; instead Φ simply makes Q and $\tilde{Q}$ massive, while itself remaining massless. Thus the only branch of moduli space in this theory is the Coulomb branch, in the form of the complex Φ plane, with a special point at the origin where the hypermultiplet is massless.

Now consider $N_f = 2$. We can expect that there will again be no obstruction to having $\langle\Phi\rangle \neq 0$; such an expectation value will make the hypermultiplets massive, preventing them from having expectation values. We can also expect that if the charged scalars do have expectation values, they will make the vector multiplet massive, preventing Φ from having an expectation value; and since only one hypermultiplet will be eaten by the vector multiplet, there should be an entire hypermultiplet — two chiral superfields — describing the moduli space. Is this true? As in the case of $\mathcal{N} = 1$ $N_f = 2$ SQED, the D-term conditions are satisfied by using the

[j]Henceforth we will not distinguish between superfields and their scalar components, since it is generally clear from context which is relevant; also we will generally write Φ to represent $\langle\Phi\rangle$.

operators M_r^s with $\det M = 0$. The conditions

$$F_\Phi^\dagger = Q_1\tilde{Q}_1 + Q_2\tilde{Q}_2 = 0 \ ; \ F_r^\dagger = \Phi Q_r = 0; \ \tilde{F}_r^\dagger = \Phi\tilde{Q}_r = 0$$

can be rewritten as

$$F_\Phi^\dagger = \mathrm{tr}\ M = 0 \ ; Q_s\tilde{F}_r^\dagger = \Phi M_r^s = 0$$

which indeed imply that either Φ or M can be nonzero, but not both simultaneously, and that there are two complex degrees of freedom in M which can be nonzero, making a single neutral hypermultiplet.

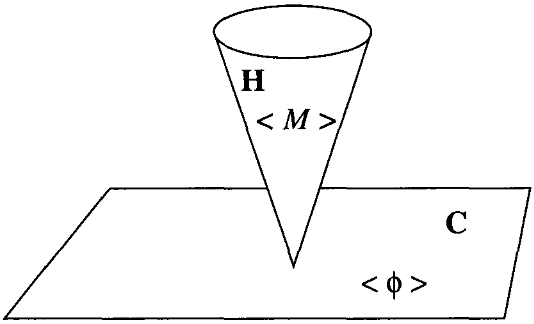


Figure 21. The classical two-complex-dimensional Higgs branch (**H**) and the one-complex-dimensional Coulomb branch (**C**) of $d = 4$ $\mathcal{N} = 2$ $U(1)$ with two charged hypermultiplets.

This last example illustrates the branch structure of these theories, indicated schematically in Fig. 21. Either Φ is nonzero, with the hypermultiplets massive and the gauge group unbroken (the Coulomb phase), or M is nonzero, with the gauge group broken (the Higgs phase) and only some massless neutral hypermultiplets remaining. In this case the Coulomb branch **C** has complex dimension 1 while the "Higgs branch" **H** has complex dimension 2 (in fact quaternionic dimension 1.) The two branches meet at the point where all the fields are massless.

<u>**Exercise:**</u> Show that for $N_f > 2$ $\mathcal{N} = 2$ $U(1)$ gauge theory, this Higgs and Coulomb branch structure continues to be found, with the quaternionic dimension of the Higgs branch being $N_f - 1$.

4.7. *Classical* $d = 3$ $\mathcal{N} = 2, 4$ *SQED*

In three dimensions, the physics is slightly more elaborate, because even the $\mathcal{N} = 2$ multiplet has a real scalar field φ. This means that even for $\mathcal{N} = 2$ $N_f = 1$ $U(1)$ gauge theory with no superpotential, there is a Coulomb branch with $\langle\varphi\rangle$ nonzero, in addition to the Higgs branch with nonzero $\langle M\rangle = \langle Q\tilde{Q}\rangle$. This is in contrast to the $d = 4$ $\mathcal{N} = 1$ $N_f = 1$ $U(1)$ gauge theory, which has only a Higgs branch. The Coulomb branch of the $d = 3$ $\mathcal{N} = 2$ theory is similar, classically, to the one we encountered in Sec. 4.6 when we considered the pure $d = 4$ $\mathcal{N} = 2$ abelian gauge theory.

The details of the moduli space are controlled partly by a new term in the Lagrangian

$$\varphi^2(|Q^2| + |\tilde{Q}|^2)\,. \tag{25}$$

The existence of this term can be inferred from the $d = 4$ $\mathcal{N} = 1$ gauge theory as follows. When we go from four dimension to three, the component of the photon A_3 becomes the scalar φ. All derivatives ∂_3 are to be discarded in this dimensional reduction, but *covariant* derivatives $D_3 = \partial_3 + iA_3$ become $i\varphi$. The interaction (25) is simply the dimensional reduction of the $d = 4$ kinetic term $|D_3Q|^2 + |D_3\tilde{Q}|^2$. For (25) to be zero, it must always be that either $\varphi = 0$ or $Q = \tilde{Q} = 0$; this gives two branches. Note that the charged chiral multiplets are massive on the Coulomb branch, as usual, because of this quartic potential. The classical moduli space is shown in Fig. 22.

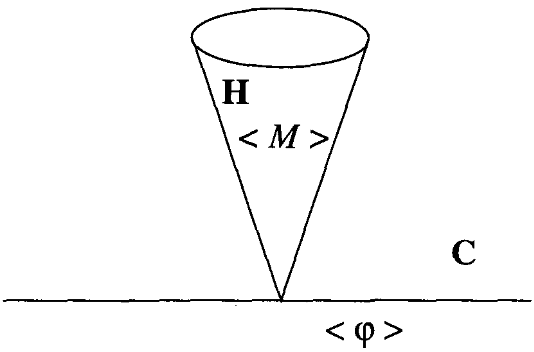


Figure 22. The classical one-complex-dimensional Higgs branch and the one-real-dimensional Coulomb branch of $d = 3$ $\mathcal{N} = 2$ $U(1)$ with $N_f = 1$.

What about the case of $\mathcal{N}=4$ supersymmetry? Here the $N_f=1$ case is not so confusing, because there is no Higgs branch; in $d=4$ $\mathcal{N}=2$ there was a one-complex-dimensional Coulomb branch, whereas here there should be (classically!) a three-real-dimensional Coulomb branch, made from the scalars $\varphi, \mathrm{Re}\,\phi, \mathrm{Im}\,\phi$. There is an $SO(3)$ symmetry acting on the scalars (which, as usual for theories with more than four supercharges, is an extended R-symmetry.)

4.8. *Quantum SQED in $d=4$*

To go further, we have to do some quantum mechanics. Let's begin with perturbation theory.

In four dimensions, perturbation theory is familiar. Just as electrons generate a positive logarithmic running for the electromagnetic coupling, via the one-loop graph above, so do scalar charged particles; and the combination of N_f chiral multiplets Q_r of charge 1 and N_f $\tilde{Q}_s$ of charge -1 gives a one-loop beta function

$$\beta_e = \frac{e^3}{16\pi^2} N_f$$

It is often convenient to write formulas not for e but for $1/e^2$:

$$\beta_{\frac{1}{e^2}} = -\frac{N_f}{8\pi^2} \Rightarrow \frac{1}{e^2(\mu)} = \frac{1}{e^2(\mu_0)} + \frac{N_f}{8\pi^2}\ln\left(\frac{\mu_0}{\mu}\right)$$

so e shrinks as we head toward the infrared. This means that e becomes large in the ultraviolet, which means that perturbation theory breaks down there, making it difficult to define the theory. We can avoid this problem by defining the theory with some additional Pauli-Villars regulator fields, N_f ghost chiral superfields of charge 1 and N_f of charge -1, all of mass M. In this case there are no charged fields above the scale M, so $\beta_e=0$; thus for $\mu > M$ the gauge coupling is a constant e_0, and

$$\frac{1}{e^2(\mu)} = \frac{1}{e_0^2} + \frac{N_f}{8\pi^2}\ln\left(\frac{M}{\mu}\right)$$

for $\mu < M$. (Remember this is only accurate at one-loop, so it only makes sense if $e_0 \ll 1$) If the fields Q and $\tilde{Q}$ have masses m then (as in real-world QED) the gauge coupling will stop running at the scale m, as illustrated in Fig. 23.

Now, the scale M was just put in to regulate the theory, while $e^2(\mu)$ is physical for low μ and should not depend on M. We therefore should define a physical scale Λ by taking it to be the value of M where the one-loop

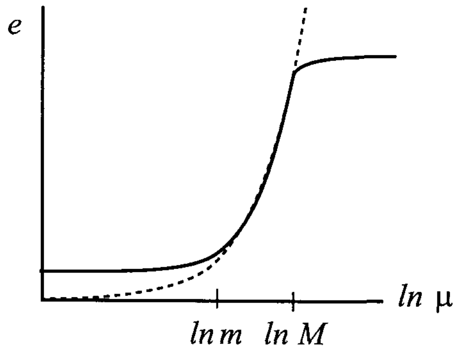


Figure 23. The coupling runs (approximately) logarithmically below M and above m.

coupling e_0 is formally infinite (though remember it will differ from the real coupling at that scale due to higher-loop effects)

$$\frac{1}{e^2(\mu)} = \frac{N_f}{8\pi^2} \ln\left(\frac{\Lambda}{\mu}\right)$$

where

$$\Lambda^{N_f} = \mu^{N_f} e^{+8\pi^2/e^2(\mu)}$$

I've been careless here: m and M are complex parameters (while μ is real[k]) so Λ should be complex also; but e^2 is real. How can we make the above expressions sensible? Clearly, we should introduce the θ angle, and rewrite the previous equation in its final form as

$$-2\pi i \tau(\mu) \equiv \frac{8\pi^2}{e^2(\mu)} - i\theta = N_f \ln\left(\frac{\Lambda}{\mu}\right)$$

which is a one-loop formula that is only sensible for $\mu \ll \Lambda$.

Note that τ is a holomorphic coupling constant. Let us verify that in perturbation theory this one-loop formula is exact! In perturbation theory the expression for τ must be a perturbative series in $e^2 \propto 1/\mathrm{Im}\ \tau$ and cannot contain $\theta \propto \mathrm{Re}\ \tau$; but that's impossible if it is to be a holomorphic expression. The only term which can appear in quantum corrections to τ must then be τ-independent, namely the one we see above.

[k]You can make it complex, actually — this is itself interesting but beyond what I can cover here.

But as before, the fact that we have found a simple formula for a holomorphic quantity by no means indicates that the physical coupling is so simple. The physical coupling gets corrections from higher loop effects (caution: as we will see, these are cancelled for theories with eight or more supersymmetry generators). As before, these can only occur in the non-holomorphic part of the theory: the Kähler potential. As we did for the coupling y in Sec. 3.1, we should find a definition of the coupling constant which is independent of field redefinitions. We'll come back to this point soon.

Let's first examine $d = 4$ $\mathcal{N} = 2$ $U(1)$ gauge theory with N_f massless hypermultiplets. We've already discussed its branch structure in Sec. 4.6; there is a Higgs branch on which some of the gauge-invariant operators $M_r^s = Q_r\tilde{Q}^s$ act as massless fields, with the others massive. There is also a Coulomb branch on which Φ has an expectation value and the term $\sqrt{2}\Phi Q_r\tilde{Q}_s$ in the superpotential gives the charged fields masses. The branch structure for $N_f > 1$ massless hypermultiplets was shown in Fig. 21. On the Coulomb branch, we can ask an interesting physical question: how does the infrared limit

$$\tau_L \equiv \lim_{\mu \to 0} \tau(\mu) \tag{26}$$

of the coupling constant τ depend on Φ? Since (1) the theory has a Φ-independent value of Λ, and (2) for any value of Φ, the coupling constant stops running at the scale Φ at which the charged fields are massive,

$$-2\pi i\tau_L = N_f \ln\left(\frac{\Lambda}{\Phi}\right) \quad [\Phi \ll \Lambda] \ .$$

This is singular only at $\Phi = 0$, where the Higgs and Coulomb branches meet and the charged fields are massless. (Recall that charged massless fields always drive the electric coupling e to zero, and thus $\tau \to i\infty$, in the infrared.) Note we cannot take $|\Phi|$ to be larger than $|\Lambda|$, since our description of the theory is not reliable there. The behavior of τ_L on the moduli space is sketched in Fig. 24.

<u>**Exercise:**</u> For $N_f = 4$, if two hypermultiplets have mass m and two have mass m', show that

$$-2\pi i\tau_L = 2\ln\left(\frac{\Lambda}{\Phi + m}\right) + 2\ln\left(\frac{\Lambda}{\Phi + m'}\right)$$

so that there are two singular points; at each singular point there are two massless hypermultiplets, which have a Higgs branch intersecting the

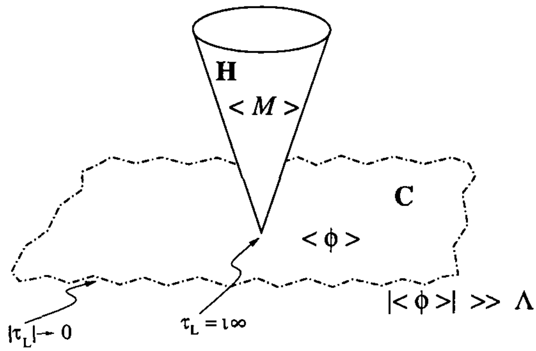


Figure 24. The quantum version of Fig. 21; the low-energy coupling $e(\mu \to 0)$ grows from 0 at the origin to ∞ at the dashed line, where the description of the theory breaks down.

Coulomb branch at that point. If each hypermultiplet has its own mass, then show that there are four singular points but *no* Higgs branches anywhere.

What does this function τ_L really tell us? Away from the singular points, for any value of Φ, the charged fields are all massive, and there is simply a pure $U(1)$ $\mathcal{N} = 2$ gauge theory in the infrared. Its effective action is of the form (23) with a nontrivial gauge kinetic function $\tau_L(\Phi)$. But like any pure abelian gauge theory, it has duality symmetries. In particular, there is the electric-magnetic transformation $\tau \to -\frac{1}{\tau}$. There is also the obvious symmetry $\tau \to \tau + 1$, which represents a shift by 2π of the θ angle. These two symmetry transformations generate a group of duality transformations of the form $SL(2, \mathbf{Z})$, the symmetry group of a torus.

A torus can be defined by taking a parallelogram and identifying opposite sides, as in Fig. 25. Ignore the size of the parallelogram by taking one side to have length 1; then the other size has a length and angle with respect to the first that can be specified by a parameter τ that lives in the upper-half of the complex plane. (For the gauge theory, the gauge coupling must be positive, so Im $\tau > 0$.) However, exchanging the two sides obviously leaves the torus unchanged ($\tau \to -\frac{1}{\tau}$) as does shifting one side by a unit of the other side ($\tau \to \tau + 1$) and any combination of these transformations.

One can therefore take the point of view that the low-energy $\mathcal{N} = 2$ $U(1)$ gauge theory should not be specified by τ. In fact, we can see this by

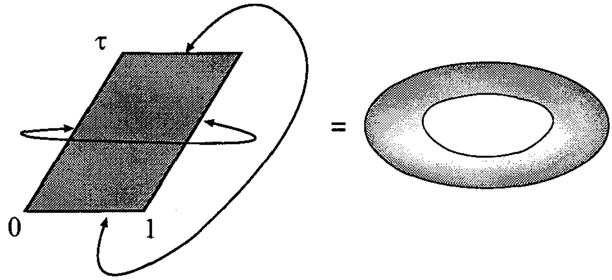


Figure 25. A torus is a parallelogram with opposite sides identified.

taking $\Phi \rightarrow \Phi e^{2\pi i}$; that is, let us circle the singular point at $\Phi = 0$. The theory obviously must come back to itself, since the physics depends only on Φ; but τ_L shifts by N_f as we make the circle. Thus τ_L does not properly characterize the theory; all values of τ_L related by $SL(2, \mathbf{Z})$ transformations are actually giving the same theory. We should therefore characterize the low-energy theory by specifying a torus! For each value of Φ, there should be a torus with parameter $\tau_L(\Phi)$ which tells us the properties of the low-energy theory. More precisely, this is a fiber bundle, with a torus fibered over the complex Φ plane, as expressed in Fig. 26. This torus is invariant under $\Phi \rightarrow \Phi e^{2\pi i}$, and becomes singular at $\Phi = 0$.

4.9. *Quantum SQED in $d = 3$*

Now let's move back to three dimensions. The gauge coupling is now dimensionful, so classically it has a negative beta function. Due to the wonders of gauge symmetry, the diagram in Fig. 27 is ultraviolet finite! But not trivial. In fact, if the fermion is masssless, and the momentum flowing through the photon line is p^μ, this graph is proportional to $\frac{1}{\sqrt{p^2}}$!

Exercise: Calculate the one-loop correction to ordinary nonsupersymmetric QED in three dimensions for N_f massless electrons.

This means that the one-loop gauge coupling in three dimensions has the form

$$\frac{1}{e^2(\mu)} = \frac{1}{e_0^2} + c_3 \frac{N_f}{\mu}$$

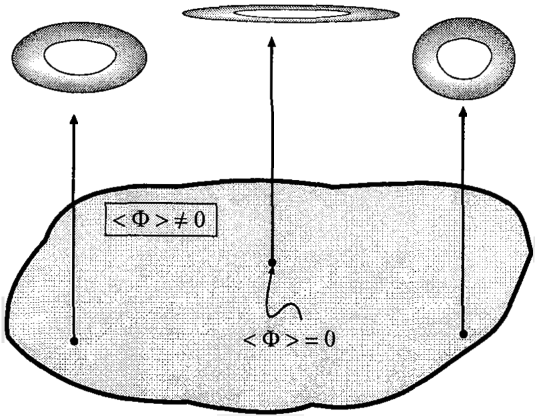


Figure 26. The low-energy gauge coupling and its dualities are best understood using a torus fibered over the Coulomb branch; at the origin, where $\tau \to i\infty$, the torus degenerates.

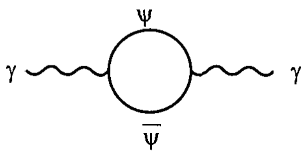


Figure 27. This one-loop graph is finite in three dimensions.

where c_3 is a positive constant, of order one, which depends on the specific theory. Notice that there is no divergence in e as $\mu \to \infty$; e goes to a constant e_0 in the ultraviolet, so in $d = 3$ supersymmetric QED is well-defined in the ultraviolet.

As always, to define a beta function we should employ a dimensionless coupling

$$\zeta \equiv \frac{e^2(\mu)}{\mu} = \left(\frac{\mu}{e_0^2} + c_3 N_f\right)^{-1}$$

which is infinite for large μ but — interestingly — goes (at one loop) to a *constant* at small μ. In other words,

$$\beta_\zeta = \frac{\mu}{e_0}^2 \left(\frac{\mu}{e_0^2} + c_3 N_f\right)^{-2} = \zeta(1 - c_3 N_f \zeta)$$

so there is a fixed point in the one-loop formula at $\zeta_* = 1/c_3 N_f$. This is illustrated in Fig. 28 and Fig. 29.

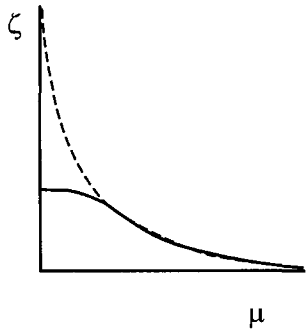


Figure 28. ζ as a function of scale μ; the dashed line shows its classical flow.

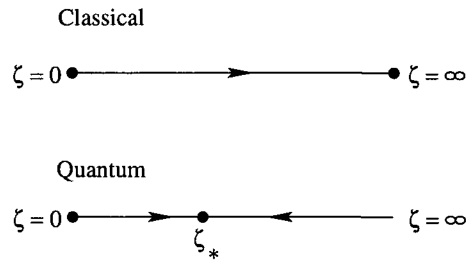


Figure 29. The coupling ξ has a quantum fixed point.

This is very interesting. Remembering that perturbation theory is an expansion in the parameter $e^2/\mu = \zeta$, we see that the one-loop formula has a fixed point *at weak coupling* if N_f is large. If this is true, then for large N_f two-loop effects such as those in Fig. 30 are always suppressed by factors of $1/N_f$ and can be neglected. Thus the large-N_f behavior of the theory is indeed given by the one-loop formula, with a fixed point at small ζ and all diagrams calculable. The theory is soluble!

Figure 30. In $d = 3$ $N_f \gg 1$ (S)QED the one-loop correction $\propto e^2 N_f$ dominates the propagator; the higher-loop corrections are suppressed by extra powers of $e^2 \propto 1/N_f$, and can be dropped.

Exercise: Show nonsupersymmetric QED in $d = 3$ is soluble and has a conformal fixed point at large N_f.

In nonsupersymmetric QED, it is believed that there is a value of N_f below which the fixed point disappears and other nonperturbative phenomena take place. In $\mathcal{N} = 2$ and $\mathcal{N} = 4$ QED, however, the fixed point visible at one loop survives for all N_f. In fact, in the latter case, there are no higher-loop corrections to the gauge coupling, so the above beta function is exact and the fixed point at small N_f is completely reliable.

Let us examine the $N_f = 1$ case in both $\mathcal{N} = 2$ and $\mathcal{N} = 4$ supersymmetry. Both of these theories are truly remarkable. As we noted, for $d = 3$ a vector boson has a electric-magnetic dual pseudoscalar σ. In $\mathcal{N} = 2$, this scalar combines with the scalar φ to make a complex field $\Sigma = \varphi + i\sigma$. The scalar σ is compact, with radius $2\pi e^2$, so classically the field Σ takes values on a cylinder of radius e^2. As $e^2 \to 0$ the field σ disappears and the cylinder becomes the φ line; when $e^2 \to \infty$ the cylinder expands to be become the entire plane, as we saw in Fig. 20.

But in the presence of charged matter, e^2 is not so simple. In particular, the terms

$$\varphi^2(|Q^2| + |\tilde{Q}|^2) + \varphi(\bar{\psi}\psi - \bar{\tilde{\psi}}\tilde{\psi})$$

imply that the charged matter has mass φ. Since $e^2(\mu)$ stops running below this scale, the low energy value e_L^2 of the gauge coupling is

$$\frac{1}{e_L^2} = \frac{1}{e_0^2} + \frac{1}{\varphi}$$

For very large $|\varphi|$ we have $e_L \approx e_0$, so the radius of the cylinder on which Σ lives is of order e_0^2 far from $\varphi = 0$. However, for very small φ the $1/e_0^2$ term can be neglected and $e_L^2 \sim \varphi$; thus the cylinder shrinks in size. At $\varphi = 0$ — where the Higgs branch meets the Coulomb branch — the cylinder shrinks to zero radius. Thus, the moduli space has the form of Fig. 31.

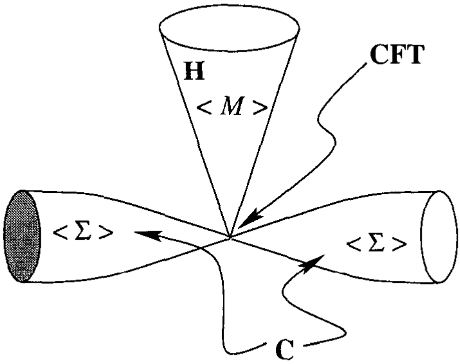


Figure 31. The quantum version of Fig. 22, which combines it with Fig. 20.

This picture should look familiar. At the meeting point of the three branches, there is a conformal field theory which looks remarkably like the $W = hXYZ$ conformal fixed point that we considered earlier (Fig. 3 and Sec. 3.4). And in fact, it is the same! In the XYZ model, the three branches had nonzero expectation values for X, Y and Z respectively. Here, the branches are the three complex planes labeled by the expectation values for $M = Q\tilde{Q}$, e^{Σ}, and $e^{-\Sigma}$. Thus we have another example of "duality"; a single conformal fixed point is the infrared physics of two different field theories, one the $W = hXYZ$ model, the other $\mathcal{N} = 2$ super-QED. The theories are different in the ultraviolet but gradually approach each other, becoming identical in the infrared, as shown in Fig. 32. This is called an "infrared" duality. Notice the Z_3 symmetry between the branches is exact in the XYZ model but is an "quantum accidental" symmetry (a property only of the infrared physics) in super-QED.

<u>**Exercise:**</u> Calculate the anomalous dimension of Q. Note the sign! Why is it allowed here?

This duality is a particle-vortex duality. Along the Higgs branch, where $\langle M\rangle \neq 0$, there are vortex solitons of *finite mass*; these are similar to the

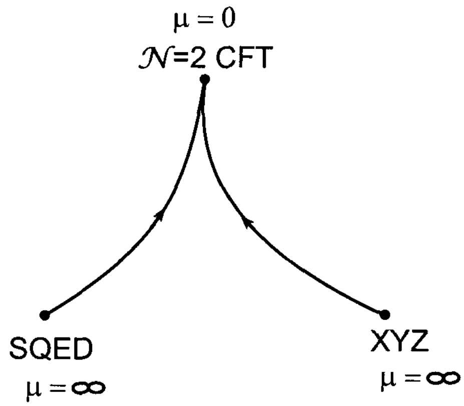


Figure 32. $\mathcal{N} = 2$ SQED and the XYZ model flow to the same fixed point.

vortices discussed in Sec. 2.6. The phases of the fields Q and $\tilde{Q}$ wind once around the circle at infinity; however, the presence of the gauge field cuts off the logarithmically divergent energy that was a feature we dwelt on in Sec. 2.6. (You can read about how this works in Nielsen and Olesen (1973).) These solitons correspond to the fields Y and Z, which have (finite) mass when $\langle X \rangle \neq 0$. Thus the vortices of the one theory correspond to the particles of the other.

<u>**Exercise:**</u> Since X and $M = Q\tilde{Q}$ are to be identified, a mass term $W = mQ\tilde{Q}$ should correspond to changing the dual theory to $W = hXYZ + mX$. The massive fields $Q, \tilde{Q}$ are logarithmically confined (as always for weakly-coupled electrically-charged particles in $d = 3$) by the light photon which remains massless. Look back at section 1.6, where we showed there are vortex solitons in the dual $W = hXYZ + mX$ theory which are logarithmically confined, and argue that it is consistent to identify Q and $\tilde{Q}$ with these solitons. Using the relation beween the gauge field and σ, try to show that the electric field surrounding the electrons $Q, \tilde{Q}$ corresponds to a variation in Σ which agrees with the properties of Y and Z near these solitons.

Now let us examine the theory with $\mathcal{N} = 4$ supersymmetry, which is even more amazing. We can obtain it from the $\mathcal{N} = 2$ case by adding a neutral chiral superfield Φ to the theory and coupling it to the other fields via the superpotential $W = \sqrt{2}\Phi Q\tilde{Q}$. This will destabilize the $\mathcal{N} = 2$ fixed point and cause it to flow to a new one, as in Fig. 33.

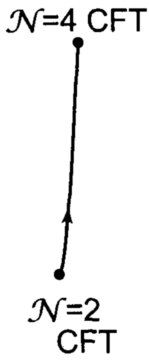


Figure 33. There is a flow linking the two theories.

Exercise: Check that the operator $\Phi Q\tilde{Q}$ is a relevant operator both at the free $\mathcal{N} = 2$ fixed point and at the infrared $\mathcal{N} = 2$ conformal fixed point.

Since the XYZ model is the same as $\mathcal{N} = 2$ SQED in the infrared, we may obtain the $\mathcal{N} = 4$ theory another way. Let us go to the far infrared of the XYZ model. We just studied what happens when we add a single field Φ and couple it to $M = Q\tilde{Q}$ in the superpotential. But we can simply change variables from SQED to XYZ; from this dual point of view, what we did was couple Φ to X. The low-energy physics of a model with $W = hXYZ + \Phi X$ should be the same as that of $\mathcal{N} = 4$ SQED. But ΦX is just a mass term which removes Φ and X from the theory, leaving Y and Z, with *no* superpotential. *Thus the dual description of the $\mathcal{N} = 4$ SQED fixed point is a free theory!*

In short, the low-energy limit of $\mathcal{N} = 4$ SQED is a conformal fixed point which can be rewritten as a free theory — a theory whose massless particles are the vortices of SQED.

From this astonishing observation, a huge number of additional duality transformations of other abelian gauge theories can be obtained. In this sense, it plays a role similar to "bosonization" (boson-fermion duality) in two dimensions, which can be used to study and solve many field theories. These three-dimensions "mirror" duality transformations, first uncovered by Intriligator and Seiberg and much studied by many other authors, are a simple yet classic example of dualities, and I strongly encourage you to study them. A summary of previous work and a number of new results on this subject appear in work I did with Kapustin (1999).

An important aside: it is essential to realize that we have here an example of a nontrivial *exact* duality, which is not merely an infrared duality. We noted that the flow from the weakly coupled XYZ to the $\mathcal{N} = 2$ fixed point is different from the flow from weakly coupled SQED to the $\mathcal{N} = 2$ fixed point. However, at the $\mathcal{N} = 2$ fixed point the two flows reach the same theory, and the operators $Q\tilde{Q}$ and X are identical there. The relevant perturbations $\Phi Q\tilde{Q}$ and ΦX may be added with arbitrarily tiny couplings; in this case the two different flows approach and nearly reach the $\mathcal{N} = 2$ fixed point, stay there for a long range of energy, and then flow out, together, along the same direction, heading for the $\mathcal{N} = 4$ fixed point. This is shown schematically in Fig. 34. In the limit where the $\mathcal{N} = 2$ fixed point is reached at arbitrarily high energies, the flow to the $\mathcal{N} = 4$ fixed point is described exactly by two different descriptions, one using the XYZ variables, the other using those of SQED. One will sometimes read in the string theory literature that "field theory has infrared dualities, but duality in string theory is exact." Clearly this is not true; as we have seen in this example, infrared dualities always imply the existence of exact dualities. You can look at my work with Kapustin (1999) for some very explicit examples.

5. Non-abelian four-dimensional gauge theory

We now turn to nonabelian gauge theories in four dimensions. This is a huge subject and we shall just scratch the surface, but hopefully this lecture will give you some sense of the immensity of this field and teach you a few of the key ideas you need to read the already existing review articles.

5.1. *The classical theory*

Let us first consider the classical pure gauge theory. The only difference from the abelian case (aside from some complications in the superfield formalism) is that the kinetic terms reflect the fact that the pure vector multiplet is self-interacting. The gauge group, a Lie group such as $SU(N)$, is generated by a Lie algebra with generators T^A, A an index running from

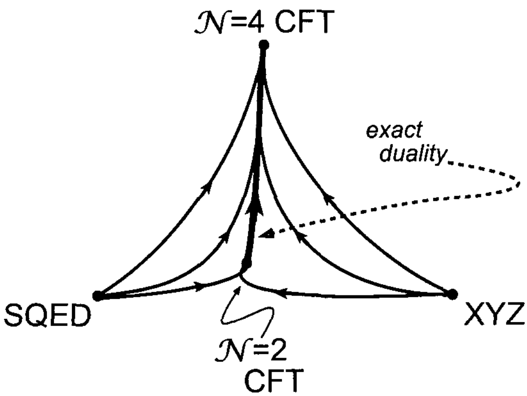


Figure 34. By adjusting couplings and scales we may obtain two exactly equivalent descriptions of the flow in Fig. 33.

1 to the dimension of the group.[1] Gauge bosons $A_\mu = A_\mu^A T^A$, gauginos $\lambda_a = \lambda_a^A T^A$, and auxiliary fields $D = D^A T^A$ are all in the adjoint representation, and the kinetic terms are the minimal ones

$$S_{gauge} = \frac{i\tau}{4\pi} \int d^4x \, \mathrm{tr} \left[-\frac{1}{4}(F^2 + iF\tilde{F}) + i\bar{\lambda}\not{D}\lambda + \frac{1}{2}D^2 \right] + \text{hermitean conjugate.} \tag{27}$$

where τ is again defined in Eq. (22) (with $e \to g$) and

$$\begin{aligned} F_{\mu\nu} = F_{\mu\nu}^A T^A &= \partial_\mu A_\nu - \partial_\nu A_\mu + i[A_\mu, A_\nu] \\ &= \left(\partial_\mu A_\nu^A - \partial_\nu A_\mu^A + f^{ABC} A_\mu^B A_\nu^C\right) T^A \,, \end{aligned} \tag{28}$$

$D_\mu \lambda_a = \partial_\mu \lambda_a + i[A_\mu, \lambda_a]$, and $D^2 \equiv \sum_A |D^A|^2$. We will often choose to represent fields in the adjoint representation using matrices T^A that are in the fundamental representation; this is usually the easiest representation

[1]For example, for $SU(N)$, A runs from 1 to $N^2 - 1$. The generators may themselves appear as $N^2 - 1$ matrices in any representation of the group. If we take T^A in the fundamental representation, then each T^A is an $N \times N$ matrix $(T^A)_i^{\bar{j}}$, where i and $\bar{j}$ are indices in the fundamental and antifundamental represention of $SU(N)$. The matrices are normalized by the condition $\mathrm{tr}(T^A T^B) = \delta^{AB}$. In the adjoint representation, $(T^A)_B^C$ is the matrix f^{ABC}, the structure constants of the group. In *any* representation, $[T^A, T^B] = i f^{ABC} T^C$.

to work with. In $SU(N)$, doing so allows us to write A_μ, λ, D as $N \times N$ hermitean traceless matrices.

The addition of charged chiral fields involves a fairly minimal change in the kinetic terms from the abelian case. If we add chiral fields Q_r and $\tilde{Q}^s$, $r, s = 1, \ldots, N_f$, in the fundamental and antifundamental representation of $SU(N)$, we obtain

$$\begin{aligned} S_{kin} &= \textstyle\sum_r \int d^dx \Big[D_\mu q_r^\dagger D^\mu q_r + i\bar{\psi}_r \not{D}\, \psi_r + F_r^\dagger F_r + \\ &\qquad \lambda\bar{\psi}_r q_r + \bar{\lambda}\psi_r q_r^\dagger + q_r^\dagger D q_r\Big] \\ &+ \textstyle\sum_s \int d^dx \Big[D_\mu \tilde{q}^{s\dagger} D^\mu \tilde{q}^s + i\bar{\tilde{\psi}}^s \not{D}\, \tilde{\psi}^s + \tilde{F}^{s\dagger}\tilde{F}^s + \\ &\qquad \lambda\bar{\tilde{\psi}}^s \tilde{q}^s + \bar{\lambda}\tilde{\psi}^s \tilde{q}^{s\dagger} + \tilde{q}^{s\dagger} D\tilde{q}^s\Big] \end{aligned}$$

where the contraction of gauge indices is in each case unique: for example in the term $q_r^\dagger D q_r$ the indices are contracted as

$$(q_r^\dagger)_{\bar{j}} D^A (T^A)_i^{\bar{j}} (q_r)^i$$

If we add a chiral superfield Φ in the adjoint representation, the kinetic terms take the same form as above, but we should interpret $\phi^\dagger D\phi$ as

$$\mathrm{tr}\ \phi^\dagger[D, \phi] = -\mathrm{tr}\ D[\phi^\dagger, \phi] = -if^{ABC} D^A \phi^{B\dagger} \phi^C$$

and similarly for the scalar-fermion-fermion terms. (Be careful not to confuse the derivative D_μ and the auxiliary field D^A!)

As in the abelian case,

- We may obtain $\mathcal{N} = 1$ gauge theories by adding arbitrary charged (and neutral) matter to the theory with arbitrary gauge-invariant holomorphic gauge kinetic and superpotential functions and an abitrary gauge-invariant Kähler potential.
- We may obtain a pure $\mathcal{N} = 2$ gauge theory by writing an $\mathcal{N} = 1$ gauge theory with a single chiral multiplet Φ in the adjoint representation, a gauge kinetic term and Kähler potential term for Φ which must be related, and zero superpotential.
- We may add matter to the $\mathcal{N} = 2$ gauge theory in the form of a hypermultiplet (two chiral multiplets Q and $\tilde{Q}$ in conjugate representations) coupled in the superpotential $W = \sqrt{2}\tilde{Q}\Phi Q$, with gauge indices contracted in the unique way. We may also add mass terms for the hypermultiplets, obtaining $W = \sqrt{2}\tilde{Q}\Phi Q + mQ\tilde{Q}$.

- Finally, if we have a massless hypermultiplet in the *adjoint* representation, so that the theory has a total of *three* chiral multiplets $\Phi_1 = \Phi, \Phi_2 = Q, \Phi_3 = \tilde{Q}$ in the adjoint, with superpotential $W = \sqrt{2}\,\mathrm{tr}\,\Phi_1[\Phi_2, \Phi_3]$, then the theory has $\mathcal{N} = 4$ supersymmetry.

Now the condition for a supersymmetric vacuum requires that $D^A = 0$, $F^i_r = 0$, $F^s_{\bar{j}} = 0$. If the superpotential is zero, then the constraints all come from

$$0 = D^A \propto (q^\dagger_r)_{\bar{j}}(T^A)^{\bar{j}}_i(q_r)^i(\tilde{q}^{s\dagger})^i(T^A)^{\bar{j}}_i(\tilde{q}_r)_{\bar{j}} = (T^A)^{\bar{j}}_i\left[(q^\dagger_r)(q_r) - \tilde{q}^{s\dagger}\tilde{q}^s\right]^i_{\bar{j}}$$

$$= \mathrm{tr}\,T^A\left[(q^\dagger_r)(q_r) - \tilde{q}^{s\dagger}\tilde{q}^s\right] \tag{29}$$

(I have written a proportional sign since the precise relation depends on the Kähler potential and gauge kinetic term, while the proportionality relation does not!)

These equations are beautifully solved in the case of $SU(N)$ with fields N_f Q in the **N** representation and $\tilde{Q}$ in the $\bar{\mathbf{N}}$ representation.[m] The hermitean matrix $\left[(q^\dagger_r)(q_r) - \tilde{q}^{s\dagger}\tilde{q}^s\right]$ can be uniquely expanded as

$$\left[(q^\dagger_r)(q_r) - \tilde{q}^{s\dagger}\tilde{q}^s\right]^i_{\bar{j}} = c_0\delta^i_{\bar{j}} + \sum_B c_B(T^B)^i_{\bar{j}}$$

Then, using the fact that

$$\mathrm{tr}\,T^AT^B = \frac{1}{2}\delta^{AB}\ ;\ \mathrm{tr}\,T^B = 0$$

we find that the conditions (29) become simply

$$\left[(q^\dagger_r)(q_r) - \tilde{q}^{s\dagger}\tilde{q}^s\right]^i_{\bar{j}} = c_0\delta^i_{\bar{j}} \tag{30}$$

for *any* c_0. Before writing any solutions to these equations, we note the following: given *any* expectation values of q and $\tilde{q}$ which are a solution to these equations, a continuously infinite class of solutions is generated by multiplying all of the q and $\tilde{q}$ fields by a complex constant. Thus there will generally be, as in the abelian case, noncompact, continuous moduli spaces

[m] In this case there is an $SU(N_f)_L$ and an $SU(N_f)_R$ symmetry acting on the Q and $\tilde{Q}$ fields respectively; there is also a baryon number under which Q and $\tilde{Q}$ have charges 1 and -1, and an *anomalous* axial symmetry (present classically but explicitly violated quantum mechanically) under which Q and $\tilde{Q}$ both have charge 1.

of vacua. As before, these vacua are not related by any symmetry, in that they have fields with different masses.[n]

<u>**Exercise:**</u> The scalars q^i_r and $q^s_{\bar{j}}$ are $N_f \times N$ and $N \times N_f$ matrices respectively. Being careful with the indices, show that the only solutions for $N_f < N$ are gauge and global symmetry transformations of the particular solution $q^i_r = v\delta^i_r$, $\tilde{q}^s_{\bar{j}} = v\delta^s_{\bar{j}}$ for $i, \bar{j} \leq N_f$. Then show that the only solutions for $N_f \geq N$ are gauge and global symmetry transformations of the particular solution $q^i_r = v\delta^i_r$, $\tilde{q}^s_{\bar{j}} = \tilde{v}\delta^s_{\bar{j}}$ for $r, s \leq N$. Note that for $N_f \geq N$, v and $\tilde{v}$ are in general different and the constant c_0 in (30) is nonzero, while for $N_f < N$ c_0 must be zero.

As another example, consider $SU(2)$ with fields Φ_n $(n = 1, 2, \ldots, N_a)$ in the adjoint representation (the **3**). Representing $(\Phi_n)^{\bar{j}}_i$ as a traceless 2×2 complex matrix, the D-term conditions are the matrix equation

$$\sum_n [\Phi^\dagger_n, \Phi_n]^{\bar{j}}_i = 0 \,.$$

In the case of $N_a = 1$, appropriate to pure $\mathcal{N} = 2$ gauge theory, the solution is clearly that Φ is diagonal

$$\langle \Phi \rangle = \begin{bmatrix} a & 0 \\ 0 & -a \end{bmatrix}$$

This breaks the $SU(2)$ gauge group to $U(1)$. In terms of the one independent gauge invariant operator which can be built from Φ

$$\langle u \rangle \equiv \langle \mathrm{tr}\, \Phi^2 \rangle = 2a^2$$

the moduli space of the theory is the complex u plane, classically. The Kähler potential for u has a singularity at $u = 0$, where the gauge group is unbroken. Let's check the counting: Φ is a triplet, and two of its components are eaten when $SU(2)$ breaks to $U(1)$, leaving one — the chiral multiplet u.

Now suppose there are three fields Φ_n, $n = 1, 2, 3$, in the **3** of $SU(2)$, and also a superpotential $W = \mathrm{tr}\, \Phi_1[\Phi_2, \Phi_3]$. This is the $\mathcal{N} = 4$ $SU(2)$

[n]More precisely, they are related by scale invariance (since the vevs are the only scales in the classical $d = 4$ gauge theory) but since scale invariance will be broken quantum mechanically, while the D-term conditions generally will not be altered, we will see that the vacua really are physically very different.

gauge theory. The easiest way to establish the solutions to the D-term and F-term constraints is to rewrite the fields as

$$\Phi_1 = X_4 + iX_7 \ , \ \Phi_2 = X_5 + iX_8, \ , \ \Phi_3 = X_6 + iX_9$$

in terms of which the potential $V(X_m) = \sum_A (D^A)^2 + \sum_n F_n^2$ can be rewritten as

$$V(X_p) \propto \sum_{p,q=4}^{9} \mathrm{tr}\left([X_p, X_q]\right)^2$$

This formulation has the advantage that the $SO(6)$ symmetry rotating the X_p is manifest; the superpotential only exhibits a $U(3)$ subgroup of that symmetry. The condition $V = 0$ implies all of the X_p are simutaneously diagonalizable. In short, the solutions are

$$X_p = \begin{bmatrix} c_p & 0 \\ 0 & -c_p \end{bmatrix}$$

with $(c_4, \cdots, c_9)$ forming a real six-vector living in a flat six-real-dimensional moduli space. Again, with the exception of the point $c_p = 0$, all of the vacua have $SU(2)$ broken to $U(1)$.[o]

Let's add a mass m tr $\Phi_2\Phi_3$ (which breaks $\mathcal{N} = 4$ but preserves $\mathcal{N} = 2$) and integrate out the massive fields. Their equations of motion, which at low momenta reduce simply to

$$\frac{\partial W}{\partial \Phi_2} = \sqrt{2}[\Phi_1, \Phi_3] + m\Phi_3 = 0; \frac{\partial W}{\partial \Phi_3} = -\sqrt{2}[\Phi_1, \Phi_2] + m\Phi_2 = 0;$$

only have a solution $\Phi_2 = \Phi_3 = 0$. When substituted back into the superpotential, this solution gives a low-energy theory with one adjoint Φ_1 and $W = 0$ — the pure $SU(2)$ $\mathcal{N} = 2$ theory. Thus it is easy to flow from the $\mathcal{N} = 4$ theory to the pure $\mathcal{N} = 2$ theory, and this was studied in the $SU(2)$ case by Seiberg and Witten (1994).

By contrast, consider adding a mass $\frac{1}{2}m$ tr Φ_3^2. The situation here is much like the XYZ model with a mass for X; the low-energy theory there had $W = Y^2Z^2$. Here the equation of motion for Φ_3 reduces to

$$\sqrt{2}[\Phi_1, \Phi_2] = m\Phi_3$$

leaving a superpotential

$$W_L = p([\Phi_1, \Phi_2])^2 \tag{31}$$

[o]This generalizes: for $SU(N)$, we have $N - 1$ such real six-vectors, corresponding to the $N - 1$ sextuplets of eigenvalues of the traceless matrices X_p.

where $p \propto 1/m$ is a coupling with classical dimension -1. Like $(YZ)^2$, it is an irrelevant operator in four dimensions, scaling to zero in the infrared.

5.2. *Beta functions and fixed points*

Now we turn to the quantum mechanics of these theories. We have already studied the abelian case in great detail at the one-loop level, and we know that the main difference in the nonabelian case will be that gauge boson loops will give a negative contribution to the beta function. The contributions to the beta function of various particles are shown in Fig. 35, from which one can see that a vector multiplet contributes a factor of $-3N$

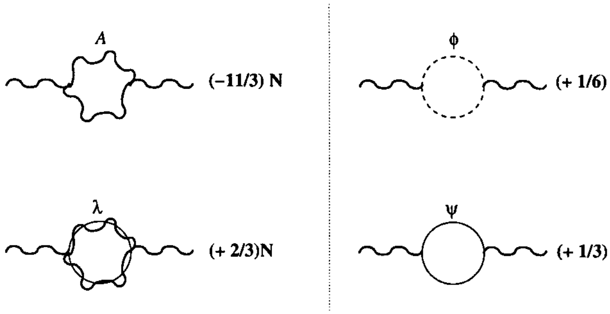


Figure 35. The contribution of various particles to the gauge beta function.

to the one-loop beta function, while a chiral multiplet in the fundamental or antifundamental representation contributes 1/2, and one in the adjoint representation gives a factor N. Thus for supersymmetric QCD with N_f quarks/squarks and N_f antiquarks/antisquarks, the beta function at one-loop is (see the abelian case)

$$\beta_g = -\frac{g^3}{8\pi^2}(3N - N_f) \Rightarrow \frac{1}{g^2(\mu)} = \frac{3N - N_f}{8\pi^2} \ln\left(\frac{\Lambda}{\mu}\right) \tag{32}$$

where (defining $b_0 = 3N - N_f$)

$$\Lambda^{b_0} = \mu^{b_0} e^{-8\pi^2/g^2 + i\theta} = \mu^{b_0} e^{2\pi i \tau} \tag{33}$$

For N_a adjoint fields the one-loop beta function has $b_0 = (3 - N_a)N$; note it vanishes for $\mathcal{N} = 4$ Yang-Mills. Just as in the abelian case, the entry of the theta angle into Λ implies there can be no perturbative corrections to these formulas.

As an aside, let me stress that the existence of the quantum mechanical holomorphic parameter Λ is very important. Although the effective superpotential is still constrained by the perturbative nonrenormalization theorem, as in non-gauge theories, the presence of Λ permits, in many cases, a nonperturbative renormalization of the superpotential. There is a lot of literature on this subject; the classic paper is that of Affleck, Dine and Seiberg (1984), although a number of others obtained similar results using somewhat less reliable techniques. There are good reviews on this subject by other authors, including ones by Intriligator and Seiberg and one by Argyres. For lack of time, in these lectures we will *only* discuss cases where there is *no* nonperturbative correction to the effective superpotential (or at least no qualitative change in its structure.) Large classes of interesting theories have this property, but we are leaving out other large classes; see the Appendix for some examples.

As before, the formula for the holomorphic coupling $g(\mu)$, and the holomorphic renormalization scale Λ, can't represent the physical properties of the theory. This is obvious from the fact that $g^2(\mu)$ blows up at small μ and can't make sense below Λ. Higher loop effects, and possibly nonperturbative effects, appearing in the nonholomorphic parts of the theory can change this formula significantly. How can we define a physical gauge coupling? A natural approach is to find a more physical definition of Λ, so let the holomorphic object in Eq. (33) be renamed $\hat{\Lambda}$, and let us attempt to define a Λ independent of field redefinitions. (The presentation here is related to recent work of Arkani-Hamed and Rattazzi, although the resulting formula is due to Novikov, Shifman, Vainshtein and Zakharov from the early 1980s.)

Recall how we defined a physical version of y, the coupling in the Wess-Zumino model. We noted that if we sent $\Phi \to a\Phi$, a a constant, this would affect both $\hat{y}$ in the superpotential and Z in the Kähler potential. Let's do the same here for the charged fields Q and $\tilde{Q}$. Suppose we multiply them all of them by a, where a is a *phase* $e^{i\alpha}$. This is equivalent to a transformation by an anomalous "axial" $U(1)$ global symmetry, under which quarks and antiquarks have the same charge. As happens in QCD, this kind of transformation is an anomalous symmetry, and shifts the θ angle; it therefore rotates Λ^{b_0} by a phase. The phase by which Λ^{b_0} rotates is a^{2N_f}. But since Q and Λ are holomorphic, it must still be true that Λ changes by a^{2N_f} even if a is *not* a phase but has $|a| \neq 1$! Therefore, since $Z \to |a|^{-2}Z$ under this transformation, only

$$(\hat{\Lambda}^{b_0})^\dagger \left[\prod_{r=1}^{N_f} Z_r\right] \left[\prod_{s=1}^{N_f} \tilde{Z}_s\right] \hat{\Lambda}^{b_0} \ ,$$

where Z_r and $\tilde{Z}_s$ are the wave function factors for Q_r and $\tilde{Q}_s$, can be invariant under these field redefinitions.

We can go further by considering R-symmetry transformations. Suppose we rotate the gluino fields λ by $e^{i\alpha}$ and the fields $q, \tilde{q}$ by $e^{i\alpha}$, so that the quarks ψ_q and antiquarks $\tilde{\psi}_q$ do not rotate at all (recall there is a $\lambda\psi_r q_r^\dagger$ interaction which fixes the charge of one field in terms of the other two.) Then the anomaly in this transformation involves only the gluinos, and the θ angle shifts by $e^{2Ni\alpha}$. (The factor $2N$ is the group theory "index" of the adjoint representation in $SU(N)$; it determines the size of the anomaly.) Again, we can generalize this by taking α to be complex, so that $|e^{i\alpha}| \neq 1$. An invariant which is unchanged by both this and the previous transformation is

$$\mu^{2b_0} e^{-16\pi^2/g^2(\mu)} = |\Lambda^2(\mu)| = (\hat{\Lambda}^{b_0})^\dagger Z_\lambda^{2N} \left[\prod_r Z_r\right] \left[\prod_s \tilde{Z}_s\right] \hat{\Lambda}^{b_0}$$

Taking a derivative with respect to μ of both sides, we obtain

$$\beta_{\frac{8\pi^2}{g^2}} = b_0 + \frac{1}{2}\sum_r \gamma_r + \frac{1}{2}\sum_s \gamma_s + N\gamma_\lambda$$

But from the kinetic terms of the gauginos it is evident that $Z_\lambda = 1/g^2(\mu)$ itself! Therefore

$$\gamma_\lambda = -\frac{\partial \ln Z_\lambda}{\partial \ln \mu} = \frac{\beta_{8\pi^2/g^2}}{8\pi^2/g^2}$$

from which we obtain the *exact* NSVZ beta function

$$\beta_{\frac{8\pi^2}{g^2}} = \frac{b_0 + \frac{1}{2}\sum_r \gamma_r + \frac{1}{2}\sum_s \gamma_s}{1 - g^2 N/8\pi^2} = -\frac{16\pi^2}{g^3}\beta_g$$

(Remember to keep track of the difference in sign between β_g and $\beta_{8\pi^2/g^2}$.) In supersymmetric QCD, where in the absence of a superpotential all charged fields are related by symmetry, and therefore have the same anomalous dimension γ_0, we may write

$$\beta_{\frac{8\pi^2}{g^2}} = \frac{3N - N_f[1-\gamma_0]}{1 - g^2 N/8\pi^2}$$

In a general theory with charged fields ϕ_i in representations R_i with $\text{tr}\ T^A T^B = T_{R_i}\delta^{AB}$, and with anomalous dimensions γ_i, we have

$$\beta_{\frac{8\pi^2}{g^2}} = \frac{3N - \sum_i T_{R_i}[1-\gamma_i]}{1 - g^2N/8\pi^2} = \frac{b_0 + \sum_i T_{R_i}\gamma_i}{1 - g^2N/8\pi^2}$$

These formulas continue to hold even when there are other gauge and matter couplings in the theory. This formula, which has a Taylor expansion in g^2, summarizes all higher-loop corrections. If there is matter in the theory, then we do not know γ_i exactly; but in the absence of matter (the pure $\mathcal{N}=1$ gauge theory) the formula is a definite function of g.[p]

We will now use this exact beta function to prove a few things. Before doing so, let's consider the one-loop contributions to the anomalous dimensions of charged fields. As we saw, trilinear terms in the superpotential give one-loop graphs as in Fig. 36 which *must* give a positive contribution,

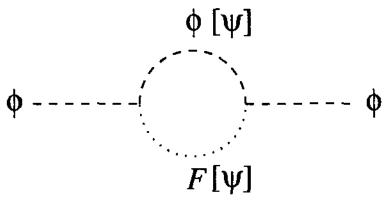


Figure 36. Typical contributions to Z_Φ from the superpotential.

if there are no gauge couplings around. To leading order in the gauge coupling, this must still be true; the one-loop graph from a superpotential term must be positive. However, there is no such constraint for the one-loop diagram in Fig. 37 involving the gauge interactions and it is a determining, crucial feature of supersymmetric gauge theories that the coefficient of this diagram has the opposite sign. (I don't know of an argument which explains this fact in physical terms.) Consequently the sign of γ will flip as the coupling constants are varied. For a gauge theory with no superpotential, the charged fields have negative anomalous dimensions.

First, let us prove that for large N_f, slightly less than $3N$, SQCD has nontrivial conformal fixed points (and does not for $N_f \geq 3N$). These are

[p]The pole in the denominator is still not fully understood, even after 20 years; notice that it becomes dominant when $g^2N \sim 8\pi^2$, an important issue for those studying large-N gauge theories at large 't Hooft coupling!

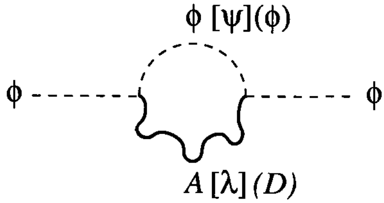


Figure 37. Contributions to Z_Φ from the gauge interactions.

sometimes called "Banks-Zaks" fixed points although they were discussed earlier; they exist also in the nonsupersymmetric case!

Exercise: By examining the one- and two-loop beta functions (given in Ellis's lectures at this school) verify that ordinary QCD has conformal fixed points when N_f and N_c are large and the theory is just barely asymptotically free. Remember to check that all higher-loop terms in the beta-function can be neglected!

To see this, note that the loop expansion is really an expansion in g^2N, so if we can trap g^2 in a region where it is of order $1/N^2$, then the loop expansion can be terminated at leading nonvanishing order. The anomalous dimension γ_0 of the superfields $Q_r, \tilde{Q}_s$ will be of the form, on general grounds,

$$-\hat{c}\frac{g^2N}{8\pi^2} + \text{order}\left[(g^2N)^2\right]$$

where $\hat{c} > 0$. For $N_f = 3N - k$, where k is order 1, the beta function takes the form

$$\beta_{\frac{8\pi^2}{g^2}} = \frac{k + \frac{N_f}{8\pi^2}\left(-\hat{c}g^2N + \text{order}\left[(g^2N)^2\right]\right)}{1 - g^2N/8\pi^2} \approx k - 3N\hat{c}\frac{g^2N}{8\pi^2}$$

where in the last expression we have dropped terms of order g^2Nk and $N(g^2N)^2$. For $k \leq 0$, that is, $N_f \geq 3N$, the beta function β_g is positive (*i.e.* $\beta_{8\pi^2/g^2} < 0$) for small g, so the gauge coupling g flows back to zero in the *infrared* and is not asymptotically free in the ultraviolet. However, if $k > 0$, and thus for $N_f < 3N$, the beta function β_g is negative at small g but has a zero at

$$g_*^2 = \frac{8\pi^2 k}{3\hat{c}N^2}$$

and therefore the coupling g^2 never gets larger than of order $1/N^2$. The above formula is therefore self-consistent in predicting that the gauge coupling flows from zero to the above fixed point value. (Note that the holomorphic coupling has no such fixed point! thus the physical properties and holomorphic properties of the theory are vastly different!) The flow is shown in Fig. 38.

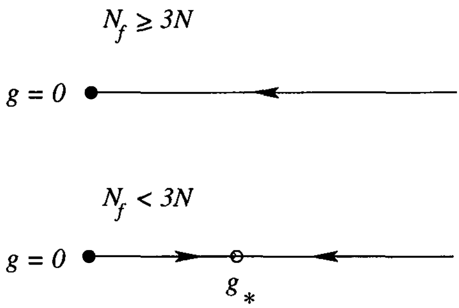


Figure 38. The gauge coupling is marginally relevant for $N_f < 3N$, with a nearby fixed point at g_*.

Next, let's prove that in SQCD for $N_f \leq \frac{3}{2}N$ there can be no such fixed points — specifically, ones in which no fields are free, and which are located at the origin of moduli space, where no fields have expectation values. In SQCD a fixed point requires

$$\beta_{\frac{8\pi^2}{g^2}} \propto b_0 + N_f\gamma_0 = 0 \;\Rightarrow \gamma_0 = \frac{3N - N_f}{N_f} \; .$$

(This in turn implies that the R-charge of Q and $\tilde{Q}$, using the earlier formula that $\dim Q = 1 + \frac{1}{2}\gamma_0 = \frac{3}{2}R_Q$, must be

$$R_Q = 1 - \frac{N}{N_f} \; ; \tag{34}$$

one may check that this particular R-symmetry is the unique nonanomalous chiral symmetry of the theory!) However, if $\gamma_0 \leq -1$, then the gauge-invariant operator $Q_r\tilde{Q}^s$ would have dimension $2(1 + \frac{1}{2}\gamma_0) \leq 1$. This is not allowed at a nontrivial fixed point, so to have such a fixed point (at least one in the simple class we have been discussing) it must be that

$$\gamma_0 > -1 \quad \Rightarrow \quad b_0 < N_f \quad \Rightarrow \quad N_f > \frac{3}{2}N$$

What happens for adjoint fields? If the number of fields is N_a, then we have

$$\beta_{\frac{8\pi^2}{g^2}} \propto (3 - N_a)N + N_a\gamma_0 = 0 \Rightarrow \gamma_0 = 1 - \frac{3}{N_a} \ .$$

This is too negative a γ_0 for $N_a = 1$, so there is no ordinary $SU(N)$ fixed point in the pure $\mathcal{N} = 2$ theory.[q] For $N_a = 2$, the anomalous dimension at any fixed point must be $-\frac{1}{2}$; that is, the dimension of Φ_1 and Φ_2 must be $\frac{3}{4}$. For $N_a = 3$ the anomalous dimension at any fixed point must be zero. If Φ_n were gauge-invariant, then by the earlier theorem this could only happen if Φ were free; but outside of the abelian case, Φ_n is not gauge-invariant and this is not a requirement.

__Exercise:__ In an $\mathcal{N} = 2$ gauge theory, the one-loop formula $\beta_{8\pi^2/g^2} = b_0$ is exact. Using the facts that (a) the anomalous dimension of the adjoint field Φ is related to that of the gauge bosons by $\mathcal{N} = 2$ supersymmetry, and (b) $\mathcal{N} = 2$ forbids hypermultiplets to have anomalous dimensions, prove that the NSVZ beta function is consistent with this statement.

Now, Seiberg has suggested that there are fixed points in $\mathcal{N} = 1$ SQCD for $3N > N_f > \frac{3}{2}N$. That is, he conjectured in 1994 that in this range there is some value g_* of the gauge coupling for which $\gamma_0(g_*) = \frac{3N-N_f}{N_f}$. (Again, we know this is true for $3N - N_f \ll N$; Seiberg's conjecture is that this continues down to much lower N_f.) If he is right, as most experts think that he is, then some remarkable and exciting phenomena immediately follow. These can be found by combining our minimal knowledge concerning the properties of these putative fixed points with the approach to the renormalization group outlined in the previous lectures.

For example, consider adding the superpotential

$$W = \hat{p} \sum_{r,s=1}^{N_f} (Q_r\tilde{Q}^s)(Q_s\tilde{Q}^r) \tag{35}$$

with gauge indices contracted inside the parentheses. Very importantly, this superpotential preserves a diagonal $SU(N_f)$ global symmetry and charge conjugation; this is enough symmetry to ensure that all of the fields share the same anomalous dimension $\gamma_0(p, \tau)$, as was true for $p = 0$. As always

[q]There are however some much more subtle fixed points discovered by Argyres and Douglas in 1995.

we should study the dimensionless coupling constant $\pi \equiv p\mu$ and ask how it scales. Classically it scales like μ (and thus has $\beta_\pi = \pi$) but quantum mechanically, —

Wait a minute. This theory, whose potential contains (scalar)^6 terms, is nonrenormalizable. Can we even discuss it?

Well, nonrenormalizable simply means that the operator in the superpotential is irrelevant, so in the ultraviolet regime the effective coupling is blowing up and perturbative diagrams in the theory don't make sense. But we're interested in the infrared anyway. We'll deal with the ultraviolet later; for now will think of $1/\hat{p}$ as setting a cutoff on the theory.[r] Perturbation theory may not converge, but we are asking perfectly valid nonperturbative infrared questions which do not depend on the details of the ultraviolet cutoff.

In particular, we know that we need to define a physical coupling π, of the form $|\pi^2| = |\hat{p}\mu|^2 Z_0^{-4}$. We see it has a beta function

$$\beta_\pi = \pi[1 + 2\gamma_0]$$

Now, this means π is irrelevant if $\gamma_0 > -\frac{1}{2}$ and is relevant if $\gamma_0 < -\frac{1}{2}$. The formula for the gauge coupling is unchanged

$$\beta_{\frac{8\pi^2}{g^2}} \propto 3N - N_f + N_f\gamma_0 \,.$$

Now, remember that γ_0 is a function of τ and π with the following properties: (1) if $g = 0$, $\pi \neq 0$ then $\gamma_0 > 0$; (2) if $\pi = 0$, $0 \neq g \ll 1$ then $\gamma_0 < 0$; and (3) there is at least one nontrivial fixed point at $g = g_*$, $\pi = 0$ with $\gamma_0 = \frac{3N-N_f}{N_f}$. Notice that at this fixed point π is irrelevant (as it is classically) if $N > 2N_f$, marginal if $N = 2N_f$, and *relevant* if $N < 2N_f$.

From this we can guess the qualitative features of the renormalization group flow. For $N > 2N_f$, the qualitative picture is given in Fig. 39. Even if $\pi \neq 0$, we still end up at the Seiberg fixed point. For $N < 2N_f$, however, there is a very different picture, as in Fig. 40. Notice that if we start at weak gauge coupling initially, π is irrelevant and flows toward zero as we would expect classically; but as we flow toward the infrared, the gauge coupling grows, γ_0 becomes more negative, and eventually the coupling π turns around and becomes *relevant*. Although at first it seems as though it will be negligible in the infrared, it in fact *dominates*. This is called a

[r]Note that we did essentially the same thing with SQED in four dimensions, which is *perturbatively renormalizable* but *nonperturbatively nonrenormalizable*, since we cannot take the cutoff on the theory to infinity without the gauge coupling diverging in the ultraviolet.

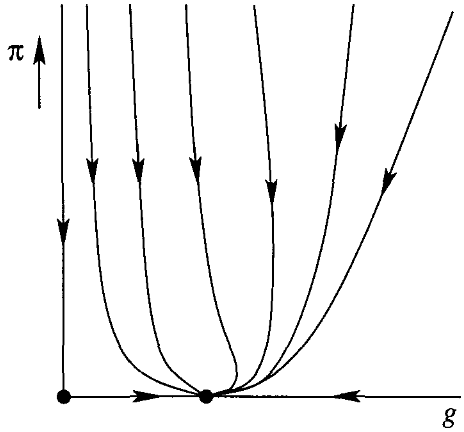


Figure 39. For $N_f > 2N$ the coupling π is irrelevant both at $g = 0$ and at $g = g_*$.

"dangerous irrelevant" operator, since although it is initially irrelevant it is dangerous to forget about it! In the infrared it becomes large, and we must be more precise about what happens when it gets there.[s]

What about $N_f = 2N$? You'll do this as an exercise, after I've done a bit more.

5.3. *Using $\mathcal{N} = 1$ language to understand $\mathcal{N} = 4$*

As another application of these ideas, let's argue that $\mathcal{N} = 4$ Yang-Mills is finite. Consider an $\mathcal{N} = 1$ gauge theory with three chiral superfields and a superpotential $W = h \text{ tr } \Phi_1[\Phi_2, \Phi_3]$. I will use canonical normalization here for the Φ_n, so $h = g$ is the $\mathcal{N} = 4$ supersymmetric theory. But let's not assume that $h = g$. For any g, h, the symmetry relating the three fields ensures they all have the same anomalous dimension γ_0, which is a single function of two couplings. The beta functions for the couplings are

$$\beta_h = \frac{3}{2} h\gamma_0 \; ; \; \beta_{g^2} = \frac{-g^4}{16\pi^2} \frac{\gamma_0}{1 - g^2 N/8\pi^2}$$

These are proportional to one another, so the conditions for a fixed point ($\beta_h = 0$ and $\beta_g = 0$) reduce to a single equation, $\gamma_0(g, h) = 0$. But this is

[s]What happens is fascinating — the theory flows to a *different* Seiberg fixed point, that given by $SU(N_f - N)$ SQCD with N_f flavors. See Leigh and Strassler (1995) for an understanding of how to treat the limit where the coupling is large.

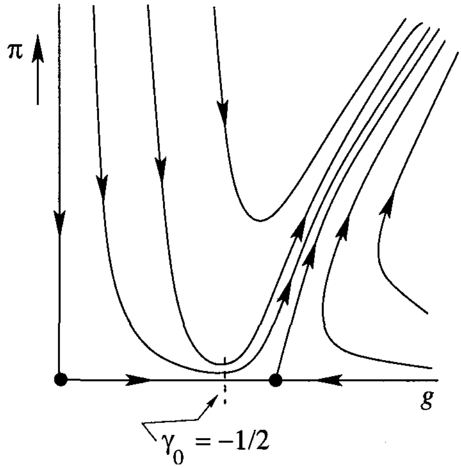


Figure 40. For $N_f < 2N$ the coupling π is *relevant* at g_*; its initial decrease is reversed once g is sufficiently large.

one equation on two variables, *so if a solution exists, it will be a part of a one-dimensional space of such solutions.*

Now, does a solution exist? We know that $\gamma(g, h = 0) < 0$ and that $\gamma_0(g = 0, h) > 0$; so yes, by continuity, there must be a curve, passing through $g = h = 0$, along which $\gamma_0 = 0$ and $\beta_h = \beta_g = 0$ (and thus perturbation theory has *no* infinities along this line.) The renormalization group flow must look like the graph in Fig. 41. Both the theory with $h = 0$ and the theory with $g = 0$ are infrared free; yet a set of nontrivial field theories lies between. Notice that we do not know, however, the precise position of the curve $\gamma_0 = 0$. In particular, we have not shown that $g = h$ gives $\gamma_0 = 0$. However, the existence of a finite theory (which is renormalization-group stable in the infrared) requires only arguments using $\mathcal{N} = 1$ symmetry. Of course, since the theory at $g = h$ has more symmetry (namely $\mathcal{N} = 4$) it is natural to expect $g = h$ to be the solution to $\gamma_0(g, h) = 0$.

The motivation for introducing this $\mathcal{N} = 1$-based reasoning is there are many $\mathcal{N} = 1$ field theories which are also finite, as one can show using similar arguments. (For example, replace the $\mathcal{N} = 4$ superpotential with $W = h \text{ tr } \Phi_1\{\Phi_2, \Phi_3\}$; the discussion is almost unchanged, except that $g = h$ is not the solution to $\gamma_0 = 0$.) The existence of these theories was discovered in the 1980s; the slick proof presented above is in Leigh and Strassler (1995).

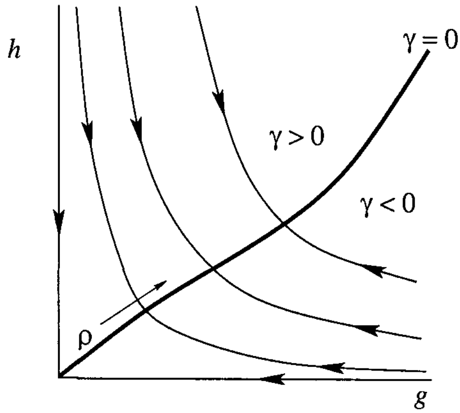


Figure 41. In some $\mathcal{N} = 1$ theories one can argue for a line of fixed points indexed by an exactly marginal coupling ρ; perturbation theory has no divergences on this line. Only in $\mathcal{N} = 4$ is the equation for this line $g = h$.

As mentioned in Sec. 3.4, the coupling which parametrizes the line of conformal fixed points (which is actually a complex line, since the couplings are complex) is called an "exactly marginal coupling." Let's call this complex coupling ρ. (In the $\mathcal{N} = 4$ case we can identify ρ as equal to the gauge coupling i/τ, but in a more general $\mathcal{N} = 1$ finite theory these will not be simply related.) Unlike λ in $\lambda\phi^4$, which is marginal at $\lambda = 0$ but irrelevant at $\lambda \neq 0$, ρ is marginal at $\rho = 0$, and remains marginal for any value of ρ. Thus ρ is a truly dimensionless coupling, indexing a continuous class of scale-invariant theories. It is very common for such classes of theories to be acted upon by duality transformations. In fact, for $\mathcal{N} = 4$ electric-magnetic duality (S-duality) acts on this coupling ρ as in Fig. 42, identifying those theories at large ρ with those at small ρ.

5.4. *The two-adjoint model*

We conclude with a discussion of a theory with two adjoint chiral multiplets, obtained from the $\mathcal{N} = 4$ gauge theory by adding a mass for the third adjoint Φ_3. It has a superpotential (31):

$$W_L = p([\Phi_1, \Phi_2])^2 \ .$$

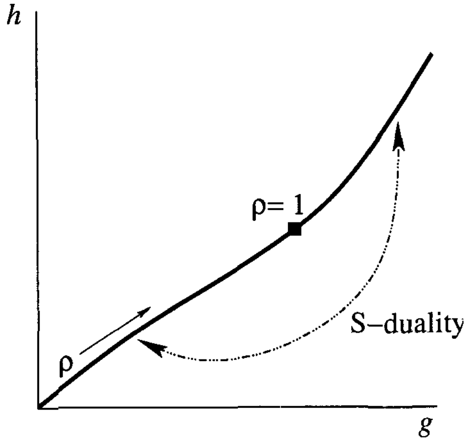


Figure 42. The action of S-duality on the line of fixed points.

This quartic superpotential is nonrenormalizable, but low-energy effective theories often are. We are interested (for the moment) in the infrared behavior, so the fact that $\pi = p\mu \propto \nu^{-1}$ (where $\nu = \frac{m}{\mu}$) blows up in the ultraviolet is not our immediate concern.

What are the beta functions for π and for g? We have

$$\beta_\pi = \pi[1 + 2\gamma_0] \ ; \ \beta_{8\pi^2/g^2} \propto 3N - 2N(1-\gamma_0) = N + 2N\gamma_0$$

and thus $\beta_\pi \propto \beta_g$. This means that, as before, the conditions for a fixed point to exist, namely $\beta_\pi = 0 = \beta_g$, reduce to a single condition (except at $g = \pi = 0$):

$$1 + 2\gamma_0 = 0 \ \Rightarrow \ \gamma_0 = -\frac{1}{2} \ .$$

Again, this is one condition on two couplings, *so any solution will be part of a one-dimensional space of solutions.* Following Seiberg, we might well expect that the theory with two adjoints and $W = 0$ has a fixed point at some g_* where $\gamma_0(g_*) = -\frac{1}{2}$. If this is true, then the renormalization group flow of the theory will look like Fig. 43.

So here our irrelevant operator has been converted into an exactly marginal one! The coupling ρ which parametrizes the line of fixed points, and on which duality symmetries might act, now has nothing to do with the gauge coupling. In fact $\rho = 0$ corresponds to $g = g_*$. Nowhere are these

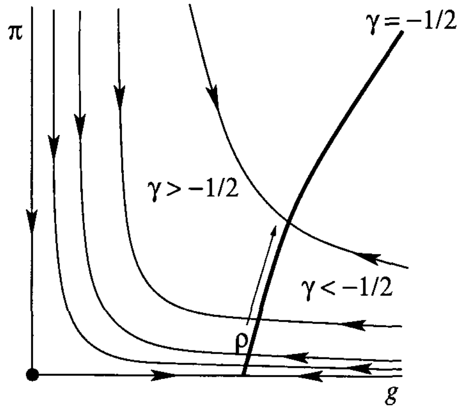


Figure 43. In the two-adjoint theory, π is marginal at $g = g_*$ and there is a line of fixed points emerging from the fixed point at $(g, \pi) = (g_*, 0)$.

conformal field theories near $g = \pi = 0$, so we have no hope of seeing them in any perturbation expansion.

Now, since the theory is nonrenormalizable, we probably should at least say something about the ultraviolet. But we know a perfectly good ultraviolet theory into which we can embed this theory, namely the $\mathcal{N} = 4$ gauge theory with Φ_3 massive. Classically, we know what this flow would look like. Just as in the XYZ model with X massive, shown in Fig. 5, the theory would start from a conformal field theory indexed by τ and flow into a classical fixed point, with a nonzero gauge coupling and $W = 0$, along the irrelevant operator $([\Phi_1, \Phi_2])^2$. But quantum mechanically the endpoint of the theory is not $W = 0$; instead, it is one of the conformal field theories we found above in the two-adjoint theory. In fact, we can expect that each $\mathcal{N} = 4$ field theory flows to a unique two-adjoint theory, along a flow which looks schematically like Fig. 44. It is natural therefore to identify ρ with i/τ, as we did in the $\mathcal{N} = 4$ case, but this "τ" is not the gauge coupling of the two-adjoint theory. Rather, we have defined here a physical mechanism for using the label τ of the $\mathcal{N} = 4$ fixed points as a label for the two-adjoint fixed points. Since S-duality acts on τ in $\mathcal{N} = 4$ we are essentially guaranteed that duality will also act on ρ in the two-adjoint theory.

Exercise: Examine the beta functions and sketch the renormalization group flow for SQCD with $N_f = 2N$ and the quartic superpotential (35).

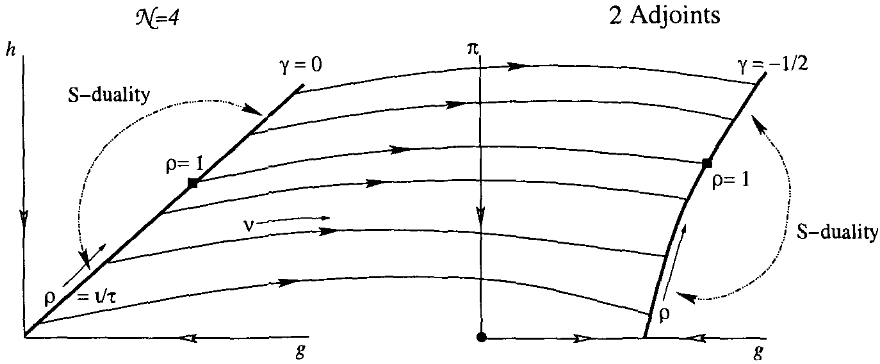


Figure 44. The two-adjoint theory inherits S-duality from the $\mathcal{N} = 4$ theory through the flow which takes one to the other.

6. Regrets

What didn't I talk about in these lectures? The list seems to be infinite. There was nothing on the phases of gauge theories; nothing on confinement; nothing on nonperturbative renormalizations of superpotentials; nothing on Seiberg duality; nothing on two, five or six dimensions; nothing on spontaneous dynamical supersymmetry breaking; nothing on exact methods for studying renormalization in theories with broken supersymmetry; nothing on D-brane or other stringy constructions of these theories; and above all, nothing on applications of this material to real-world physics! But there are good reviews on almost all of these subjects. By contrast, there are no reviews on the material I have discussed here, which is necessary for an understanding of duality, plays a very important role in the AdS/CFT correspondance, and (as Ann Nelson and I have suggested) may even be responsible for the pattern of quark and lepton masses and for the low rate of proton decay. So I hope that this somewhat unorthodox introduction to this subject will serve you well; and I hope I have convinced you that this is a profound and fascinating subject, where much is to be learned and much remains to be understood.

7. Appendix: Comments on $\mathcal{N} = 1$ $SU(2)$ SQCD

We will focus our attention on the case of $SU(2)$. This case is a bit special because the **2** and $\bar{\mathbf{2}}$ representation are identical [you already know this from quantum mechanics: there is no conjugate-spin-1/2 representation of $SU(2)$] so we actually should combine the Q_r and $\tilde{Q}^s$ into $2N_f$ fields Q_u, with an $SU(2N_f)$ global symmetry, and a D-term condition

$$\left[\sum_{u=1}^{2N_f}(q_u^\dagger)(q_u)\right]^i_{\bar{j}} = c_0\delta^i_{\bar{j}} \tag{36}$$

One solution to this condition is $q_1 = (\sqrt{c_0/2}, 0)$, $q_2 = (0, \sqrt{c_0/2})$, with all others zero. As in the abelian case, it is most convenient to express this result using gauge invariant combinations of the chiral superfields. The $2N_f \times 2N_f$ antisymmetric matrix of gauge-invariants $M_{uv} = Q^i_u Q^j_v \epsilon_{ij}$ has $M_{12} = -M_{21} = c_0/2$, with all other components zero.

In fact, all solutions to the condition (36) can be written as $SU(2N_f)$-flavor and $SU(2)$-gauge rotations of the above particular solution. The gauge rotations leave M_{uv} invariant, and the flavor rotations leave invariant the fact that it has rank at most two, with either zero or two equal non-vanishing eigenvalues.

Note that unless $M_{uv} = 0$, the gauge group is completely broken. Let's check this is the case for $N_f = 1$. There are two chiral fields Q_1 and Q_2, each in the **2** of $SU(2)$, for a total of four complex fields. Three of these must be eaten by the three gauge bosons if $SU(2)$ is completely broken. Consequently, there should be one remaining. Indeed, there is only one (unconstrained) field M_{12}. Let's check it for $N_f = 2$: in this case there are six fields M_{uv} $(u, v = 1, 2, 3, 4)$ but also a single constraint that the rank must be 2, not 4, which is the condition $\mathrm{Pf}(M) = 0$. (The Pfaffian is just the square root of the determinant, and is defined as $\epsilon_{uvwx}M^{uv}M^{wx}$ in this case.) This leaves five unconstrained fields. Initially there are four doublets Q_1, Q_2, Q_3, Q_4 for a total of eight fields, with three being eaten when the gauge group is broken; this too leaves five.

Exercise: For $SU(3)$ the Q_r and $\tilde{Q}^s$ are in distinct representations. The allowed operators are $M_r^s = Q_r\tilde{Q}^s$, $B = QQQ$ and $\tilde{B} = \tilde{Q}\tilde{Q}\tilde{Q}$ (indices suppressed.) Show that the conditions we have just obtained from the $SU(N)$ D-terms imply that for $N_f < N$ the rank of M is N_f or less; for $N_f = N$ $\det M = B\tilde{B}$; and for $N_f > N$ the rank of M is N or less. Show also that for $N_f \geq N$ there are branches with $B \neq 0$ but $M = 0$, and that $B\tilde{B}$ always equals a subdeterminant of rank N of M.

What then happens quantum mechanically for $SU(2)$? Let's note that for $N_f = 1, 2, 3, 4, 5$, the nonanomalous R-charge of the Q_r is $-1, 0, \frac{1}{3}, \frac{1}{2}, \frac{3}{5}$; for $N_f \geq 6$ the theory is no longer asymptotically free. For $N_f = 4, 5$ we might well have a conformal fixed point in the infrared, but not for $N_f < 4$. We should check for renormalizations. The classical superpotential is zero; the one-loop holomorphic gauge coupling is renormalized; but neither can get any further perturbative renormalization for the reasons we discussed earlier. All of the higher-loop effects in the NSVZ beta function come through the Kähler potential. However, we did not check that nonperturbative effects were absent. In particular, while the *perturbative* superpotential cannot depend on the theta angle, this is not true *nonperturbatively*. We should therefore look for a superpotential of the form

$$W_{nonpert}(M_{uv}, \hat{\Lambda})$$

which is invariant under all of the global symmetries. The only globally-symmetric holomorphic object which we can build from M is its Pfaffian Pf M, which has dimension $2N_f$ and has R-charge $2(N_f - 2)$, and its powers. The superpotential has dimension 3 and R-charge 2, so its form is very highly constrained; in fact

$$W_{nonpert} = c \left(\frac{\text{Pf } M}{\hat{\Lambda}_{N_f}^{6-N_f}} \right)^{1/(N_f-2)} \tag{37}$$

where c is a constant, is the *only* possibility. (This was pointed out by Affleck, Dine and Seiberg in 1984.) You should check that this formula is also consistent with the *anomalous* $U(1)$ symmetries which we used to write the physical version of Λ. For this reason, the above formula even holds for $N_f = 0$, where there is no anomaly-free R-symmetry.

Now let us examine whether the coefficient c can ever be nonzero. Affleck, Dine and Seiberg pointed out that c is in fact nonzero in the case $N_f = 1$; they showed that an instanton effect does indeed give a mass to the fermion in the multiplet $M \equiv M_{12}$, and calculated it, showing that

$$W_{N_f=1} = \frac{\Lambda_1^5}{M}$$

This is rather strange; the potential

$$V(M) \sim \frac{1}{|M|^2}$$

blows up at small M (though the Kähler potential cannot be calculated there) and runs gradually to zero as $M \to \infty$ (where the gauge theory is broken at a very high scale, and thus at weak coupling, where the Kähler potential is easy to calculate.) In short, this theory has no supersymmetric vacuum except at $M = \infty$; it has a runaway instability!

However, if we add a mass for the two doublets

$$W_{classical} = mQ_1Q_2$$

then the effective superpotential becomes

$$W_{full} = \frac{\Lambda_1^5}{M} + mM$$

which has supersymmetic minima

$$M^2 = m\Lambda^5 \tag{38}$$

or in other words two vacua, $M = \pm\sqrt{m\Lambda^5}$. Notice that the superpotential in Eq. (37), for $N_f = 0$, gives $W = \pm c\sqrt{\Lambda_0^6}$, which, for $c = 2$ and the matching condition $\Lambda_0^6 = m\Lambda_1^5$, is consistent with (38). The interpretation of this result, originally due to Witten (1980), is that the pure $\mathcal{N} = 1$ $SU(2)$ gauge theory has a fermion bilinear condensate

$$\langle\lambda\lambda\rangle \propto \sqrt{\Lambda_0^6}$$

which breaks a discrete chiral symmetry, somewhat analogous to QCD's breaking of chiral symmetries, and generates a nonzero superpotential $W \propto \lambda\lambda$.

What about $N_f = 2$? In this case the theory has six mesons M_{uv} subject to the constraint Pf $M = 0$. There can be no nontrivial superpotential here built just from M, but Seiberg (1994) pointed out that it was useful to implement this constraint using a Lagrange multiplier field X, of R-charge 2 and dimension -1, in the tree-level superpotential:

$$W_{classical} = X(\text{Pf } M)$$

Then $\frac{\partial W}{\partial X} = \text{Pf } M = 0$ defines the classical moduli space. However, quantum mechanically we are allowed by the symmetries to add

$$W_{nonpert} = cX\Lambda_2^4$$

which means

$$\frac{\partial W}{\partial X} = \text{Pf } M + c\Lambda_2^4 = 0$$

so the classical moduli space is modified quantum mechanically. In particular, the symmetric point Pf $M = 0$ is removed! This means that the chiral $SU(4)$ symmetry is nowhere restored on the moduli space — there is quantum breaking of a chiral symmetry in this theory!

<u>**Exercise:**</u> By adding mass terms for the fields Q_3, Q_4 and comparing with the $N_f = 1$ case, show that c cannot be zero.

And for $N_f = 3$? Here the proposed quantum superpotential

$$W = \frac{\text{Pf } M}{\Lambda_3^3}$$

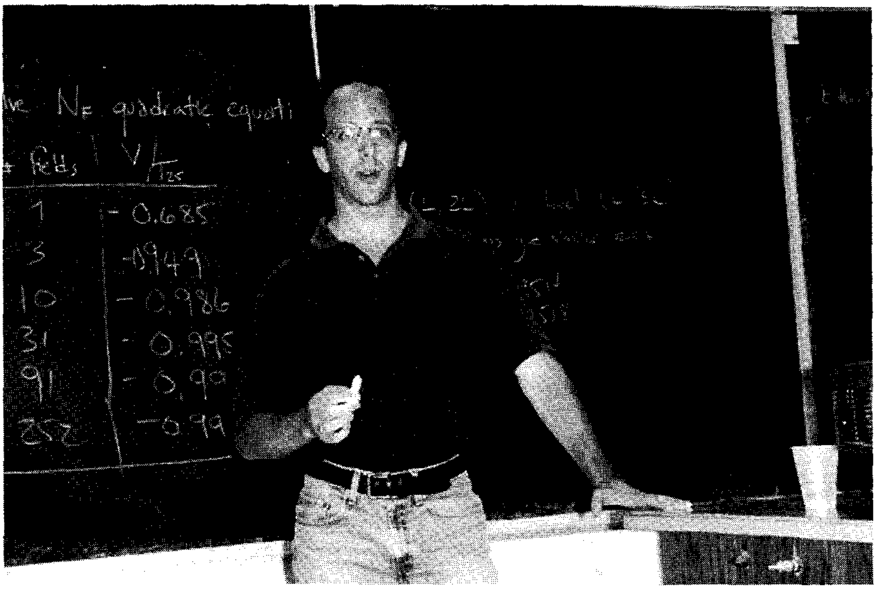
is exactly right for imposing the classical constraint that M have rank 2. The interpretation Seiberg gave is that the M fields are mesons built from confined quarks, and they have an XYZ-like superpotential quantum mechanically, one which is marginally irrelevant. In the infrared, the M fields are free, and at the origin, the $SU(6)$ symmetry is *unbroken*. This is the first example known of confinement *without* chiral symmetry breaking.

For $N_f = 4, 5$, the proposed superpotential is singular at $M = 0$, and cannot be valid there. Seiberg (1994) therefore suggested that there are nontrivial infrared fixed points at $M = 0$ for $N_f = 4, 5$. The evidence in favor of this suggestion is now very strong, although it is still not really proven beyond a shadow of doubt. Personally I don't doubt it, but I would love to see a conclusive proof someday.

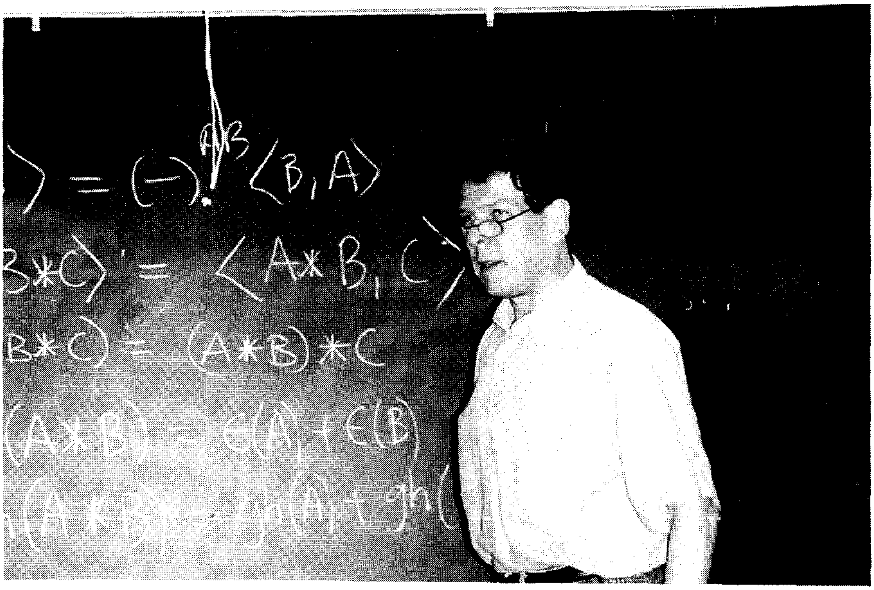
8. Suggested reading

There are many great papers, and many excellent reviews, for you to look at in your further explorations of this subject. I learned supersymmetry, the Wess-Zumino model, non-renormalization theorems, and so forth from Wess and Bagger and from West; both books have advantages and problems. Philip Argyres has a set of very useful lectures; they can be accessed from his website. Renormalization you must learn from many places; no one book does it all well. The two classic papers of Seiberg and Witten (1994) on duality in $\mathcal{N} = 2$ and the various papers of Seiberg on holomorphy and on duality (1993-1994) in $\mathcal{N} = 1$ are must-reads for everyone. There are pedagogical reviews (try Bilal (1995) for $\mathcal{N} = 2$, Intriligator and Seiberg (1995) for $\mathcal{N} = 1$) that unpack these papers somewhat. Three-dimensional supersymmetric abelian gauge theories were first studied in papers by Seiberg and Witten (1996) and by Intriligator and Seiberg (1996); see also de Boer et al. (1996, 1997), and Aharony et al. (1997). Vortex solutions appear in Nielsen and Olesen, and earlier in work of Abrikosov in the context of superconductivity. Duality is best understood by first studying the classic work on the Ising model, and by reading a lovely paper on bosonization by Burgess and Quevedo (1995). The work of the author and Kapustin (1998) follows in this spirit and points in new directions. You can also get a quick tour of duality (though not as quick as in these lectures) and vortices in my Trieste 2001 lectures. The papers of Shifman and Vainshtein, many cowritten with Novikov and Zakharov (1980-1988), painstakingly explored and finally drained the swamp surrounding the distinction between the holomorphic and physical gauge couplings. The work of Leigh and the author (1995) on exactly marginal couplings builds on their results, as well as on related results in two dimensions (see for example Martinec (1989) and Lerche, Vafa and Warner (1989).) A summary and list of references concerning recent refinements in the study of beta functions can be found in an appendix of a paper by Nelson and the author (2002).

This is as sketchy a bibliography as can be imagined; there are literally hundreds of interesting papers which are relevant to these lecture notes. Well, such is the fate of most papers that we write; we may love them dearly, but it is wise to remember that the next generation of students will never read them.



WASHINGTON TAYLOR



BARTON ZWIEBACH

This page intentionally left blank

D-BRANES, TACHYONS, AND STRING FIELD THEORY

WASHINGTON TAYLOR AND BARTON ZWIEBACH
Center for Theoretical Physics,
Massachusetts Institute of Technology,
Cambridge, MA 02139 USA
E-mail: wati@mit.edu, zwiebach@mitlns.mit.edu

In these notes we provide a pedagogical introduction to the subject of tachyon condensation in Witten's cubic bosonic open string field theory. We use both the low-energy Yang-Mills description and the language of string field theory to explain the problem of tachyon condensation on unstable D-branes. We give a self-contained introduction to open string field theory using both conformal field theory and overlap integrals. Our main subjects are the Sen conjectures on tachyon condensation in open string field theory and the evidence that supports these conjectures. We conclude with a discussion of vacuum string field theory and projectors of the star-algebra of open string fields. We comment on the possible role of string field theory in the construction of a nonperturbative formulation of string theory that captures all possible string backgrounds.

1. Introduction

The last seven years have been a very exciting time for string theory. A new understanding of nonperturbative objects in string theory, such as D-branes, has led to exciting new developments that relate string theory to physical systems such as black holes and supersymmetric gauge field theories. It has also led to the discovery of unexpected relationships between Yang-Mills theories and quantum theories of gravity such as closed superstring theories and M-theory. Finally, the analysis of unstable D-branes has elucidated the long-standing mysteries associated with the open string tachyon.

The study of unstable D-branes and tachyons has also led to the realization that string field theory contains significant non-perturbative information. This has been somewhat of a surprise. Certain forms of string field theory were known since the early 1990's, but there was no concrete evidence that they could be used to give a non-perturbative definition of string theory. The study of tachyon condensation, however, has changed our perspective. These lecture notes give an introduction to string field theory and review recent work in which unstable D-branes and their associated tachyons are described using string field theory. As we will discuss here, this work suggests that open string field theory, or some successor of it, may give a complete definition of string theory in which all possible backgrounds can be obtained from a single set of degrees of freedom. Such a formulation appears to be necessary to address questions related to vacuum selection and string cosmology.

In the rest of this section we will review briefly the current status of string theory as a whole, and summarize the goals of this set of lectures. In section 2 we review some basic aspects of D-branes. In section 3, we describe a particular D-brane configuration which exhibits a tachyonic instability. This tachyon can be seen in the low-energy Yang-Mills description of the D-brane system. We also describe a set of conjectures made by Sen in 1999, which stated that the tachyonic instability of the open bosonic string is the instability of the space-filling D25-brane. Sen suggested that open string field theory could be used to give an analytic description of this instability. In section 4 we give an introduction to Witten's bosonic open string field theory (OSFT). Section 5 gives a more detailed analytic description of this theory using the language of conformal field theory. Section 6 describes the string field theory using the oscillator approach and overlap integrals. The two approaches to OSFT described in these two sections give complementary ways of analyzing problems in string field theory. In section 7 we summarize evidence from string field theory for Sen's conjectures. In section 8 we describe "vacuum string field theory," a new version of open string field theory which arises when one attempts to directly formulate the theory around the classically stable vacuum where the D-brane has disappeared. This section also discusses important structures in string field theory, such as projectors of the star algebra of open string fields. Section 9 contains concluding remarks.

Much new work has been done in this area since these lectures were presented at TASI in 2001. Except for some references to more recent developments which are related to the topics covered, these lecture notes primarily cover work done before summer of 2001. Previous articles reviewing related work include those of Ohmori,[1] de Smet,[2] Aref'eva *et al.*,[3] Bonora *et al.*,[4] and Taylor.[5] There are a number of major related areas which we do not cover significantly or at all in these lectures. We do not have any substantial discussion on the dynamic process of tachyon decay; there has been quite a bit of work on this subject[6] since the time of these lectures in 2001. We do not discuss the Moyal approach to SFT taken recently by Bars and collaborators;[7,8,9] this work is an interesting alternative to the level-truncation method primarily used here. We also do not discuss in any detail the alternative boundary string field theory (BSFT) approach to OSFT. The BSFT approach is well suited to derive certain concrete results regarding the tachyon vacuum[10]—for example, using this approach the energy of the tachyon vacuum can be computed exactly. On the other hand, BSFT is not a completely well-defined framework, as massive string fields cannot yet be consistently incorporated into the theory.

1.1. *The status of string theory: a brief review*

To understand the significance of developments over the last seven years, it is useful to recall the status of string theory in early 1995. At that time it was clearly understood that there were five distinct ways in which a supersymmetric string could be quantized to give a microscopic definition of a theory of quantum gravity in ten dimensions. Each of these quantum string theories gives a set of rules for calculating scattering amplitudes of string states; these states describe gravitational quanta and other massless and massive particles moving in a ten-dimensional spacetime. The five superstring theories are known as the type IIA, IIB, I, heterotic $SO(32)$, and heterotic $E_8 \times E_8$ theories. While these string theories give perturbative descriptions of quantum gravity, there was little understanding in 1995 of nonperturbative aspects of these theories.

In the years between 1995 and 2002, several new ideas dramatically transformed our understanding of string theory. We now briefly summarize these ideas and mention some aspects of these developments relevant to the main topic of these lectures.

Dualities: The five different perturbative formulations of superstring theory are all related to one another through duality symmetries,[11,12] whereby the degrees of freedom in one theory can be described through a duality transformation in terms of the degrees of freedom of another theory. Some of these duality symmetries are nonperturbative, in the sense that the string coupling g in one theory is related to the inverse string coupling $1/g$ in the dual theory. The web of dualities that relate the different theories gives a picture in which, rather than describing five distinct fundamental theories, each superstring theory appears to be a particular perturbative limit of a single, still unknown, underlying theoretical structure.

M-theory: In addition to the five perturbative string theories, the web of dualities also seems to include a limit which describes a quantum theory of gravity in eleven dimensions. This new theory has been dubbed "M-theory". Although no covariant definition for M-theory has been given, this theory can be related to type IIA and heterotic $E_8 \times E_8$ string theories through compactification on a circle S^1 and the space S^1/Z_2, respectively.[13,12,14] In the relation to type IIA, for example, the compactification radius R_{11} of M-theory is equal to the product $g_s l_s$ of the string coupling g_s and the string length l_s. Thus, M-theory in flat space, which arises in the limit $R_{11} \to \infty$, can be thought of as the strong coupling limit of type IIA string theory. The field theory limit of M-theory is eleven-dimensional supergravity. It is also suspected that M-theory may be formulated as a quantum theory of membranes in eleven dimensions.[13]

Branes: In addition to strings, all five superstring theories, as well as M-theory, contain extended objects of various dimensionalities known as "branes". M-theory has M2-branes and M5-branes, which have two and five dimensions of spatial extent, respectively. (A string is a one-brane, since it has one spatial dimension.) The different superstring theories each have different sets of (stable) D-branes, special branes that are defined by Dirichlet-type boundary conditions on strings. In particular, the IIA/IIB superstring theories contain (stable) D-branes of all even/odd dimensions. Each superstring theory also has a fundamental string and a Neveu-Schwarz five-brane. The branes of one theory can be related to the branes of another through the duality transformations mentioned above. Using an appropriate sequence of dualities, any brane can be mapped to any other brane, including the string itself. This suggests that none of these objects are really any more fundamental than any others; this idea is known as "brane democracy".

M(atrix) theory and AdS/CFT: It is a remarkable consequence of the above developments that for certain asymptotic space-time backgrounds, M-theory and string theory can be completely described through supersymmetric quantum mechanics and field theories related to the low-energy description of systems of branes. The M(atrix) model of M-theory is a simple supersymmetric matrix quantum mechanics, and it is believed to capture (in light-cone coordinates) all of the physics of M-theory in asymptotically flat spacetime. In the AdS/CFT correspondence, certain maximally supersymmetric Yang-Mills theories can be used to describe closed superstring theories in asymptotic spacetime backgrounds that are the product of anti-de Sitter space and a sphere. It is believed that the Yang-Mills theories and the matrix model of M-theory, each give true nonperturbative descriptions of quantum gravity in the corresponding spacetime geometry. For reviews of M(atrix) theory and AdS/CFT, see Taylor[15] and Aharony*et. al.*[16]

Unstable D-branes and open string tachyons: This is in large part the subject of these lectures. The most recent chapter in our new understanding of nonperturbative effects in string theory has been the incorporation of unstable branes and open string tachyons into the overall framework of the theory. It has turned out that an understanding of unstable D-branes is necessary to properly describe all D-branes. This is natural from the point of view of K-theory, where brane configurations which are equivalent under the annihilation of unstable branes are identified.[17] The long-mysterious tachyon instability of open string theory has finally been given a physical interpretation: it is the instability of the D-brane that supports the existence of open strings. The instability disappears in the tachyon vacuum, in which the D-brane decays. Moreover, the belief that D-branes are solitonic solutions of string theory has been confirmed: starting with the appropriate tachyonic field theory of unstable space-filling branes, one can describe lower dimensional D-branes as solitonic solutions. Lower dimensional D-branes are thereby essentially obtained as solitons of the tachyon field theory, so, in some sense, lower-dimensional D-branes can be thought of as being made of tachyons! It has also been shown that the physics of unstable D-branes is captured by string field theory, thus making it a candidate for a non-perturbative formulation of string theory capable of describing changes of the string background.

The set of ideas just summarized have greatly increased our understanding of nonperturbative aspects of string theory. In particular M(atrix) theory and the AdS/CFT correspondences provide nonperturbative definitions of M-theory and string theory in certain asymptotic space-time backgrounds which can be used, in principle, to calculate any local result in quantum gravity. Through string field theory we have a possibly nonperturbative definition of the theory that appears to capture many open string theory backgrounds. The existing formulations of string field theory are not manifestly background independent because a background must be selected to write the theory. Nevertheless, as we discuss in these lectures, the theory describes multiple distinct backgrounds in terms of a common set of variables, so it embodies, at least partially, physical background independence. It remains to be seen if the theory incorporates full physical background independence; this requires an ability to describe all possible open string backgrounds, as well as all possible closed string backgrounds.

1.2. *The goal of these lectures*

The goal of these lectures is to describe progress towards a nonperturbative formulation of string theory that implements the physics of background independence. Open string field theory, as applied to tachyon condensation and related matters, has shown itself capable of describing non-perturbative objects in string theory, and it has demonstrated an ability to represent various open string backgrounds.

A completely background independent formulation of string theory may be needed to address fundamental questions such as: What is string theory/M-theory? How is the vacuum of string theory selected? (*i.e.*, Why can the observable low-energy universe be accurately described by the standard model of particle physics in four space-time dimensions with an apparently small but nonzero positive cosmological constant?), and other questions of a cosmological nature. Obviously, aspiring to address these questions is an ambitious undertaking, but we believe that attaining a better understanding of string field theory is a useful step in this direction. More concretely, in these lectures we will describe recent progress on open string field theory. It may be useful here to recall some basic aspects of open and closed strings and the relationship between them.

Closed strings, which are topologically equivalent to a circle S^1, give rise upon quantization to a massless set of states associated with the graviton $g_{\mu\nu}$, the dilaton φ, and the antisymmetric two-form $B_{\mu\nu}$, as well as an infinite family of massive states. For the supersymmetric closed string, further massless fields appear within the graviton supermultiplet—these are the Ramond-Ramond p-form fields $A^{(p)}_{\mu_1\cdots\mu_p}$ and the gravitini $\psi_{\mu\alpha}$. Thus, the quantum theory of closed strings is naturally associated with a theory of gravity in space-time. On the other hand, open strings, which are topologically equivalent to an interval $[0,\pi]$, give rise under quantization to a massless gauge field A_μ in space-time. The supersymmetric open string also has a massless gaugino field ψ_α. It is now understood that the endpoints of open strings must lie on a Dirichlet p-brane (Dp-brane), and that the massless open string fields describe the fluctuations of the D-brane and the gauge field living on the world-volume of the D-brane.

It may seem, therefore, that open and closed strings are quite distinct, and describe disjoint aspects of the physics in a fixed background space-time that contains some family of D-branes. At tree level, the closed strings indeed describe gravitational physics in the bulk space-time, while the open strings describe the D-brane dynamics. At the quantum level, however, the physics of open and closed strings are deeply connected. Indeed, historically open strings were discovered first through the form of their scattering amplitudes.[18] Looking at one-loop processes for open strings led to the first discovery of closed strings, which appeared as *poles* in nonplanar one-loop open string diagrams.[19,20] The fact that open string diagrams naturally contain closed string intermediate states indicates that in some sense all closed string interactions are implicitly defined by the open string diagrams. This connection underlies many of the important recent developments in string theory. In particular, the M(atrix) theory and AdS/CFT correspondences between gauge theories and quantum gravity are essentially limits in which closed string physics in a fixed space-time background is captured by the Yang-Mills limit of an open string theory on a family of branes (D0-branes for M(atrix) theory, D3-branes for the CFT that describes $AdS_5 \times S^5$, etc.)

Since quantum gravity theories in certain fixed space-time backgrounds can be described by field theory limits of open strings, we may ask if a global change of the space-time background can be described as well. If M(atrix) theory or AdS/CFT allowed for this description, it would indicate that these models may have background-independent generalizations. Unfortunately, such background changes involve the generally intractable addition of an infinite number of nonrenormalizable interactions to the field theories in question. One tractable situation arises for the addition of a constant background $B_{\mu\nu}$ field in space-time (perhaps because this closed string background is gauge equivalent to the open string background of a D-brane with a magnetic field). In the associated Yang-Mills theory, this change in the background field corresponds to replacing products of open string fields with a noncommutative star-product. The resulting theory is a noncommutative Yang-Mills theory. Such noncommutative theories are the only well-understood example of a situation where adding an infinite number of apparently nonrenormalizable terms to a field theory action leads to a sensible modification of quantum field theory (for a review of noncommutative field theory and its connection to string theory, see Douglas and Nekrasov[21]).

String field theory is a nonperturbative formulation of string theory in which the infinite family of fields associated with string excitations are described by a space-time field theory action. For open strings on a D-brane configuration, this field theory contains Yang-Mills fields and an entire hierarchy of massive string fields. Integrating out all the massive fields from the string field theory action gives rise to a nonabelian Born-Infeld action for the D-branes, which includes an infinite set of higher-order terms that arise from string theory corrections to the simple Yang-Mills action. Like the case of noncommutative field theory discussed above, the new terms appearing in this action are apparently nonrenormalizable, but the combination of terms must work together to form a sensible theory.

In the 1980's, a great deal of work was done to formulate string field theory for open and closed, bosonic and supersymmetric string theories. All of this work was based on the BRST approach to string quantization.[22,23,24,25] For the open bosonic string Witten[26] constructed an extremely elegant string field theory based on the Chern-Simons action. This cubic open string field theory (OSFT) is the primary focus of the work described in these lectures. Although this theory can be described in a simple abstract language, practical computations rapidly become complicated. The formulation of bosonic closed string field theory was completed in the early 1990s.[27,28,29,30] This theory is the natural counterpart of Witten's open string field theory, but it is more technically challenging because of its nonpolynomiality. A nonpolynomial string field theory is also required to describe in a non-singular fashion open and closed string fields.[31] For open superstrings, a cubic formulation[32] encountered some difficulties[33,34] (for which there are some proposed resolutions[35,36]), but the nonpolynomial formulation of Berkovits[37] appears to be fully consistent. Despite a substantial amount of work in string field theory in the early 90's, little insight was gained at the time concerning non-perturbative physics. Work on this subject stalled out until open string field theory was used to test the tachyon conjectures beginning in 1999.[38]

One simple feature of the 26-dimensional bosonic string has been problematic since the early days of string theory: both the open and closed bosonic strings have tachyons in their spectra, indicating that the usual perturbative vacua of these theories are unstable. In 1999, Ashoke Sen had a remarkable insight into the nature of the open bosonic string tachyon.[39] He observed that the open bosonic string theory (the so-called Veneziano model) represents open strings that end on a space-filling D25-brane. He pointed out that this D-brane is unstable, as it does not carry any conserved charge, and he suggested that the open string tachyon is in fact the unstable mode of the D25-brane. This led him to conjecture that open string field theory could be used to precisely determine a new vacuum for the open string, namely one in which the D25-brane is annihilated through condensation of the tachyonic unstable mode. Sen made several precise conjectures regarding the details of the string field theory description of this new open string vacuum. As we describe in these lectures, there is now overwhelming evidence that Sen's picture is correct, demonstrating that string field theory accurately describes the nonperturbative physics of D-branes. This new nonperturbative application of string field theory has sparked a new wave of work on open string field theory, revealing many remarkable new structures. In particular, string field theory now provides a concrete framework in which disconnected string backgrounds can emerge from the equations of motion of a single underlying theory. Although so far this can only be shown explicitly in the open string context, this work paves the way for a deeper understanding of background-independence in quantum theories of gravity.

2. D-branes

In this section we briefly review some basic features of D-branes. The concepts developed here will be useful to describe tachyonic D-brane configurations in the following section. For more detailed reviews of D-branes, see the reviews of Polchinski,[40] and of Taylor.[41]

2.1. *D-branes and Ramond-Ramond charges*

D-branes can be understood from many points of view. In these lectures we primarily focus on the viewpoint motivated by the recent work on tachyon condensation, which is that that D-branes are solitons in string field theory. The original realization of the importance of D-branes in string theory stemmed from Polchinski's realization that D-branes could be described in two complementary fashions: *a*) as extended extremal black brane solutions of supergravity that carry conserved charges, and *b*) as hypersurfaces on which strings have Dirichlet boundary conditions. We now discuss these two viewpoints briefly.

a) The ten-dimensional type IIA and IIB supergravity theories each have a set of $(p+1)$-form fields $A^{(p+1)}_{\mu_1 \cdots \mu_{(p+1)}}$ in the supergraviton multiplet, with p even/odd for type IIA/IIB supergravity. These are the Ramond-Ramond (RR) fields in the massless superstring spectrum. For each of these $(p+1)$-form fields, there is a solution of the supergravity field equations that is invariant under $(p+1)$-dimensional Lorentz transformations, and which has the form of an extremal black hole solution in the $9-p$ spatial directions that are not affected by these Lorentz transformations.[42] These "black p-brane" solutions carry charge under the RR fields $A^{(p+1)}$, and are BPS states in the supergravity theory that preserve half the supersymmetry of the theory. These solutions represent the gravitational and gauge backgrounds created by the branes, in a way similar to that in which the Schwarzschild solution represents the gravitational background of a point mass, or the Coulomb field represents the electric field of a point charge.

b) In type IIA and IIB string theory, it is possible to consider open strings with Dirichlet boundary conditions on some number $9 - p$ of the spatial coordinates $x^{\mu}(\sigma)$. The locus of points defined by such Dirichlet boundary conditions defines a $(p+1)$-dimensional hypersurface Σ_{p+1} in the ten-dimensional spacetime. When p is even/odd in type IIA/IIB string theory, the spectrum of the resulting quantum open string theory contains a massless set of fields $A_{\alpha}, \alpha = 0, 1, \ldots, p$ and $X^a, a = p+1, \ldots, 9$. These fields can be associated with a gauge field living on the hypersurface Σ_{p+1}, and a set of degrees of freedom describing the transverse fluctuations of this hypersurface in spacetime, respectively. Thus, the quantum fluctuations of the open string describe a fluctuating $(p+1)$-dimensional hypersurface in spacetime — a Dirichlet-brane, or "D-brane".

The remarkable insight of Polchinski[43] in 1995 was the observation that the stable Dirichlet-branes of superstring theory carry Ramond-Ramond charges, and therefore should be described in the low-energy supergravity limit of string theory by precisely the black p-branes discussed in *a*). This connection between the string and supergravity descriptions of these nonperturbative objects paved the way to a dramatic series of new developments in string theory, including connections between string theory and supersymmetric gauge theories, string constructions of black holes, and new approaches to string phenomenology. The bosonic D-branes on which we concentrate attention in these lectures do not carry conserved charges, and thus are not associated with supergravity solutions as in *a)*; rather, these D-branes can be described through open bosonic strings with some Dirichlet boundary conditions as in *b)*.

2.2. *Born-Infeld and super Yang-Mills D-brane actions*

In this subsection we briefly review the low-energy super Yang-Mills description of the dynamics of one or more D-branes. As discussed in the previous subsection, the massless open string modes on a Dp-brane in type IIA or IIB superstring theory describe a $(p+1)$-component gauge field A_{α}, $9-p$ transverse scalar fields X^a, and a set of massless fermionic gaugino fields. The scalar fields X^a describe small fluctuations of the D-brane around a flat hypersurface. If the D-brane geometry is sufficiently far from flat, it is useful to describe the D-brane configuration by a general embedding $X^\mu(\xi)$, where ξ^α are $p+1$ coordinates on the Dp-brane world-volume $\Sigma_{(p+1)}$, and X^μ are ten functions giving a map from $\Sigma_{(p+1)}$ into the space-time manifold $\mathbf{R}^{9,1}$. Just as the Einstein equations which govern the geometry of spacetime arise from the condition that the one-loop contribution to the closed string beta function vanishes, a set of equations of motion for a general Dp-brane geometry and associated world-volume gauge field can be derived from a calculation of the one-loop open string beta function.[44] These equations of motion arise from the classical Born-Infeld action:

$$S = -T_p \int d^{p+1}\xi \; e^{-\varphi} \sqrt{-\det(G_{\alpha\beta} + B_{\alpha\beta} + 2\pi\alpha' F_{\alpha\beta})} + S_{\mathrm{CS}} + \text{fermions} \tag{1}$$

where G, B, and φ are the pullbacks of the ten-dimensional metric, antisymmetric tensor, and dilaton to the D-brane world-volume, while F is the field strength of the world-volume $U(1)$ gauge field A_α. S_{CS} represents a set of Chern-Simons terms which will be discussed in the following subsection. This action can be verified by a perturbative string calculation,[40] which also gives a precise expression for the brane tension

$$\tau_p = \frac{T_p}{g_s} = \frac{1}{g_s\sqrt{\alpha'}} \frac{1}{(2\pi\sqrt{\alpha'})^p} \tag{2}$$

where $g_s = e^{\langle\varphi\rangle}$ is the closed string coupling, equal to the exponential of the dilaton expectation value.

A particular limit of the Born-Infeld action (1) is useful to describe many low-energy aspects of D-brane dynamics. Take the background space-time $G_{\mu\nu} = \eta_{\mu\nu}$ to be flat, and all other supergravity fields $(B_{\mu\nu}, A^{(p+1)}_{\mu_1\cdots\mu_{p+1}})$ to vanish. We then assume that the D-brane is approximately flat, and is close to the hypersurface $X^a = 0, a > p$, so that we may make the static gauge choice $X^\alpha = \xi^\alpha$. We furthermore take the low-energy limit in which $\partial_\alpha X^a$ and $2\pi\alpha' F_{\alpha\beta}$ are small and of the same order. The action (1) can then be expanded as

$$S = -\tau_p V_p - \frac{1}{4g_{\mathrm{YM}}^2} \int d^{p+1}\xi \left(F_{\alpha\beta}F^{\alpha\beta} + \frac{2}{(2\pi\alpha')^2} \partial_\alpha X^a \partial^\alpha X^a \right) + \cdots \tag{3}$$

where V_p is the p-brane world-volume and the coupling g_{YM} is given by

$$g_{\mathrm{YM}}^2 = \frac{1}{4\pi^2\alpha'^2\tau_p} = \frac{g_s}{\sqrt{\alpha'}}(2\pi\sqrt{\alpha'})^{p-2}\,. \tag{4}$$

Ignoring fermionic terms, the second term in the right-hand side of (3) is simply the reduction to $(p+1)$ dimensions of the ten-dimensional $\mathcal{N}=1$ super Yang-Mills action:

$$S = \frac{1}{g_{\mathrm{YM}}^2} \int d^{10}\xi \left(-\frac{1}{4} F_{\mu\nu} F^{\mu\nu} + \frac{i}{2} \bar{\psi} \Gamma^\mu \partial_\mu \psi \right) \tag{5}$$

where for $\alpha, \beta \leq p$, $F_{\alpha\beta}$ is the world-volume $U(1)$ field strength, and for $a > p, \alpha \leq p$, $F_{\alpha a} \to \partial_\alpha X^a/(2\pi\alpha')$.

When multiple Dp-branes are present, the D-brane action is modified in a fairly simple fashion.[45] Consider a system of N coincident D-branes. For every pair of branes $\{i, j\}$ there is a set of massless fields

$$(A_\alpha)_i^{\ j}, \quad (X^a)_i^{\ j}, \tag{6}$$

associated with strings stretching from the ith brane to the jth brane; the indices i, j are known as Chan-Paton indices. Treating the fields (6) as N-by-N matrices, and letting Tr denote the trace operation for such matrices, the multiple brane analogue of the Born-Infeld action (1) takes the schematic form

$$S \sim \int \mathrm{Tr} \sqrt{-\det\left(G + B + 2\pi\alpha' F\right)}\,. \tag{7}$$

In order to properly define this nonabelian analog of the Born-Infeld action (NBI), it is necessary to resolve the ordering ambiguities in (7). Since the spacetime coordinates X^a associated with the D-brane positions in space-time become themselves matrix-valued, even evaluating the pullbacks $G_{\alpha\beta}, B_{\alpha\beta}$ involves resolving ordering issues. Much work has been done recently to resolve these ordering ambiguities[46] but there is still no known definition of the nonabelian Born-Infeld theory (7) which is valid to all orders.

The nonabelian Born-Infeld action (7) becomes much simpler when, once again, the background space-time is assumed to be flat and we take the low-energy limit , leading to the nonabelian $U(N)$ super Yang-Mills action in $p+1$ dimensions. This action is the reduction to $p+1$ dimensions of the ten-dimensional $U(N)$ super Yang-Mills action (analogous to (5)). In this reduction, for $\alpha, \beta \leq p$, $F_{\alpha\beta}$ is the world-volume $U(N)$ field strength, and for $a > p, \alpha \leq p$, $F_{\alpha a} \to \partial_\alpha X^a$, where now A_α, X^a, and $F_{\alpha\beta}$ are $N \times N$ matrices. Since the derivatives ∂_a are set to zero in the dimensional reduction, we furthermore have, for $a, b > p$, $F_{ab} \to -i[X^a, X^b]$.

The low-energy description of a system of N coincident flat D-branes is thus given by $U(N)$ super Yang-Mills theory in the appropriate dimension.

This connection between D-brane actions in string theory and super Yang-Mills theory has led to many new developments, including new insights into supersymmetric field theories, the M(atrix) theory and AdS/CFT correspondences, and brane world scenarios.

2.3. *Branes from branes*

In this subsection we describe a remarkable feature of D-brane systems: one or more D-branes of a fixed dimension can be used to construct additional D-branes of higher or lower dimension.

In our discussion of the D-brane action (1), we mentioned a group of terms S_{CS} which we did not describe explicitly. For a single Dp-brane, these Chern-Simons terms can be combined into a single expression of the form

$$S_{\text{CS}} \sim \int_{\Sigma_{p+1}} \mathcal{A}\, e^{F+B}\,, \tag{8}$$

where $\mathcal{A} = \sum_k A^{(k)}$ represents a formal sum over all the Ramond-Ramond fields $A^{(k)}$ of various dimensions. In this integral, for each term $A^{(k)}$, the nonvanishing contribution to (8) is given by expanding the exponential of $F + B$ to order $(p+1-k)/2$, where the dimension of the resulting form saturates the dimension of the brane. For example, on a Dp-brane, there is a coupling of the form

$$\int_{\Sigma_{(p+1)}} A^{(p-1)} \wedge F\,. \tag{9}$$

This coupling implies that the $U(1)$ field strength on the Dp-brane couples to the RR field associated with $(p-2)$-branes. Thus, we can associate magnetic fields on a Dp-brane with dissolved $(p-2)$-branes living on the Dp-brane. This result generalizes to a system of multiple Dp-branes, in which case a trace is included on the right-hand side of (8). For example, on N compact Dp-branes wrapped on a p-torus, the flux

$$\frac{1}{2\pi}\int \text{Tr}\, F_{\alpha\beta}, \tag{10}$$

of the magnetic field over a two-cycle on the torus is quantized and measures the number of units of D$(p-2)$-brane charge on the Dp-branes that threads the cycle integrated over. Thus, these branes are encoded in the field strength $F_{\alpha\beta}$. The object in (10) is the relevant component of the first Chern class of the $U(N)$ bundle described by the gauge field on the N

branes. Similarly,

$$\frac{1}{8\pi^2}\int \mathrm{Tr}\, F \wedge F \tag{11}$$

encodes D$(p-4)$-brane charge on the Dp-branes, etc..

Just as lower-dimensional branes can be described in terms of the degrees of freedom associated with a system of N Dp-branes through the field strength $F_{\alpha\beta}$, higher-dimensional branes can be described by a system of N Dp-branes in terms of the commutators of the matrix-valued scalar fields X^a. Just as $\frac{1}{2\pi}F$ measures $(p-2)$-brane charge, the matrix

$$2\pi i[X^a, X^b] \tag{12}$$

measures $(p+2)$-brane charge.[41,47,48] The charge (12) should be interpreted as a form of local charge density. Just as the N positions of the Dp-branes are replaced by matrices in the nonabelian theory, so the locations of the charges become matrix-valued. The trace of (12) vanishes for finite sized matrices because the net Dp-brane charge of a finite-size brane configuration in flat spacetime vanishes. Higher multipole moments of the brane charge, however, have a natural definition in terms of traces of the charge matrix multiplied by powers of the scalars X^a, and generically are nonvanishing.

A simple example of the mechanism by which a system of multiple Dp-branes form a higher-dimensional brane is given by the matrix sphere. If we take a system of D0-branes with scalar matrices X^a given by

$$X^a = \frac{2r}{N}J^a, \qquad a = 1, 2, 3 \tag{13}$$

where J^a are the generators of $SU(2)$ in the N-dimensional representation, then we have a configuration corresponding to the "matrix sphere". This is a D2-brane of spherical geometry living on the locus of points satisfying $x^2 + y^2 + z^2 = r^2$. The "local" D2-brane charge of this brane is given by (12); here, for example, the D2-brane charge in the x-y plane is proportional to the matrix X^3 (z), as one would expect from the geometry of a spherical brane. The D2-brane configuration given by (13) is rotationally invariant (up to a gauge transformation). The restriction of the brane to the desired locus of points can be seen from the relation $(X^1)^2 + (X^2)^2 + (X^3)^2 = r^2 \mathbb{1} + \mathcal{O}(N^{-2})$.

2.4. *T-duality*

We conclude our discussion of D-branes with a brief description of T-duality. T-duality is a perturbative and nonperturbative symmetry which relates the type IIA and type IIB string theories. This duality symmetry was in fact crucial in the original discovery of D-branes.[43] A more detailed discussion of T-duality can be found in the textbook by Polchinski.[49] Using T-duality, we construct an explicit example of a brane within a brane encoded in super Yang-Mills theory, illustrating the ideas of the previous subsection. This example will be used in the following section to construct an analogous configuration with a tachyon.

Consider type IIA string theory on a spacetime of the form $M^9 \times S^1$ where M^9 is a generic 9-manifold of Lorentz signature, and S^1 is a circle of radius R. T-duality is the statement that this theory is precisely equivalent, even at the perturbative level, to type IIB string theory on the spacetime $M^9 \times (S^1)'$, where $(S^1)'$ is a circle of radius $R' = \alpha'/R$.

T-duality symmetry is most easily understood in the case of closed strings, where it amounts to an exchange of winding and momentum modes of the string. The string winding modes on S^1 have energy $R|m|/\alpha'$, where m is the winding number. The T-dual momentum modes on $(S^1)'$ have energy $|n|/R'$, where n is the momentum quantum number. These two sets of values coincide when m and n run over all possible integers. It is in fact straightforward to check that the full spectrum of closed string states is unchanged under T-duality. For the case of open strings, T-duality maps an open string with Neumann boundary conditions on S^1 to an open string with Dirichlet boundary conditions on $(S^1)'$, and vice versa. Thus, a Dp-brane which is wrapped around the circle S^1 is mapped under T-duality to a D$(p-1)$-brane which is localized to a point on the circle $(S^1)'$. Under T-duality the low-energy $(p+1)$-dimensional Yang-Mills theory on the p-brane is replaced by a p-dimensional Yang-Mills theory on the dual $(p-1)$-brane. Mathematically, the covariant derivative operator in the direction S^1 is replaced under T-duality with an adjoint scalar field X^a. Formally, this adjoint scalar field is an infinite size matrix,[50] which contains information about the open strings wrapped an arbitrary number of times around the compact direction $(S^1)'$.

We can summarize the relevant mappings under T-duality in the following table

$$
\begin{array}{rcr}
\mathrm{IIA}/S^1 & \leftrightarrow & \mathrm{IIB}/(S^1)' \\
R & \leftrightarrow & R' = \alpha'/R \\
\text{Dirichlet/Neumann b.c.'s} & \leftrightarrow & \text{Neumann/Dirichlet b.c.'s} \\
p\text{-brane} & \leftrightarrow & (p \pm 1)\text{-brane} \\
2\pi\alpha'(i\partial_a + A_a) & \leftrightarrow & X^a
\end{array}
$$

The phenomena by which field strengths in one brane describe lower- or higher-dimensional branes can be easily understood using T-duality. The following simple example may help to clarify this connection. (For a more detailed discussion using this point of view, see Taylor.[41]) constant magnetic

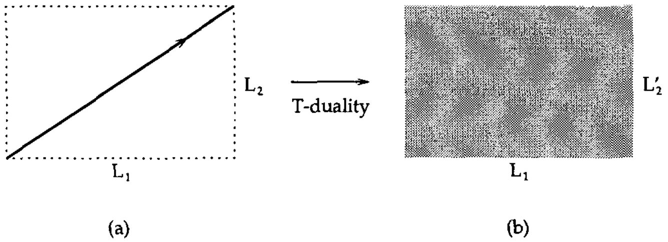


Figure 1. T-duality takes a diagonal D1-brane on a two-torus (a) to a D2-brane on the dual torus with constant magnetic flux encoding an embedded D0-brane (b).

Consider a D1-brane wrapped diagonally on a two-torus T^2 with sides of length $L_1 = L$ and $L_2 = 2\pi R$. (Figure 1(a)). This configuration is described in terms of the world-volume Yang-Mills theory on a D1-brane stretched in the L_1 direction through a transverse scalar field

$$X^2 = 2\pi R\xi_1/L\,. \tag{14}$$

To be technically precise, this scalar field should be treated as an $\infty \times \infty$ matrix[50] whose (n, m) entry is associated with strings that connect the nth and mth images of the D1-brane on the covering space of S^1. The diagonal elements $X^2_{n,n}$ of this infinite matrix are given by $2\pi R(\xi_1 + nL)/L$, while all off-diagonal elements vanish. While the resulting matrix-valued function of ξ_1 is not periodic, it is periodic up to a gauge transformation

$$X^2(L) = VX^2(0)V^{-1} \tag{15}$$

where V is the shift matrix with nonzero elements $V_{n,n+1} = 1$.

Under T-duality in the x^2 direction the infinite matrix X^2_{nm} becomes the Fourier mode representation of a gauge field on a dual D2-brane:

$$A_2 = \frac{1}{R'L}\xi_1 \,. \tag{16}$$

The magnetic flux associated with this gauge field is

$$F_{12} = \frac{1}{R'L}\,, \tag{17}$$

so that

$$\frac{1}{2\pi}\int F_{12}\, d\xi^1\, d\xi^2 = 1\,. \tag{18}$$

Note that the boundary condition (15) on the infinite matrix X^2 transforms under T-duality to the boundary condition on the gauge field

$$\begin{aligned} A_2(L, x_2) &= e^{2\pi i \xi_2/L'_2}\left(A_2(0, x_2) + i\partial_2\right) e^{-2\pi i \xi_2/L'_2} \\ &= A_2(0, x_2) + \frac{2\pi}{L'_2}, \end{aligned} \tag{19}$$

which (16) clearly satisfies. The off-diagonal elements of the shift matrix V in (15) describe winding modes which correspond after T-duality to the first Fourier mode $e^{2\pi i \xi_2/L'_2}$. The boundary condition on the gauge fields in the ξ_2 direction is trivial, which simplifies the T-duality map; a similar construction can be done with a nontrivial boundary condition in both directions, although the configuration looks more complicated in the D1-brane picture.

This construction gives a simple Yang-Mills description of the mapping of D-brane charges under T-duality: the initial configuration described above has charges associated with a single D1-brane wrapped around each of the directions of the 2-torus: $\mathrm{D1}_1$+ $\mathrm{D1}_2$. Under T-duality, these D1-branes are mapped to a D2-brane and a D0-brane respectively: $\mathrm{D2}_{12}$+ D0. The flux integral (18) is the representation in the D2-brane world-volume Yang-Mills theory of the charge associated with a D0-brane which has been uniformly distributed over the surface of the D2-brane, just as in (10).

3. Tachyons and D-branes

We now turn to the subject of tachyons. Certain D-brane configurations are unstable, both in supersymmetric and nonsupersymmetric string theories. This instability is manifested as a tachyon, that is, as a state with $M^2 < 0$ in the spectrum of open strings that end on the D-brane. We will explicitly describe the tachyonic mode in the case of the open bosonic string in Section 4.1; this open bosonic string tachyon will be the focal point of most of the developments described in these notes. In this section we list some elementary D-brane configurations where tachyons arise, and we describe a particular situation in which the tachyon can be seen in the low-energy Yang-Mills description of the D-branes. This Yang-Mills background with a tachyon provides a simple field-theory model of a system analogous to the more complicated string field theory tachyon we describe in the later part of these notes. This simpler model may be useful to keep in mind in the later analysis.

3.1. *D-brane configurations with tachyonic instabilities*

Some simple examples of unstable D-brane configurations where the open string contains a tachyon include the following:

Brane-antibrane: A pair of parallel Dp-branes with opposite orientation in type IIA or IIB string theory which are separated by a distance $d \ll l_s$ give rise to a tachyon in the spectrum of open strings stretched between the branes.[51] The difference in orientation of the branes means that the two branes are really a brane and antibrane, carrying equal but opposite RR charges. Since the net RR charge is 0, the brane and antibrane can annihilate, leaving an uncharged vacuum configuration.

Wrong-dimension branes: In type IIA/IIB string theory, a Dp-brane of even/odd spatial dimension p is a stable BPS state with nonzero RR charge. On the other hand, a Dp-brane of the *wrong* dimension (*i.e.*, odd/even for IIA/IIB) carries no charges under the classical IIA/IIB supergravity fields, and has a tachyon in the open string spectrum. Such a brane can decay into the vacuum without violating charge conservation.

Bosonic D-branes: Like the wrong-dimension branes of IIA/IIB string theory, a Dp-brane of any dimension in the bosonic string theory carries no conserved charge and has a tachyon in the open string spectrum. Again, such a brane can decay into the vacuum without violating charge conservation.

3.2. *Example: tachyon in low-energy field theory of two D-branes*

In order to illustrate the physical behavior of tachyonic configurations, we consider in this subsection a simple example[52,53] where a brane-antibrane tachyon can be seen in the context of the low-energy Yang-Mills theory.

The system we want to consider is a simple generalization of the (D2 + D0)-brane configuration we described using Yang-Mills theory in section 2.4. Consider a pair of D2-branes wrapped on a two-torus, one of which has a D0-brane embedded in it as a constant positive magnetic flux, and the other of which has an anti-D0-brane within it described by a constant negative magnetic flux. We take the two dimensions of the torus to be L_1, L_2. Following the discussion of Section 2.4, this configuration is equivalent under T-duality in the L_2 direction to a pair of crossed D1-branes (see Figure 2).

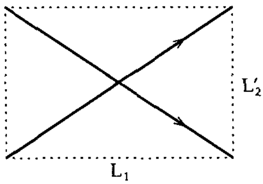


Figure 2. A pair of crossed D1-branes, T-dual to a pair of D2-branes with uniformly embedded D0- and anti-D0-branes.

The Born-Infeld energy of this configuration is

$$\begin{aligned} E_{\text{BI}} &= 2\sqrt{(\tau_2 L_1 L_2)^2 + \tau_0^2} \\ &= \frac{1}{g}\left[\frac{2L_1L_2}{\sqrt{2\pi}} + \frac{(2\pi)^{3/2}}{L_1L_2} + \cdots\right], \end{aligned} \tag{20}$$

in units where $2\pi\alpha' = 1$. This can be computed either directly from the Born-Infeld action on the D2-branes (the abelian theory can be used since the matrices are diagonal), or by simply using the Pythagorean theorem in the T-dual D1-brane picture. The second term in the last line corresponds to the Yang-Mills approximation. In this approximation (dropping the D2-brane energy) the energy is

$$E_{\text{YM}} = \frac{\tau_2}{4}\int \text{Tr}\ F_{\alpha\beta}F^{\alpha\beta} = \frac{1}{4\sqrt{2\pi}g}\int \text{Tr}\ F_{\alpha\beta}F^{\alpha\beta}. \tag{21}$$

We are interested in studying this configuration in the Yang-Mills approximation, in which we have a $U(2)$ theory on T^2 with field strength

$$F_{12} = \begin{pmatrix} \frac{2\pi}{L_1 L_2} & 0 \\ 0 & -\frac{2\pi}{L_1 L_2} \end{pmatrix} = \frac{2\pi}{L_1 L_2}\tau_3 \,. \tag{22}$$

This field strength can be realized as the curvature of a linear gauge field

$$A_1 = 0, \qquad A_2 = \frac{2\pi}{L_1 L_2}\xi\tau_3 \,, \tag{23}$$

which satisfies the boundary conditions

$$A_j(L, \xi_2) = \Omega(i\partial_j + A_j(0, \xi_2))\Omega^{-1} \,, \tag{24}$$

where

$$\Omega = e^{2\pi i(\xi_1/L_2)\tau_3} \,. \tag{25}$$

It is easy to check that this configuration indeed satisfies

$$E_{\rm YM} = \frac{1}{2g}\frac{(2\pi)^{3/2}}{L_1 L_2}{\rm Tr}\;\tau_3^2 = \frac{1}{g}\frac{(2\pi)^{3/2}}{L_1 L_2} \,, \tag{26}$$

as desired from (20). Since

$${\rm Tr}\; F_{\alpha\beta} = 0, \tag{27}$$

the gauge field we are considering is in the same topological equivalence class as $F = 0$. This corresponds to the fact that the D0-brane and anti-D0-brane can annihilate. To understand the appearance of the tachyon, we can consider the spectrum of excitations δA_α around the background (23).[52] The eigenvectors of the quadratic mass terms in this background are described by torus theta functions which satisfy boundary conditions related to (24). There are precisely two elements in the spectrum with the negative eigenvalue $-4\pi/L_1 L_2$. These theta functions[52] are tachyonic modes of the theory which are associated with the annihilation of the positive and negative fluxes that encode the D0- and anti-D0-brane. These tachyonic modes are perhaps easiest to understand in the dual configuration, where they provide a direction of instability in which the two crossed D1-branes reconnect as in Figure 3.

It is also interesting to note that in the T-dual picture the tachyonic modes of the gauge field have support which is localized near the two intersection points and take the off-diagonal form

$$\delta A_t \sim \begin{pmatrix} 0 & \star \\ \star & 0 \end{pmatrix} \,, \tag{28}$$

Figure 3. The brane-antibrane instability of a D0-$\bar{\text{D0}}$ system embedded in two D2-branes, as seen in the T-dual D1-brane picture.

which naturally encodes our geometric understanding that the tachyonic modes reconnect the two D1-branes near each intersection point.

The full Yang-Mills action around the background (23) can be written as a quartic function of the mass eigenstates around this background. Written in terms of these modes, there are nontrivial cubic and quartic terms which couple the tachyonic modes to all the massive modes in the system. If we integrate out the massive modes, we know from the topological reasoning above that an effective potential arises for the tachyonic mode A_t, with a maximum value of (26) and a minimum value of 0. This system is highly analogous to the bosonic open string tachyon we will discuss in the remainder of these lectures. Our current understanding of the bosonic string through bosonic string field theory is analogous to that of someone who only knows the Yang-Mills theory around the background (23) in terms of a complicated quartic action for an infinite family of modes. Without knowledge of the topological structure of the theory, and given only a list of the coefficients in the quartic action, such an individual would have to systematically calculate the tachyon effective potential by explicitly integrating out all the massive modes one by one. This would give a numerical approximation to the minimum of the effective potential, which could be made arbitrarily good by raising the mass of the cutoff at which the effective action is computed. It may be helpful to keep this example in mind in the following sections, where an analogous tachyonic system is considered in string field theory. For further discussion of this unstable configuration in Yang-Mills theory, see Refs. 52, 53, 54, 55.

3.3. *The Sen conjectures*

The existence of the tachyonic mode in the open bosonic string indicates that the standard choice of perturbative vacuum for this theory is unstable. In the early days of the subject, there was some suggestion that this tachyon could condense, leading to a more stable vacuum.[56] Kostelecky and Samuel argued early on that the stable vacuum could be identified in string field theory in a systematic way,[57] however there was no clear physical picture for the significance of this stable vacuum. In 1999, Ashoke Sen reconsidered the problem of tachyons in string field theory. Sen suggested that the open bosonic string should really be thought of as living on a D25-brane, and hence that the perturbative vacuum for this string theory should have a nonzero vacuum energy associated with the tension of this D25-brane. He suggested that the tachyon is simply the instability mode of the D25-brane, which carries no conserved charge and hence is not expected to be stable, as discussed in section 3.1. More precisely, Sen conjectured that the following three statements are true:[39]

(1) The tachyon potential has a locally stable minimum, whose energy density $\mathcal{E}$, measured with respect to that of the unstable critical point, is equal to minus the tension of the D25-brane:

$$\mathcal{E} = -T_{25} \,. \tag{29}$$

(2) Lower-dimensional D-branes are solitonic solutions of the string theory on the background of a D25-brane.

(3) The locally stable vacuum of the system is the closed string vacuum. In this vacuum the D25-brane is absent and no conventional open string excitations exist.

It was also suggested by Sen that open string field theory was a natural setup to test the above conjectures. He sharpened the first conjecture by suggesting that Witten's OSFT should precisely reproduce the tension of the D25-brane, which he expressed in terms of the open string coupling constant g which appears in the formulation of open string field theory:

$$T_{25} = \frac{1}{2\pi^2 g^2} \,. \tag{30}$$

We will give the instructive derivation of this result in section 7.

Our first encounter with the tachyon conjectures will happen in section 5, where we calculate the first nontrivial term in the tachyon potential, find a minimum, and discover that, even with this rough approximation, the calculated $\mathcal{E}$ gives about 70% of the expected answer. In Section 7 of these lectures we systematically explore the evidence for these conjectures in Witten's OSFT. First, however, we need to develop the technical tools to do specific calculations in string field theory.

4. Witten's cubic open string field theory

In this and in the following two sections we give a detailed description of Witten's open string field theory.[26] This section contains a general introduction to this string field theory. Subsection 4.1 reviews the quantization of the open bosonic string in 26 dimensions and sets notation. Subsection 4.2 gives a heuristic introduction to open string field theory, which follows Witten's original paper. In subsection 4.3 we discuss the algebraic structure of open string field theory which emerges naturally in the context of conformal field theory. This discussion also develops the properties of the twist operator Ω which reverses the orientation of open strings.

The work in the present section prepares the ground for sections 5 and 6, in which precise definitions of the open bosonic SFT are given using conformal field theory and the mode decomposition of overlap equations. For further background on Witten's OSFT see the reviews of LeClair, Peskin and Preitschopf,[58] of Thorn,[59] and of Gaberdiel and Zwiebach.[60]

4.1. *The bosonic open string*

In this subsection we review the quantization of the open bosonic string. For further details see the textbooks by Green, Schwarz, and Witten[61] and by Polchinski.[49] The bosonic open string can be quantized using the BRST approach starting from the action

$$S = -\frac{1}{4\pi\alpha'}\int \sqrt{-\gamma}\gamma^{ab}\partial_a X^\mu \partial_b X_\mu , \tag{31}$$

where γ is an auxiliary dynamical metric on the world-sheet. This action can be gauge-fixed to conformal gauge $\gamma_{ab} \sim \delta_{ab}$, introducing at the same time ghost and antighost fields $c^\pm(\sigma), b_{\pm\pm}(\sigma)$. The gauge-fixed action is

$$S = -\frac{1}{4\pi\alpha'}\int \partial_a X^\mu \partial^a X_\mu + \frac{1}{\pi}\int \left(b_{++}\partial_- c^+ + b_{--}\partial_+ c^-\right) . \tag{32}$$

The matter fields X^μ can be expanded in modes using

$$X^\mu(\sigma,\tau) = x_0^\mu + 2p^\mu \tau + i\sqrt{2}\sum_{n\neq 0}\frac{1}{n}\alpha_n^\mu \cos(n\sigma)e^{-in\tau} , \tag{33}$$

where we have fixed $l_s = \sqrt{2\alpha'} = \sqrt{2}$, so that $\alpha' = 1$. In the quantum theory, x_0^μ and p^μ obey the canonical commutation relations

$$[x_0^\mu, p^\nu] = i\eta^{\mu\nu} . \tag{34}$$

The α_n^μ's with negative/positive values of n become raising/lowering operators for the oscillator modes on the string. They satisfy the Hermiticity conditions $(\alpha_n^\mu)^\dagger = \alpha_{-n}^\mu$ and the commutation relations

$$[\alpha_m^\mu, \alpha_n^\nu] = m\eta^{\mu\nu}\delta_{m+n,0} . \tag{35}$$

We will often use the canonically normalized oscillators:

$$a_n^\mu = \frac{1}{\sqrt{n}}\alpha_n^\mu , \quad n \geq 1 , \tag{36}$$

which obey the commutation relations

$$[a_m^\mu, a_n^{\nu\dagger}] = \eta^{\mu\nu}\delta_{m,n} . \tag{37}$$

We will also frequently use position modes x_n for $n \neq 0$ and lowering and raising operators $a_0, a_0^\dagger$ for the zero modes. These are related to the modes in (33) through (dropping space-time indices)

$$\begin{aligned} x_n &= \frac{i}{\sqrt{2n}}(a_n - a_n^\dagger) \\ x_0 &= \frac{i}{2}(a_0 - a_0^\dagger) \end{aligned} \tag{38}$$

The ghost and antighost fields can be decomposed into modes through

$$\begin{aligned} c^\pm(\sigma) &= \sum_n c_n e^{\mp in\sigma} \\ b_{\pm\pm}(\sigma) &= \sum_n b_n e^{\mp in\sigma} . \end{aligned} \tag{39}$$

The ghost and antighost modes satisfy the anticommutation relations

$$\begin{aligned} \{c_n, b_m\} &= \delta_{n+m,0} \\ \{c_n, c_m\} &= \{b_n, b_m\} = 0 . \end{aligned} \tag{40}$$

A general state in the open string Fock space can be written in the form

$$\alpha_{-n_1}^{\mu_1} \cdots \alpha_{-n_i}^{\mu_i} \; c_{-m_1} \cdots c_{-m_j} \; b_{-p_1} \cdots b_{-p_l} \; |0; k\rangle \tag{41}$$

where

$$n_i \geq 1 , \quad m_i \geq -1 , \quad \text{and} \quad p_i \geq 2 , \tag{42}$$

since $|0;k\rangle$ is annihilated by

$$\begin{aligned} b_n|0;k\rangle &= 0, \quad n \geq -1 \\ c_n|0;k\rangle &= 0, \quad n \geq 2\,, \\ \alpha_n^\mu|0;k\rangle &= 0, \quad n \geq 1\,. \end{aligned} \tag{43}$$

The state $|0;k\rangle$ is a momentum eigenstate:

$$p^\mu|0;k\rangle = k^\mu|0;k\rangle\,. \tag{44}$$

The zero-momentum state $|0;0\rangle$ is the SL(2,R) invariant vacuum; we will often write it simply as $|0\rangle$. This vacuum is defined to have ghost number 0, and it is normalized by the equation

$$\langle 0;k|c_{-1}c_0c_1|0;k'\rangle = (2\pi)^{26}\delta(k-k') \tag{45}$$

For string field theory we will also find it convenient to work with the vacua of ghost number one and two:

$$\begin{aligned} G=1: &\quad |0_1\rangle = c_1|0\rangle \\ G=2: &\quad |0_2\rangle = c_0c_1|0\rangle\,. \end{aligned} \tag{46}$$

In the notation of Polchinski,[49] these two vacua are written as

$$\begin{aligned} |0_1\rangle &= |0\rangle_m \otimes |\downarrow\rangle \\ |0_2\rangle &= |0\rangle_m \otimes |\uparrow\rangle\,, \end{aligned} \tag{47}$$

where $|0\rangle_m$ is the matter vacuum and $|\downarrow\rangle, |\uparrow\rangle$ are the ghost vacua annihilated by b_0, c_0.

The BRST operator of this theory is given by

$$Q_B = \sum_{n=-\infty}^{\infty} c_n L_{-n}^{(\mathrm{m})} + \sum_{n,m=-\infty}^{\infty} \frac{(m-n)}{2} : c_m c_n b_{-m-n} : - c_0 \tag{48}$$

where the matter Virasoro operators are given by

$$L_q^{(\mathrm{m})} = \begin{cases} \frac{1}{2}\sum_n \alpha_{q-n}^\mu \alpha_{\mu\, n}, & q \neq 0 \\ p^2 + \sum_{n=1}^\infty \alpha_{-n}^\mu \alpha_{\mu\, n}\,. \end{cases} \tag{49}$$

4.2. *Witten's cubic bosonic SFT*

The discussion of the previous subsection leads to a systematic quantization of the open bosonic string in the conformal field theory framework. Using this approach it is possible, in principle, to calculate an arbitrary perturbative on-shell scattering amplitude for physical string states. To study tachyon condensation in string theory, however, we require a nonperturbative, off-shell formalism for the theory— a string field theory.

A very simple form for the off-shell open bosonic string field theory action was proposed by Witten in 1986:[26]

$$S = -\frac{1}{2}\int \Psi \star Q\Psi - \frac{g}{3}\int \Psi \star \Psi \star \Psi . \tag{50}$$

This action has the general form of a Chern-Simons theory on a 3-manifold, although for string field theory there is no explicit interpretation of the integration in terms of a concrete 3-manifold. In Eq. (50), g is interpreted as the (open) string coupling constant. The field Ψ is a string field, which takes values in a graded algebra $\mathcal{A}$. Associated with the algebra $\mathcal{A}$ there is a star product

$$\star : \mathcal{A} \otimes \mathcal{A} \rightarrow \mathcal{A}, \tag{51}$$

under which the degree G is additive ($G_{\Psi\star\Phi} = G_\Psi + G_\Phi$). There is also a BRST operator

$$Q : \mathcal{A} \rightarrow \mathcal{A}, \tag{52}$$

of degree one ($G_{Q\Psi} = 1 + G_\Psi$). String fields can be integrated using

$$\int : \mathcal{A} \rightarrow \mathbf{C}. \tag{53}$$

This integral vanishes for all Ψ with degree $G_\Psi \neq 3$. Thus, the action (50) is only nonvanishing for a string field Ψ of degree 1.

The elements $Q, \star, \int$ that define the string field theory are assumed to satisfy the following axioms:

(a) Nilpotency of Q: $Q^2\Psi = 0, \quad \forall \Psi \in \mathcal{A}$.

(b) $\int Q\Psi = 0, \quad \forall \Psi \in \mathcal{A}$.

(c) Derivation property of Q:
$Q(\Psi \star \Phi) = (Q\Psi) \star \Phi + (-1)^{G_\Psi}\Psi \star (Q\Phi), \quad \forall \Psi, \Phi \in \mathcal{A}$.

(d) Cyclicity: $\int \Psi \star \Phi = (-1)^{G_\Psi G_\Phi} \int \Phi \star \Psi, \quad \forall \Psi, \Phi \in \mathcal{A}$.

(e) Associativity: $(\Phi \star \Psi) \star \Xi = \Phi \star (\Psi \star \Xi), \quad \forall \Phi, \Psi, \Xi \in \mathcal{A}$.

When these axioms are satisfied, the action (50) is invariant under the gauge transformations

$$\delta\Psi = Q\Lambda + \Psi \star \Lambda - \Lambda \star \Psi , \tag{54}$$

for any gauge parameter $\Lambda \in \mathcal{A}$ with degree 0.

When the string coupling g is taken to vanish, the equation of motion for the theory defined by (50) simply becomes $Q\Psi = 0$, and the gauge transformations (54) simply become

$$\delta\Psi = Q\Lambda\,. \tag{55}$$

Thus, when $g = 0$ this string field theory gives precisely the structure needed to describe the free bosonic string. The motivation for introducing the extra structure in (50) was to find a simple interacting extension of the free theory, consistent with the perturbative expansion of open bosonic string theory.

Witten presented this formal structure and argued that all the needed axioms are satisfied when $\mathcal{A}$ is taken to be the space of string fields

$$\mathcal{A} = \{\Psi[x(\sigma); c(\sigma), b(\sigma)]\} \tag{56}$$

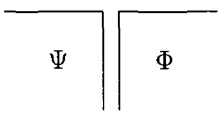
which can be described as functionals of the matter, ghost and antighost fields describing an open string in 26 dimensions with $0 \leq \sigma \leq \pi$. Such a string field can be written as a formal sum over open string Fock space states with coefficients given by an infinite family of space-time fields

$$\Psi = \int d^{26}p\ \left[\phi(p)\ |0_1;p\rangle + A_\mu(p)\ \alpha^\mu_{-1}|0_1;p\rangle + \cdots\right] \tag{57}$$

Each Fock space state is associated with a given string functional, just as the states of a harmonic oscillator are associated with wavefunctions of a particle in one dimension. For example, the matter ground state $|0\rangle_m$ annihilated by a_n for all $n \geq 1$ is associated (up to a constant C) with the functional of matter modes

$$|0\rangle_m \to C \exp\left(-\frac{1}{2}\sum_{n>0}^{\infty} n x_n^2\right)\,. \tag{58}$$

For Witten's cubic string field theory, the BRST operator Q in (50) is the usual open string BRST operator Q_B, given in (48), and the degree associated with a Fock space state is the ghost number of that state. The star product $\star$ acts on a pair of functionals Ψ, Φ by gluing the right half of one string to the left half of the other using a delta function interaction



This star product factorizes into separate matter and ghost parts. In the matter sector, the star product is given by the formal functional integral

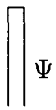
$$
\begin{aligned}
&(\Psi \star \Phi)\,[z(\sigma)] \qquad\qquad (59)\\
&\equiv \int \prod_{0\leq\tilde{\tau}\leq\frac{\pi}{2}} dy(\tilde{\tau})\; dx(\pi-\tilde{\tau}) \prod_{\frac{\pi}{2}\leq\tau\leq\pi} \delta[x(\tau)-y(\pi-\tau)]\; \Psi[x(\tau)]\Phi[y(\tau)]\,,
\end{aligned}
$$

$$
\begin{aligned}
x(\tau) &= z(\tau) \quad \text{for} \quad 0 \leq \tau \leq \frac{\pi}{2}\,,\\
y(\tau) &= z(\tau) \quad \text{for} \quad \frac{\pi}{2} \leq \tau \leq \pi\,.
\end{aligned}
$$

Similarly, the integral over a string field factorizes into matter and ghost parts, and in the matter sector is given by

$$\int \Psi = \int \prod_{0\leq\sigma\leq\pi} dx(\sigma) \prod_{0\leq\tau\leq\frac{\pi}{2}} \delta[x(\tau)-x(\pi-\tau)]\; \Psi[x(\tau)]\,. \qquad (60)$$

This corresponds to gluing the left and right halves of the string together with a delta function interaction



The ghost sector of the theory is defined in a similar fashion, but has an anomaly due to the curvature of the Riemann surface that describes the three-string vertex. The ghost sector can be described either in terms of fermionic ghost fields $c(\sigma), b(\sigma)$ or through bosonization in terms of a single bosonic scalar field $\phi_g(\sigma)$. From the functional point of view of Eqs. (59, 60), it is easiest to describe the ghost sector in the bosonized language. In this language, the ghost fields $b(\sigma)$ and $c(\sigma)$ are replaced by the scalar field $\phi_g(\sigma)$, and the star product in the ghost sector is given by (59) with an extra insertion of $\exp(3i\phi_g(\pi/2)/2)$ inside the integral. Similarly, the integration of a string field in the ghost sector is given by (60) with an insertion of $\exp(-3i\phi_g(\pi/2)/2)$ inside the integral. Witten first described the cubic string field theory using bosonized ghosts. While this approach is useful for some purposes, we will use fermionic ghost fields in the remainder of these lecture notes. With the fermionic ghosts, there is no need for insertions at the string midpoint.

The expressions (59, 60) may seem rather formal, as they are written in terms of functional integrals. These expressions, however, can be given precise meaning when described in terms of creation and annihilation operators acting on the string Fock space. In Sections 5 and 6 we give a more precise definition of the string field theory action using conformal field theory and the countable mode decomposition of the string.

4.3. *Algebraic structure of OSFT*

Here we discuss an approach to the algebraic structure of OSFT that is inspired by conformal field theory. This approach can be used to write rather general string actions, including those whose interactions are not based on delta-function overlaps. In this language the string action takes the form

$$S(\Phi) = -\frac{1}{g^2}\left[\frac{1}{2}\langle \Phi\,, Q\,\Phi\rangle + \frac{1}{3}\langle \Phi\,, \Phi * \Phi\rangle\right]. \tag{61}$$

Here g is the open string coupling constant, the string field Φ is a state in the total matter plus ghost CFT, Q is the kinetic operator, $*$ denotes a multiplication or star-product, and $\langle\cdot\,,\cdot\rangle$ is a bilinear inner product on the state space of the CFT. We will discuss the relationship between the action (61) and the form of the action (50) used in the previous subsection shortly.

The kinetic operator Q satisfies the following identities

$$\begin{aligned} &Q^2 A = 0, \\ &Q(A * B) = (QA) * B + (-1)^A A * (QB)\,, \\ &\langle QA, B\rangle = -(-)^A\langle A, QB\rangle\,. \end{aligned} \tag{62}$$

The first equation is the nilpotency condition, the second states that Q is a derivation of the star product, and the third states that Q is self-adjoint. There are also identities involving the inner product and the star operation

$$\begin{aligned} &\langle A, B\rangle = (-)^{AB}\langle B, A\rangle\,, \\ &\langle A\,, B * C\rangle = \langle A * B\,, C\rangle \\ &A * (B * C) = (A * B) * C\,. \end{aligned} \tag{63}$$

In the sign factors, the exponents $A, B, \cdots$ denote the Grassmanality of the state, and should be read as $(-)^A \equiv (-)^{\epsilon(A)}$ where $\epsilon(A) = 0\,(\text{mod } 2)$ when A is Grassmann even, and $\epsilon(A) = 1\,(\text{mod } 2)$ when A Grassmann odd. The first property above is a symmetry condition, the second indicates that the inner product has a cyclicity property analogous to the similar property of the trace operation. Finally, the last equation is the statement that the star product is associative.

Finally, we also have that the star operation is an *even* product of degree zero (as before, we identify degree with ghost number). In plain english, this means that both the grassmanality and the ghost number of the star product of two string fields is obtained from those of the string fields without any additional offset:

$$\begin{aligned}\epsilon(A * B) &= \epsilon(A) + \epsilon(B)\,,\\ \mathrm{gh}(A * B) &= \mathrm{gh}(A) + \mathrm{gh}(B)\,.\end{aligned} \tag{64}$$

In this language Q is an odd operator of degree one:

$$\begin{aligned}\epsilon(QA) &= \epsilon(A) + 1\,,\\ \mathrm{gh}(QA) &= \mathrm{gh}(A) + 1\,.\end{aligned} \tag{65}$$

In the conventions we shall work the SL(2,R) vacuum $|0\rangle$ is assigned ghost number zero. The Grassmanality $\epsilon(A)$ of a string field A is an integer mod 2. In open string field theory, Grassmanality and ghost number (degree) are related because Grassmann odd operators carry odd units of ghost number.

The algebraic structure discussed here is very similar, but not identical to that in section 4.2. The string field Φ and the action (61) can be related to the string field Ψ and the action (50) of the previous section by taking

$$\Phi = g\Psi \tag{66}$$

and by relating the inner product used here to the integral used in (50) through

$$\langle A\,, B\rangle = \int A \star B\,. \tag{67}$$

The first two conditions in (62) are then clearly equivalent to properties (a) and (c) of section 4.2. You can also readily see that the first two properties in (63) hold given properties (d) and (e). Property (b), however, does not have a counterpart in this formalism. A counterpart exists if we assume the existence of a suitable identity string field $\mathcal{I}$, as we will discuss at the end of this subsection.

Throughout these lectures we will go back and forth between the CFT notation with string field Φ and action (61) and the oscillator description with string field Ψ and action (50) (which we rewrite more explicitly as (148) in section 6). While we could have chosen to use one notation and neglect the other, both formalisms are used extensively in the literature, and some results are more easily expressed in one notation than the other. When in doubt, the reader should return to the previous paragraph to see how the two notations are related.

Let us now deduce some basic properties of the string field, in particular its ghost number and its Grassmanality. The Grassmanality of Φ can be deduced from the condition that the kinetic term of the string action must be non-vanishing. Using the above properties we have

$$\langle \Phi, Q\Phi \rangle = (-1)^{\Phi(1+\Phi)} \langle Q\Phi, \Phi \rangle = \langle Q\Phi, \Phi \rangle = -(-1)^{\Phi} \langle \Phi, Q\Phi \rangle . \tag{68}$$

It is clear that the string field Φ must be Grassmann odd. At this point we must use some CFT knowledge to decide on the Grassmanality of the SL(2,R) vacuum and on the ghost number of the string field. For bosonic strings we have that zero momentum tachyon states are of the form $tc_1|0\rangle$, where c_1 is a ghost field oscillator. Since this oscillator is Grassmann odd, and the string field is also Grassmann odd, we must declare the SL(2,R) vacuum to be Grassmann even. Thus

$$|0\rangle \text{ is a Grassmann even state of ghost number zero.} \tag{69}$$

Since the c_1 oscillator carries ghost number one, we also deduce that the open string field must have ghost number one.

$$|\Phi\rangle \text{ is a Grassmann odd state of ghost number one.} \tag{70}$$

Equations (62), (63), (64), (65), and (70) guarantee that the string field action is invariant under the gauge transformations:

$$\delta\Phi = Q\Lambda + \Phi * \Lambda - \Lambda * \Phi \,, \tag{71}$$

for any Grassmann-even ghost-number zero state Λ. Moreover, variation of the action gives the field equation

$$Q\Phi + \Phi * \Phi = 0 \,. \tag{72}$$

Exercise Verify that the string action in (61) is gauge invariant under the transformations (71).

It is convenient to use the above structures to define a multilinear object that given three string fields yields a number:

$$\langle A \,, B \,, C \rangle \equiv \langle A \,, B * C \,\rangle \tag{73}$$

The middle equation in (63) implies the *cyclicity* of the multilinear form. A small calculation immediately gives:

$$\langle A \,, B \,, C \rangle = (-)^{A(B+C)} \langle B \,, C \,, A \rangle \tag{74}$$

A basic consistency check of the signs above is that the cubic term $\langle \Phi , \Phi , \Phi \rangle$ in the action (62) is strictly cyclic for odd Φ, and therefore does not vanish.

Open string theory has additional algebraic structure that sometimes plays a crucial role. One such structure arises from the twist operation, which reverses the parametrization of a string. From the algebraic viewpoint this is summarized by the existence of an operator Ω that satisfies the following properties:

$$\begin{aligned} &\Omega(QA) = Q(\Omega A) \\ &\langle \Omega A , \Omega B \rangle = \langle A , B \rangle \\ &\Omega\,(A * B) = (-)^{AB+1}\,\Omega(B) * \Omega(A)\,. \end{aligned} \tag{75}$$

The first property means that the BRST operator has zero twist, or does not change the twist property of the states it acts on. The second property states that the bilinear form is twist invariant. The third property is crucial. Up to signs, twisting the star product of string fields amounts to multiplying the twisted states in *opposite order.* This change of order is a simple consequence of the basic multiplication rule where the second half of the first string must be glued to the first half of the second one. The sign factor is also important. For the string field Φ, which is grassmann odd, it gives

$$\Omega(\Phi * \Phi) = +\,(\Omega\Phi) * (\Omega\Phi) \tag{76}$$

with the plus sign. This result, together with the first two equations in (75) immediately implies that the string field action in (62) is twist invariant:

$$S(\Omega\Phi) = S(\Phi)\,. \tag{77}$$

This invariance under twist transformations allows one to construct new string theories by truncating the spectrum to the subset of states that are twist even. Moreover, in solving the string field equations it will be possible to find consistent solutions by restricting oneself to the twist even subspace of the string field.

Exercise. Letting Ω_A denote the Ω eigenvalue of A, show that

$$\langle A, B, C\rangle = \Omega_A\Omega_B\Omega_C(-1)^{AB+BC+CA+1}\langle C, B, A\rangle\,. \tag{78}$$

Exercise. Let $\Omega A_\pm = \pm A$ and $\epsilon(A_\pm) = 1$. Show that

$$\langle A_+, A_+, A_-\rangle = 0\,. \tag{79}$$

Exercise. We will see later that the star product of the vacuum with itself is the vacuum plus Virasoro descendents:

$$|0\rangle * |0\rangle = |0\rangle + \cdots \tag{80}$$

Show that this implies that the vacuum is twist odd:

$$\Omega|0\rangle = -|0\rangle. \tag{81}$$

The star algebra may have an identity element $\mathcal{I}$. If $\mathcal{I}$ exists, it is presumed to satisfy

$$\mathcal{I} * A = A * \mathcal{I} = A\,, \tag{82}$$

for all states A. Some properties of $\mathcal{I}$ are immediately deduced from the above definition:

$$\mathcal{I} \text{ is Grassmann even, ghost number zero, twist odd string field.} \tag{83}$$

The twist odd property follows from the twist property of products

$$\Omega(\mathcal{I} * A) = (-1)^{0\cdot A+1}(\Omega A) * (\Omega\mathcal{I}) = -(\Omega A) * (\Omega\mathcal{I})\,. \tag{84}$$

Since the left hand side is also just (ΩA) it must follow that

$$\Omega\mathcal{I} = -\mathcal{I}\,. \tag{85}$$

This is consistent with the fact that the SL(2,R) vacuum is also twist odd. Indeed the identity string field is just the vacuum plus Virasoro descendents of the vacuum, as we shall see in Section 8.4.

Any derivation D of the star algebra should annihilate the identity:

$$D(\mathcal{I} * A) = (D\mathcal{I}) * A + \mathcal{I} * DA = (D\mathcal{I}) * A + DA\,. \tag{86}$$

Since the left hand side also equals DA, one concludes that $(D\mathcal{I}) * A = 0$ for all A, and thus the expectation that $D\mathcal{I} = 0$. In the star algebra of open strings an identity state has been identified[62,63,64] that satisfies the expected properties for most well-behaved states.

Finally, if the identity string field is annihilated by the derivation Q, then

$$\langle Q\Psi\,, \mathcal{I}\rangle = -(-)^{\Psi}\langle\Psi, Q\mathcal{I}\rangle = 0\,. \tag{87}$$

The identification (67) then yields

$$0 = \int Q\Psi \star \mathcal{I} = \int Q\Psi\,, \tag{88}$$

which is property (b) in the axiomatic formulation of OSFT discussed in section 4.2.

5. String field theory: conformal field theory approach

A direct conformal field theory evaluation of the string action is perhaps the most economical way to proceed in the case of simple computations. We will explain this definition of the action, using at the same time the example of the action restricted to only the tachyon field to illustrate the definitions. The string action, written before in (61) is given by

$$S(\Phi) = -\frac{1}{g^2}\left[\frac{1}{2}\langle\,\Phi\,,\,Q\,\Phi\rangle + \frac{1}{3}\langle\,\Phi\,,\,\Phi * \Phi\rangle\right]. \tag{89}$$

This OSFT action can be used to describe the spacetime field theory of any D-brane. For example, for a Dp-brane we would have an underlying conformal field theory of a field X^0 and p fields X^i with Neumann boundary conditions, and $(25-p)$ fields X^a with Dirichlet boundary conditions. In our computations, we will assume that the brane has unit volume, in which case the mass M of the brane coincides with its tension. One can show that, in units where $\alpha' = 1$,

$$M = \frac{1}{2\pi^2}\frac{1}{g^2}\,. \tag{90}$$

We will prove this result in section 7.1.

We will evaluate the OSFT action by truncating the string field down to the zero momentum tachyon. The systematic approximation of the full theory by successive level truncation is described in detail in Section 7.3. In the level expansion this zero momentum tachyon is assigned level zero. The tachyon vertex operator is $e^{ipX(z)}c(z)$ and the associated state is $c_1|0;p\rangle$. The zero momentum tachyon state is $c_1|0;0\rangle$ or in simpler notation $c_1|0\rangle$. Since we have

$$L_0 c_1|0\rangle = -c_1|0\rangle\,, \tag{91}$$

the level ℓ of a state is related to the L_0 eigenvalue as

$$\ell = L_0 + 1\,. \tag{92}$$

The string field truncated to the zero momentum tachyon is written as

$$|T\rangle = t\,c_1|0\rangle\,, \tag{93}$$

where the variable t denotes the expectation value of the tachyon field, and it is a spacetime constant. The variable t is related to the tachyon field ϕ in the expansion (57) through

$$t = g\phi(0)\,. \tag{94}$$

As we alternate between notation Ψ and Φ for the string field, we will use ϕ and t for the zero-momentum tachyon. After truncating to just the tachyon degree of freedom t the tachyon potential $V(t)$ is just minus $S(|T\rangle)$ and thus

$$V(t) = -S(|T\rangle) = M(2\pi^2)\Big(\frac{1}{2}\langle T, QT\rangle + \frac{1}{3}\langle T, T, T\rangle\Big)\,. \tag{95}$$

In fact, it is convenient to define the ratio

$$f(t) \equiv \frac{V(t)}{M} = (2\pi^2)\Big(\frac{1}{2}\langle T, QT\rangle + \frac{1}{3}\langle T, T, T\rangle\Big)\,. \tag{96}$$

The function $f(t)$ is a rescaled version of the tachyon potential. By construction, $f(t)$ has a quadratic term and a cubic term, so $f(t=0) = 0$. The Sen conjecture requires that $f(t)$ have a critical point at $t = t^*$ that satisfies

$$f(t^*) = -1\,. \tag{97}$$

This is indeed equivalent to saying that the energy difference between the D-brane vacuum and the stable vacuum equals the energy M of the D-brane. It suggests strongly that the stable vacuum is a vacuum without a D-brane. It is perhaps useful to remark that $V(t)$ as obtained directly from the OSFT action does not convey the true gravitational picture where absolute vacuum energies are important. The vacuum with the D-brane, namely at $t = 0$ has a positive cosmological constant, or vacuum energy. This is in fact the D-brane energy. As the theory rolls to the stable vacuum, the vacuum energy goes to zero. Thus the tachyon potential $V(t)$ is missing an additive constant, which becomes important when coupling to gravity (which we will not consider in the present lectures). Such a constant term at least morally belongs in a more general OSFT action where the disk partition function would naturally appear as a field independent contribution to the string action. This disk partition function calculated with the boundary condition appropriate to the D-brane is in fact proportional to the D-brane energy.

5.1. *Kinetic term computations*

Let us begin the computation of the string action truncated to the tachyon by evaluating $\langle T, QT\rangle$. To this end we need to use the normalization condition

$$\langle 0|c_{-1}c_0c_1|0\rangle = 1\,, \tag{98}$$

which is appropriate if we compactify all coordinates (including time) into circles of unit circumference. Indeed, comparing with (45), we see that the right-hand side of (98) should have a $(2\pi)^{26}\delta(0)$, which is equivalent to the full spacetime volume V. In our full compactification, $V = 1$. The compactification of time is only a formal trick that facilitates computations but is not strictly necessary.

Exercise: Given $c(z) = \sum_n \frac{c_n}{z^{n-1}}$ show that

$$\langle 0|c(z_1)c(z_2)c(z_3)|0\rangle = (z_1 - z_2)(z_1 - z_3)(z_2 - z_3)\,. \tag{99}$$

Now that we must compute precisely we should make clear the CFT definition of the inner product

Definition: $\langle A, B\rangle = \langle bpz(A)|B\rangle$. Here $bpz : \mathcal{H} \to \mathcal{H}^*$ is BPZ conjugation, which we review next.

Given a primary field $\phi(z)$ of dimension d, it has a mode expansion

$$\phi(z) = \sum_n \frac{\phi_n}{z^{n+d}} \quad \to \quad \phi_n = \oint \frac{dz}{2\pi i} z^{n+d-1}\phi(z)\,. \tag{100}$$

We define

$$bpz(\phi_n) \equiv \oint \frac{dt}{2\pi i} t^{n+d-1}\phi(t)\,, \qquad \text{with} \quad t = -\frac{1}{z}\,. \tag{101}$$

Note that this simply defines the BPZ conjugation of the oscillator with the same formula as the oscillator itself (100) but referred to a coordinate at $z = \infty$. This integral is evaluated by using the transformation law

$$\phi(t)(dt)^d = \phi(z)(dz)^d\,. \tag{102}$$

We therefore get

$$bpz(\phi_n) \equiv -\oint \frac{dz}{2\pi i}\frac{1}{z^2}\left(-\frac{1}{z}\right)^{n+d-1}\phi(z)(z^2)^d\,. \tag{103}$$

The minus sign in front arises from a reversal of contour of integration (a contour circling $t = 0$ clockwise circles $z = 0$ counterclockwise). Moreover the transformation law was used to reexpress $\phi(t)$ in terms of the field $\phi(z)$ whose mode expansion is given. Simplifying the integral one finds

$$bpz(\phi_n) = (-1)^{n+d}\oint \frac{dz}{2\pi i} z^{-n+d-1}\phi(z) = (-1)^{n+d}\phi_{-n}\,. \tag{104}$$

In summary, we have shown that

$$bpz(\phi_n) = (-1)^{n+d}\phi_{-n}\,. \tag{105}$$

This equation defines BPZ conjugation when we supplement it with the rule

$$bpz\Big(\phi_{n_1}\cdots\phi_{n_p}|0\rangle\Big) = \langle 0|bpz(\phi_{n_1})\cdots bpz(\phi_{n_p})\,. \tag{106}$$

This formula is correct as stated also when the oscillators are anticommuting. The only condition for its validity is that the various modes with mode numbers of the same sign must commute (or anticommute). Otherwise BPZ conjugation produces a sequence of oscillators in *reverse* order.

A nontrivial example of the above rules arises when we calculate the BPZ conjugates of the modes L_n of the stress tensor. Although the stress tensor $T(z)$ is not a primary field, it transforms as a primary under SL(2,C) transformations, and therefore it does transform as a dimension two primary under the inversion needed in the definition of BPZ. Thus we have

$$bpz(L_n) = (-1)^n L_{-n}\,, \tag{107}$$

and for a string of oscillators we must write

$$bpz\Big(L_{n_1}\cdots L_{n_p}|0\rangle\Big) = \langle 0|bpz(L_{n_p})\cdots bpz(L_{n_1})\,. \tag{108}$$

Since $c(z)$ has dimension minus one, $bpz(c_1) = (-1)^{1+1}c_{-1} = c_{-1}$, so $bpz(c_1|0\rangle) = \langle 0|c_{-1}$. With this we have

$$\langle T, QT\rangle = t^2\langle 0|c_{-1}Qc_1|0\rangle\,. \tag{109}$$

Because of the form of the inner product only the term c_0L_0 in Q can contribute and we have

$$\langle T, QT\rangle = t^2\langle 0|c_{-1}c_0L_0c_1|0\rangle = -t^2\langle 0|c_{-1}c_0c_1|0\rangle = -t^2\,. \tag{110}$$

This completes the computation of the quadratic term in the tachyon potential. The negative sign obtained is the expected one, showing the instability of the $t = 0$ field configuration.

5.2. *Interaction term computation*

To compute the interaction of three tachyons we must explain how the three vertex is defined in CFT language. Consider three states A, B, and C and their associated vertex operators $\mathcal{O}_A$, $\mathcal{O}_B$, and $\mathcal{O}_C$. We define

$$\langle A, B, C\rangle \equiv \Big\langle f_1^D\circ\mathcal{O}_A(0),\, f_2^D\circ\mathcal{O}_B(0),\, f_3^D\circ\mathcal{O}_C(0)\Big\rangle_D \tag{111}$$

Here the right hand side denotes the CFT correlator of the conformal transforms of the vertex operators $\mathcal{O}_A$, $\mathcal{O}_B$, and $\mathcal{O}_C$. The conformal transforms are specified by the functions f_i as we explain now. Let there be three canonical coordinates ξ_i, with $i = 1,2,3$. The three functions $f_i(\xi_i)$ define maps from the upper half disks $\Im(\xi_i) \geq 0, |\xi_i| \leq 1$ into a disk D, with the points $\xi_i = 0$ being taken into points in the boundary of the disk D. The meaning of the conformal map of operators is that: *$f_i \circ \mathcal{O}_A(0)$ is the operator $\mathcal{O}_A(\xi_i = 0)$ expressed in terms of local operators at $f_i(\xi_i = 0)$*. The disk D may have the form of a unit disk, or can be the (conformally equivalent) upper half plane, or any other arbitrary form. Of course, the unit disk and the upper half plane are especially convenient for explicit computations.

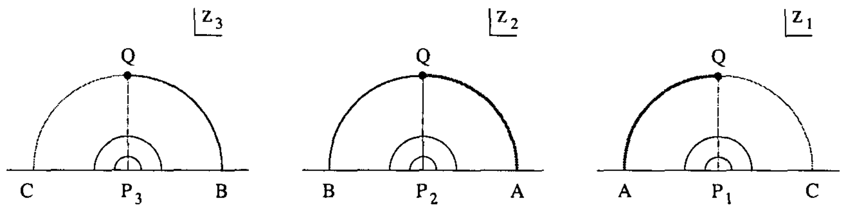


Figure 4. Representation of the cubic vertex as the gluing of 3 half-disks.

For the SFT at hand, the picture is given in Fig.4. The worldsheets of the three strings are represented as the unit half-disks $\{|\xi_i| \leq 1, \Im\xi \geq 0\}$, $i = 1,2,3$, in three copies of the complex plane. The boundaries $|\xi_i| = 1$ in the respective upper half-disks are *the* strings. Thus the point $\xi_i = i$ is the string midpoint. The interaction defining the vertex is built by gluing the three half-disks to form a single disk. This is done by the half-string identifications:

$$\begin{aligned}
\xi_1\xi_2 &= -1\,, \quad \text{for } |\xi_1| = 1, \;\; \Re(\xi_1) \leq 0\,, \\
\xi_2\xi_3 &= -1\,, \quad \text{for } |\xi_2| = 1, \;\; \Re(\xi_2) \leq 0\,, \\
\xi_3\xi_1 &= -1\,, \quad \text{for } |\xi_3| = 1, \;\; \Re(\xi_3) \leq 0\,.
\end{aligned} \tag{112}$$

Note that the common interaction point Q, is indeed $\xi_i = i$ (for $i = 1,2,3$), namely the mid-point of each open string $|\xi_i| = 1$, $\Im(\xi_i) \geq 0$. The left-half of the first string is glued with the right-half of the second string, and the same is repeated cyclically. This construction defines a specific 'three-punctured disk', a genus zero Riemann surface with a boundary, three marked points (punctures) on this boundary, and a choice of local coordinates ξ_i around each puncture.

The calculation of the functions $f_i^D(\xi)$ require a choice of disk D. We begin with the case when the disk D is simply chosen to be the interior of the unit disk $|w| < 1$, as shown in Fig. 5. In this case the functions $f_i^{D_w} \equiv f_i$ must map each half-disk to a 120° wedge of this unit disk. To construct the explicit maps that send ξ_i to the w plane, one notices that the SL(2,C) transformation

$$h(z) = \frac{1+i\xi}{1-i\xi}, \tag{113}$$

maps the unit upper–half disk $\{|\xi| \leq 1, \Im\xi \geq 0\}$ to the 'right' half–disk $\{|h| \leq 1, \Re h \geq 0\}$, with $z = 0$ going to $h(0) = 1$. Thus the functions

$$\begin{aligned} f_1(\xi_1) &= e^{\frac{2\pi i}{3}} \left(\frac{1+i\xi_1}{1-i\xi_1}\right)^{\frac{2}{3}}, \\ f_2(\xi_2) &= \left(\frac{1+i\xi_2}{1-i\xi_2}\right)^{\frac{2}{3}}, \\ f(\xi_3) &= e^{-\frac{2\pi i}{3}} \left(\frac{1+i\xi_3}{1-i\xi_3}\right)^{\frac{2}{3}}, \end{aligned} \tag{114}$$

will send the three half-disks to three wedges in the w plane of Fig. 5, with punctures at $e^{\frac{2\pi i}{3}}$, 1, and $e^{-\frac{2\pi i}{3}}$ respectively. This specification of the functions $f_i(\xi_i)$ gives the definition of the cubic vertex. In this representation cyclicity (*i.e.*, $\langle\Phi_1, \Phi_2, \Phi_3\rangle = \langle\Phi_2, \Phi_3, \Phi_1\rangle$) is manifest by construction. By SL(2,C) invariance, there are many other possible representations that give exactly the same off–shell amplitudes.

A useful choice is obtained by mapping the interacting w disk symmetrically to the upper half z-plane H. This is the convention that we shall mostly be using. We can therefore define the functions f_i^H by composition of the earlier maps f_i (that send the half-disks to the w unit disk) with the map $h^{-1}(w) = -i\,\frac{w-1}{w+1}$ that takes this unit disk to the upper–half–plane, with the three punctures on the real axis (Fig. 6),

$$\begin{aligned} f_1^H(\xi_1) &\equiv h^{-1} \circ f_1(\xi_1) = S(f_3^H(\xi_1)) \\ &= \sqrt{3} + \frac{8}{3}\,\xi_1 + \frac{16}{9}\,\sqrt{3}\,{\xi_1}^2 + \frac{248}{81}\,{\xi_1}^3 + O\left({\xi_1}^4\right) . \\ f_2^H(\xi_2) &\equiv h^{-1} \circ f_2(\xi_2) = S(f_1^H(\xi_2)) = \tan\left(\frac{2}{3}\,\arctan(\xi_2)\right) \\ &= \frac{2}{3}\,\xi_2 - \frac{10}{81}\,{\xi_2}^3 + O\left({z_2}^5\right) . \end{aligned}$$

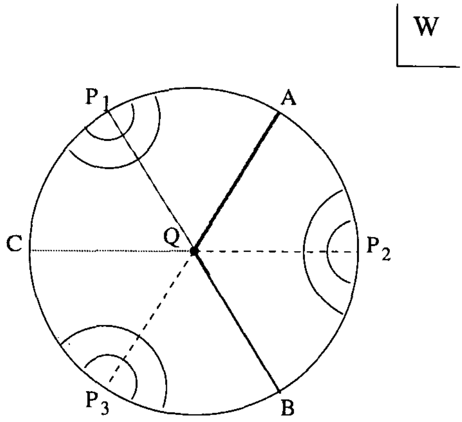


Figure 5. Representation of the cubic vertex as a 3–punctured unit disk.

$$
\begin{aligned}
f_3^H(\xi_3) &\equiv h^{-1} \circ f_3\,(\xi_3) = S(f_2^H(\xi_3)) \\
&= -\sqrt{3} + \frac{8}{3}\,\xi_3 - \frac{16}{9}\,\sqrt{3}\,{\xi_3}^2 + \frac{248}{81}\,{\xi_3}^3 + O({\xi_3}^4)\,. \qquad (115)
\end{aligned}
$$

The three punctures are at $f_1^H(0) = +\sqrt{3}, f_2^H(0) = 0, f_3^H(0) = -\sqrt{3}$, and the SL(2,R) map $S(z) = \frac{z-\sqrt{3}}{1+\sqrt{3}z}$ cycles them (thus $S \circ S \circ S(z) = z$).

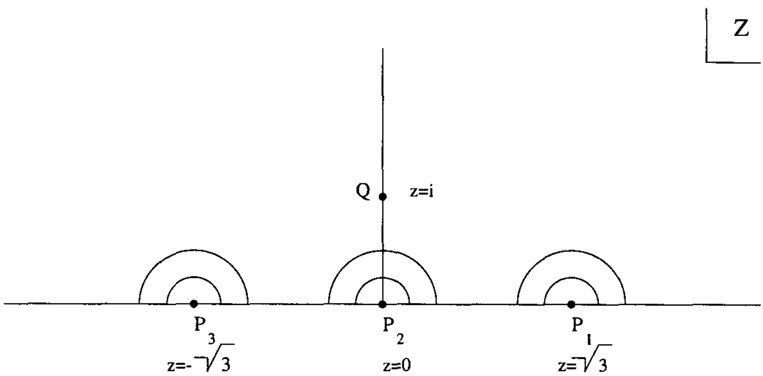


Figure 6. Representation of the cubic vertex as the upper–half plane with 3 punctures on the real axis.

This completes the definition of the string field theory action. When the disk D is presented as a unit disk the functions f_i in (111) are the functions given in equation (114). When the disk D is presented as the upper half plane H the relevant functions in (111) are the functions f_i^H given in (115) above.

Exercise: Verify explicitly by a *by-hand* calculation that the first two terms in the expansion of f_1^H and f_3^H, as well as the first term in f_2^H are correct.

Let us now return to the computation of the tachyon action. For our string field $|T\rangle = tc_1|0\rangle$ the interaction term $\langle T, T, T\rangle$ will be given by

$$\langle T, T, T\rangle = t^3\langle c_1, c_1, c_1\rangle\,. \tag{116}$$

Since the vertex operator associated to $c_1|0\rangle$ is $c(z)$, using (111) we write:

$$\langle T, T, T\rangle = t^3\langle f_1^H \circ c(0), f_2^H \circ c(0), f_3^H \circ c(0)\rangle_H \tag{117}$$

Since the field $c(z)$ is a primary of dimension minus one, we have

$$\frac{c(z)}{dz} = \frac{c(\xi)}{d\xi} \quad \to c(\xi) = \frac{c(z)}{\frac{dz}{d\xi}} \tag{118}$$

Therefore

$$f \circ c(0) \equiv c(\xi = 0) = \frac{c(f(0))}{f'(0)}\,. \tag{119}$$

Using equations (115) to read the values of $f_1^H(0)$ and $\frac{df_1^H}{d\xi}(0)$ we therefore get, for example,

$$f_1^H \circ c(0) = \frac{c(f_1^H(0))}{f_1^{H\prime}(0)} = \frac{c(\sqrt{3})}{8/3}\,. \tag{120}$$

The other two insertions are dealt with similarly, and we find

$$\begin{aligned}\langle T, T, T\rangle &= t^3\Big\langle \frac{c(\sqrt{3})}{8/3}, \frac{c(0)}{2/3}, \frac{c(-\sqrt{3})}{8/3}\Big\rangle_H \\ &= \frac{3^3}{2^7}\langle c(\sqrt{3})c(0)c(-\sqrt{3})\rangle_H = \frac{3^4\sqrt{3}}{2^6}\,,\end{aligned} \tag{121}$$

where in the last step we made use of (99). Our answer is therefore

$$\langle T, T, T\rangle = \frac{81\sqrt{3}}{64}t^3 \equiv t^3K^3\,, \quad \langle c_1, c_1, c_1\rangle = \frac{81\sqrt{3}}{64} = K^3\,. \tag{122}$$

This completes the calculation of an interaction term.

5.3. *A first test of the tachyon conjecture*

Having obtained the kinetic term of the tachyon truncated action in (110) and the cubic term in (122) we are now in a position to evaluate the function $f(t)$ in (96):

$$f(t) = 2\pi^2\Big(-\frac{1}{2}t^2 + \frac{1}{3}K^3t^3\Big)\,. \tag{123}$$

We must now find the (locally) stable critical point $t = t^*$ of this potential and evaluate the value of $f(t^*)$. It is clear the answer will not be the precise one $f = -1$, since we have truncated the string field dramatically. Nevertheless, we hope to get an answer that is reasonably close, if level expansion is supposed to make sense.

The equation of motion is

$$-t^* + t^{*2}K^3 = 0 \quad \rightarrow \quad t^* = \frac{1}{K^3}\,, \tag{124}$$

and substituting back we find

$$f(t^*) = -\frac{1}{3}\frac{\pi^2}{K^6} = -\pi^2\frac{2^{12}}{3^{10}} = -\pi^2\frac{4096}{59049} \simeq -0.684. \tag{125}$$

Thus is this simplest approximation, where we only kept the tachyon zero mode we have found that the critical point cancels about 70% of the D-brane energy. In section (7.3) we discuss the extension of this calculation to include massive string modes.

5.4. *String vertex in the CFT approach: Neumannology*

When doing explicit computations in OSFT we need to consider interactions of fields other than the tachyon. The explicit computation of the previous section becomes a lot more involved for massive fields, and it is useful to find an automated procedure to deal with such calculations. One such procedure is based on conformal field theory conservation laws. This is a very effective method, but we will not review it here since its explanation in Rastelli and Zwiebach[65] is self-contained. Another approach uses the explicit Fock representations of the string vertex. This will be our subject of interest here. We will provide a self-contained derivation of the Neumann coefficients that define the three string vertex both in the matter and in the ghost sector. In fact, our construction will be general and applies to three string interactions other than the one used in OSFT. We will determine the full structure of the three string vertex, except for the matter zero modes.

In the Fock space representation of the vertex, we must find a state $\langle V_3| \in \mathcal{H}^* \otimes \mathcal{H}^* \otimes \mathcal{H}^*$ such that for any Fock space states A, B and C one finds that

$$\langle A, B, C\rangle \equiv \langle V_3|A\rangle_{(1)}|B\rangle_{(2)}|B\rangle_{(3)} . \tag{126}$$

Since we provided in (111) a definition of the left hand side of the above equation, the vertex $\langle V_3|$ is implicitly defined. Our procedure will be general in that the functions $f_r(\xi)$ that map the canonical half-disks to the upper half plane will be kept arbitrary. There is a natural ansatz for the vertex:

$$\begin{aligned}\langle V_3| = \mathcal{N}((\langle 0|c_{-1}c_0)^{(3)})((\langle 0|c_{-1}c_0)^{(2)})((\langle 0|c_{-1}c_0)^{(1)}) \\ \exp\Big(-\frac{1}{2}\sum_{r,s}\sum_{n,m\geq 1} \alpha_m^{(r)} N_{mn}^{rs}\, \alpha_n^{(s)}\Big) \exp\Big(\sum_{r,s}\sum_{\substack{m\geq 0\\ n\geq 1}} b_m^{(r)} X_{mn}^{rs}\, c_n^{(s)}\Big) .\end{aligned} \tag{127}$$

Here $\mathcal{N}$ is a normalization factor, which will be determined shortly. In fact, its determination is essentially the tachyon computation of the previous section. Moreover, note that the nontrivial oscillator dependence in the matter sector is in the form of an exponential of a quadratic form. This is a general result that follows from the free field property of the matter CFT. Having just a quadratic form is possible also for the ghost sector, but it requires a careful choice of vacua. This is because there is a sum rule regarding ghost number– if the vacua are not chosen conveniently, extra linear ghost factors are necessary in the vertex. Since the vertex state $\langle V_3|$ is a bra we use out-vacua, in particular the vacua $\langle 0|c_{-1}c_0$. This is quite convenient because the ghost number conservation law is satisfied when each of the states A, B and C in (126) is of ghost number one. Indeed in each of the three state spaces we must have a total ghost number of three– two are supplied by the out-vacuum, and one by the in-state. This clearly allows the nontrivial ghost dependence of the vertex to be just a pure exponential with zero ghost number. A final point concerns the sum restrictions over the ghost oscillators. These simply arise because only oscillators that do not kill the vacua $\langle 0|c_{-1}c_0$ should appear in the exponential. Thus for the antighost oscillators b_m we find $m \geq 0$ and for the ghost oscillators c_n we find $n \geq 1$

The normalization factor $\mathcal{N}$ can be determined by finding the overlap of the vertex with three zero momentum tachyons $c_1|0\rangle$. In this case we have

$$\langle c_1, c_1, c_1\rangle = \langle V_3|c_1\rangle_{(1)}|c_1\rangle_{(2)}|c_1\rangle_{(3)} = \mathcal{N} , \tag{128}$$

since all oscillators in the exponentials kill the zero momentum tachyon. In (122) we found the value of this constant for the case of the OSFT vertex

$$\mathcal{N} = K^3 = \frac{3^{9/2}}{2^6}\,. \tag{129}$$

The calculation in the general case is not any more complicated and it is a good exercise!

Exercise: Show that for arbitrary functions $f_i(\xi)$, $i = 1,2,3$, that map half-disks to the UHP, the constant $\mathcal{N}$ in the vertex (127) is given by:

$$\mathcal{N} = \frac{(f_1(0) - f_2(0))(f_1(0) - f_3(0))(f_2(0) - f_3(0))}{f_1'(0)f_1'(0)f_1'(0)}\,. \tag{130}$$

Our goal now is to find explicit expressions for the Neumann coefficients N^{rs}_{mn} and X^{rs}_{mn} in terms of the functions f_i that define the vertex.

We begin with the matter sector, where the following conventions are used

$$i\partial X(z) = \sum \frac{\alpha_n}{z^{n+1}}\,, \quad \alpha_n = \oint \frac{dz}{2\pi i}\, z^n\, i\partial X\,, \tag{131}$$

$$\langle i\partial X(z)\; i\partial X(w)\rangle = \frac{1}{(z-w)^2}\,, \quad [\alpha_n, \alpha_m] = n\delta_{m+n,0}\,. \tag{132}$$

To find the matter Neumann coefficients we evaluate

$$M = \langle V_3|\; R\left(i\partial X^{(r)}(z)\; i\partial X^{(s)}(w)\right)\; c_1^{(1)}|0\rangle_{(1)}\, c_1^{(2)}|0\rangle_{(2)}\, c_1^{(3)}|0\rangle_{(3)} \tag{133}$$

in two different ways. In here $R(\dots)$ denotes radial ordering, necessary when $r = s$. For our first computation we use the mode expansion (131) of the conformal fields to find that

$$M = \langle V_3|\Big(\sum_{m,n} \frac{1}{z^{-m+1}}\frac{1}{w^{-n+1}}\,\alpha^{(r)}_{-m}\,\alpha^{(s)}_{-n} + \frac{\delta^{rs}}{(z-w)^2}\Big)c_1^{(1)}|0\rangle_{(1)}\, c_1^{(2)}|0\rangle_{(2)}\, c_1^{(3)}|0\rangle_{(3)} \tag{134}$$

and the oscillator form (127) of the vertex to obtain

$$M = -\mathcal{N}\sum_{m,n} z^{m-1}w^{n-1}\, mnN^{rs}_{mn} + \mathcal{N}\frac{\delta^{rs}}{(z-w)^2}\,. \tag{135}$$

In the second evaluation we first rewrite M as

$$M = \langle V_3|\; i\partial X^{(r)}(z)\; i\partial X^{(s)}(w)\, c^{(1)}(0)c^{(2)}(0)c^{(3)}(0)|0\rangle_{(1)}\, |0\rangle_{(2)}\, |0\rangle_{(3)}\,, \tag{136}$$

and reinterpret as a correlator, in the spirit of (111):

$$M = \Big\langle f_r \circ i\partial X(z)\, f_s \circ i\partial X(w)\, f_1 \circ c(0)\, f_2 \circ c(0)\, f_3 \circ c(0) \Big\rangle \,. \tag{137}$$

The ghost part of this correlator gives the factor $\mathcal{N}$. The matter part, using $i\partial X(z) = i\partial X(f(z))\frac{df}{dz}$, and (132) finally gives

$$M = \mathcal{N}\, f_r'(z)\, f_s'(w) \Big\langle\, i\partial X(f_r(z))\; i\partial X(f_s(w))\,\Big\rangle = \mathcal{N}\frac{f_r'(z)\, f_s'(w)}{(f_r(z)-f_s(w))^2}\,. \tag{138}$$

Equating the results (135) and (138) of the two evaluations of M we obtain:

$$\sum_{m,n} z^{m-1} w^{n-1}\, mn N^{rs}_{mn} - \frac{\delta^{rs}}{(z-w)^2} = -\frac{f_r'(z)\, f_s'(w)}{(f_r(z)-f_s(w))^2}\,. \tag{139}$$

It is now simple to pick up the coefficients N^{rs}_{mn} by contour integration over small circles surrounding $z=0$ and $w=0$. The second term on the left-hand side gives no contribution, and one finally finds

$$N^{rs}_{mn} = -\frac{1}{mn}\oint_0 \frac{dz}{2\pi i}\frac{1}{z^m}\oint_0 \frac{dw}{2\pi i}\frac{1}{w^n}\,\frac{f_r'(z)\, f_s'(w)}{(f_r(z)-f_s(w))^2}\,. \tag{140}$$

This is the desired expression for the Neumann coefficients of the matter sector. They can be used for an arbitrary vertex. The above contour integrals are straightforward to compute and they can be easily done by a computer in a series expansion. In terms of residues the expression above is equivalent to

$$N^{rs}_{mn} = -\frac{1}{mn}\mathrm{Res}_{z=0}\mathrm{Res}_{w=0}\Big[\frac{1}{z^m}\frac{1}{w^n}\,\frac{f_r'(z)\, f_s'(w)}{(f_r(z)-f_s(w))^2}\Big] \tag{141}$$

Exercise: Show that the contour integrals in (140) can be evaluated in any order. Do this both for the case when $r \neq s$ and for the case when $r = s$.

We now turn to the calculation of the ghost Neumann coefficients X^{rs}_{mn}. For this we need mode expansions and two point functions for the ghost CFT:

$$c(z) = \sum_n \frac{c_n}{z^{n-1}}\,, \quad b(z) = \sum_n \frac{b_n}{z^{n+2}}\,, \quad \langle c(z) b(w)\rangle = \frac{1}{z-w}\,. \tag{142}$$

The strategy is once more based on the computation of a certain expression in two different ways. Indeed, we consider the overlap

$$G = \langle V_3|\, R\Big(b^{(s)}(z)\; c^{(r)}(w)\Big)\; |\, c_1^{(1)}|0\rangle_{(1)}\, c_1^{(2)}|0\rangle_{(2)}\, c_1^{(3)}|0\rangle_{(3)} \tag{143}$$

and first evaluate it by using the mode expansion of the antighost and ghost fields, and then the explicit expression for the vertex in (127). In this way we find

$$G = \langle V_3| \Big(\sum_{m,n} \frac{1}{z^{-n+2}} \frac{1}{w^{-m-1}} b_{-n}^{(s)} c_{-m}^{(r)} + \frac{w}{z(z-w)} \Big) c_1^{(1)}|0\rangle_{(1)} c_1^{(2)}|0\rangle_{(2)} c_1^{(3)}|0\rangle_{(3)}$$

$$= \mathcal{N} \sum_{m,n} z^{-n+2} w^{-m-1} X_{mn}^{rs} + \mathcal{N} \frac{w}{z(z-w)} . \tag{144}$$

In the second computation G is interpreted as a correlator and we have

$$G = \Big\langle f_s \circ b(z)\; f_r \circ c(w)\; f_1 \circ c(0)\, f_2 \circ c(0)\, f_3 \circ c(0) \Big\rangle$$

$$= \frac{(f_s'(z))^2}{f_r'(w)} \frac{1}{f_1'(0) f_2'(0) f_3'(0)} \Big\langle b(f_s(z))\; c(f_r(w))\; c(f_1(0))\, c(f_2(0))\, c(f_3(0)) \Big\rangle , \tag{145}$$

where we used the standard conformal maps of the relevant operators, all of which are primary. The final correlator is in the upper half plane and all the field arguments refer to the coordinates in the upper half plane. The correlator can be calculated by using OPE's, but it is simpler to use the singularity structure and derive the normalization from a special configuration. Note, for example, that there must be zeroes when any pair of c fields approach each other. In particular, this will include a factor $(f_1(0) - f_2(0))(f_1(0) - f_3(0))(f_2(0) - f_3(0))$ as in $\mathcal{N}$ (see (130)). We will also have poles when the antighost approaches any ghost. These considerations imply that

$$G = \mathcal{N} \frac{(f_s'(z))^2}{f_r'(w)} \frac{1}{f_s(z) - f_r(w)} \frac{\prod_{I=1}^3 (f_r(w) - f_I(0))}{\prod_{J=1}^3 (f_s(z) - f_J(0))} . \tag{146}$$

We can now equate the results obtained in (144) and (146). Picking up the coefficients via contour integration, and noting that the second term on the right-hand side of (144) does not contribute for the relevant values of m and n, we find

$$X_{mn}^{rs} = \oint \frac{dz}{2\pi i} \frac{1}{z^{n-1}} \oint \frac{dw}{2\pi i} \frac{1}{w^{m+2}} \frac{(f_s'(z))^2}{f_r'(w)} \frac{1}{f_s(z) - f_r(w)} \frac{\prod_{I=1}^3 (f_r(w) - f_I(0))}{\prod_{J=1}^3 (f_s(z) - f_J(0)} . \tag{147}$$

This is the general result for the ghost Neumann coefficients. Again, for any vertex they are easily calculated by power series expansions and picking up residues. For particular vertices one can simplify somewhat the above expressions and find interesting relations. In fact, a fair amount of work can be done for the OSFT vertex in simplifying the above results. One can show that the matrices N and X are related, and while no closed form expressions are known for the coefficients, they can be generated quite efficiently from the simpler expressions.

Since any specific Neumann coefficient can be calculated exactly with a finite number of operations, the exact computation of $\langle A, B, C\rangle$ for any three Fock space states A, B, and C requires a finite number of operations, as well.

6. SFT action: oscillator approach

In this section, we give a more detailed discussion of Witten's open bosonic string field theory from the oscillator point of view. The main goal of this section is to explicitly formulate the OSFT action in the string Fock space, where the action (50) takes the form

$$S = -\frac{1}{2}\langle V_2|\Psi, Q\Psi\rangle - \frac{g}{3}\langle V_3|\Psi, \Psi, \Psi\rangle\,. \tag{148}$$

In this expression, $\langle V_2|$ and $\langle V_3|$ are elements of the two-fold and three-fold product of the dual Fock space $(\mathcal{H}^*)^2$ and $(\mathcal{H}^*)^3$, respectively. These objects defined in terms of the string Fock space give a rigorous definition to the abstract action (50) through the replacement

$$\begin{aligned}\langle V_2|A, B\rangle &\to \int A \star B\\ \langle V_3|A, B, C\rangle &\to \int A \star B \star C\,.\end{aligned}$$

Subsection 6.1 is a warmup, in which we review some basic features of the simple harmonic oscillator and discuss squeezed states. In subsection 6.2 we relate modes on the full string to modes on half strings, giving formulae needed to compute the three-string vertex. In subsection 6.3 we derive the two-string vertex in oscillator form, and in subsection 6.4 we give an explicit formula for the three-string vertex. In subsection 6.5 we put these pieces together and discuss the calculation of the full SFT action.

6.1. *Squeezed states and the simple harmonic oscillator*

Let us consider a simple harmonic oscillator with annihilation operator

$$a = -i\left(\sqrt{\frac{\alpha}{2}}x + \frac{1}{\sqrt{2\alpha}}\partial_x\right) \tag{149}$$

where α is an arbitrary constant. The oscillator ground state is associated with the wavefunction

$$|0\rangle \to \left(\frac{\alpha}{\pi}\right)^{1/4} e^{-\alpha x^2/2}\,. \tag{150}$$

In the harmonic oscillator basis $|n\rangle$, the Dirac position basis states $|x\rangle$ have a squeezed state form

$$|x\rangle = \left(\frac{\alpha}{\pi}\right)^{1/4} \exp\left(-\frac{\alpha}{2}x^2 - i\sqrt{2\alpha}a^\dagger x + \frac{1}{2}(a^\dagger)^2\right)|0\rangle\,. \tag{151}$$

A general wavefunction is associated with a state through the correspondence

$$f(x) \to \int_{-\infty}^{\infty} dx\, f(x)|x\rangle\,. \tag{152}$$

In particular, we have

$$\delta(x) \to \left(\frac{\alpha}{\pi}\right)^{1/4} \exp\left(\frac{1}{2}(a^\dagger)^2\right)|0\rangle\,, \tag{153}$$

$$1 \to \int dx\,|x\rangle = \left(\frac{4\pi}{\alpha}\right)^{1/4} \exp\left(-\frac{1}{2}(a^\dagger)^2\right)|0\rangle\,.$$

This shows that the delta and constant functions both have squeezed state representations in terms of the harmonic oscillator basis. The norm of a squeezed state

$$|s\rangle = \exp\left(\frac{1}{2}s(a^\dagger)^2\right)|0\rangle \tag{154}$$

is given by

$$\langle s|s\rangle = \frac{1}{\sqrt{1-s^2}}\,. \tag{155}$$

The states (153) are non-normalizable, but since they have $s = \pm 1$, they are right on the border of normalizability. The states (153) can be used to calculate, just like we do with the Dirac basis states $|x\rangle$, which lie outside the single-particle Hilbert space.

It will be useful for us to generalize the foregoing considerations in several ways. A particularly simple generalization arises when we consider a pair of degrees of freedom x, y described by a two-harmonic oscillator Fock space basis. In such a basis, repeating the preceding analysis leads us to a function-state correspondence for the delta functions relating x, y of the form

$$\delta(x \pm y) \to \exp\left(\pm a^\dagger_{(x)} a^\dagger_{(y)}\right)\left(|0\rangle_x \otimes |0\rangle_y\right)\,. \tag{156}$$

Note that this result is independent of α; like the δ function, the resulting state is again non-normalizable. we will find these squeezed state expressions very useful in describing the two- and three-string vertices of Witten's open string field theory. It is worth pointing out here that there are several ways of deriving (156). The most straightforward way is to carry out a two-dimensional Gaussian integral analogous to (153). We can also derive (156) indirectly, however (at least up to an overall constant) from the following argument. From the general result that delta functions give squeezed states, we expect that up to an overall constant

$$\delta(x \pm y) \rightarrow \tag{157}$$
$$|D_\pm\rangle = \exp\left(\pm\frac{1}{2}\left[Aa^\dagger_{(x)}a^\dagger_{(x)} + 2Ba^\dagger_{(x)}a^\dagger_{(y)} + Ca^\dagger_{(y)}a^\dagger_{(y)}\right]\right)(|0\rangle_x \otimes |0\rangle_y)\ .$$

The state associated with the delta function must satisfy the constraints

$$\begin{aligned}(x \pm y)|D_\pm\rangle &= 0 \\ (p_x \mp p_y)|D_\pm\rangle &= 0\,.\end{aligned} \tag{158}$$

Rewriting x, p_x in terms of $a_{(x)}, a^\dagger_{(x)}$ and similarly for y, p_y, these conditions impose the constraints

$$\begin{aligned}\left[(A \pm B - 1)a^\dagger_{(x)} + (B \pm C \pm 1)a^\dagger_{(y)}\right]|D_\pm\rangle &= 0 \\ \left[(A \mp B + 1)a^\dagger_{(x)} + (B \mp C \mp 1)a^\dagger_{(y)}\right]|D_\pm\rangle &= 0\,,\end{aligned} \tag{159}$$

from which it follows that $A = C = 0$ and $B = \pm 1$, reproducing (156) up to an overall constant. We will use this indirect method, following Gross and Jevicki, to derive the three-string vertex in subsection 6.4.

6.2. *Half-string modes*

For many computations it is useful to think of the string as being "split" into a left half and a right half. Formally, the string field can be expressed as a functional $\Psi[L, R]$, where L, R describe the left and right parts of the string. This is a very appealing idea, since it leads to a simple picture of the star product in terms of matrix multiplication

$$(\Psi \star \Phi)[L, R] = \int \mathcal{D}A\ \Psi[L, A]\Phi[A, R]\,. \tag{160}$$

While there has been quite a bit of work aimed at making this "split string" formalism precise,[66,67,68] the technical details in this approach become quite complicated when one attempts to precisely deal with the string midpoint where the left and right parts of the string attach. In particular, the BRST operator Q_B becomes rather awkward in this formulation.

Nonetheless, some of the structure of the star product encoded in the three-string vertex is easiest to understand using the half-string formalism, and many formulae related to the 3-string vertex are most easily expressed in terms of the linear map from full-string modes to half-string modes. In this subsection we discuss this linear map, encoded in a matrix X, which we use in subsection 6.4 to give an explicit formulae for the three-string vertex.

Recall that the matter fields are expanded in modes through

$$x(\sigma) = x_0 + \sqrt{2}\sum_{n=1}^{\infty} x_n \cos n\sigma\,. \tag{161}$$

(We suppress Lorentz indices in most of this section for clarity.) We are interested in considering an analogous expansion of the left and right halves of the string. We expand in odd modes with Neumann boundary conditions at the ends of the string, and Dirichlet boundary conditions at the string midpoint:

$$\begin{aligned} l(\sigma) &= x(\sigma) = \sqrt{2}\sum_{k=0}^{\infty} l_{2k+1}\cos(2k+1)\sigma, \quad \sigma < \pi/2 \\ r(\sigma) &= x(\pi-\sigma) = \sqrt{2}\sum_{k=0}^{\infty} r_{2k+1}\cos(2k+1)\sigma, \quad \sigma < \pi/2\,. \end{aligned} \tag{162}$$

Note that there are subtleties associated with the midpoint in this expansion. For example, while we have taken $l(\pi/2)$ to formally vanish, by choosing coefficients like $l_{2k+1} = (-1)^k 2\sqrt{2}a/(2k+1)\pi$ we have $l(\sigma) = a, \forall\sigma < \pi/2$, so $\lim_{\sigma\to\pi/2_-} = a$. These subtleties become important when dealing with the full theory, but are not important in the calculation we carry out below of the three-string vertex.

Let us define the quantities

$$\begin{aligned} X_{2k+1,2n} = X_{2n,2k+1} &= \frac{4(-1)^{k+n}(2k+1)}{\pi\left((2k+1)^2-4n^2\right)} \qquad (n\neq 0)\,, \\ X_{2k+1,0} = X_{0,2k+1} &= \frac{2\sqrt{2}(-1)^k}{\pi(2k+1)}\,. \end{aligned} \tag{163}$$

The matrix

$$X = \begin{pmatrix} 0 & X_{2k+1,2n} \\ X_{2n,2k+1} & 0 \end{pmatrix} \equiv \begin{pmatrix} 0 & X_{oe} \\ X_{eo} & 0 \end{pmatrix} \tag{164}$$

where e, o refer to the set of even and odd indices respectively, is manifestly symmetric and turns out to be orthogonal: $X = X^T = X^{-1}$. Performing a Fourier decomposition we can relate the full-string and half-string modes through

$$\begin{aligned} x_{2n+1} &= \tfrac{1}{2}\left(l_{2n+1} - r_{2n+1}\right), \qquad (165) \\ x_{2n} &= \tfrac{1}{2}\sum_{k=0}^{\infty} X_{2n,2k+1}\left(l_{2k+1} + r_{2k+1}\right), \end{aligned}$$

We can invert (165) to derive

$$\begin{aligned} l_{2k+1} &= x_{2k+1} + \sum_{n=0}^{\infty} X_{2k+1,2n} x_{2n}, \qquad (166) \\ r_{2k+1} &= -x_{2k+1} + \sum_{n=0}^{\infty} X_{2k+1,2n} x_{2n}. \end{aligned}$$

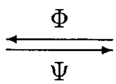
One must be careful with the order of summation in sequences of coefficients which do not go to zero faster than $1/n$: different orders of summation may give different results. Fortunately, such associativity anomalies are not relevant for the calculations we do here.

6.3. *The two-string vertex* $\langle V_2|$

We can immediately apply the oscillator formulae from subsection 6.1 to calculate the two-string vertex. Using the mode decomposition (161), we associate the string field functional $\Psi[x(\sigma)]$ with a function $\Psi(\{x_n\})$ of the infinite family of string oscillator mode amplitudes. The overlap integral combining (60) and (59) can then be expressed in modes as

$$\int \Psi \star \Phi = \int \prod_{n=0}^{\infty} dx_n dy_n \; \delta(x_n - (-1)^n y_n) \Psi(\{x_n\}) \Phi(\{y_n\}). \qquad (167)$$

Geometrically this just encodes the overlap condition $x(\sigma) = y(\pi - \sigma)$ described through



It follows from (156) that we can write the two-string vertex as the squeezed state

$$\langle V_2|_{\text{matter}} = (\langle 0| \otimes \langle 0|) \exp\Big(\sum_{n,m=0}^{\infty} -a_n^{(1)} C_{nm} a_m^{(2)} \Big), \qquad (168)$$

where $C_{nm} = \delta_{nm}(-1)^n$ is an infinite-size matrix connecting the oscillator modes of the two single-string Fock spaces, and the sum is taken over all oscillator modes, including zero. In the expression (168), we have used the formalism in which $|0\rangle$ is the vacuum annihilated by a_0. To translate this expression into a momentum basis, we use only $n, m > 0$, and replace

$$(\langle 0| \otimes \langle 0|) \exp\left(-a_0^{(1)} a_0^{(2)}\right) \to \int d^{26}p \, (\langle 0; p| \otimes \langle 0; -p|) \, . \tag{169}$$

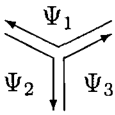
The extension of this analysis to ghosts is straightforward. For the ghost and antighost respectively, the overlap conditions corresponding with $x_1(\sigma) = x_2(\pi - \sigma)$ are[62] $c_1(\sigma) = -c_2(\pi - \sigma)$ and $b_1(\sigma) = b_2(\pi - \sigma)$. This leads to the overall formula for the two-string vertex

$$\begin{aligned} \langle V_2| = & \int d^{26}p \, (\langle 0; p| \otimes \langle 0; -p|) \, (c_0^{(1)} + c_0^{(2)}) \qquad (170) \\ & \times \exp\left(- \sum_{n=1}^{\infty} (-1)^n [a_n^{(1)} a_n^{(2)} + c_n^{(1)} b_n^{(2)} + c_n^{(2)} b_n^{(1)}] \right) . \end{aligned}$$

This expression for the two-string vertex can also be derived directly from the conformal field theory approach, computing the two-point function of an arbitrary pair of states on the disk.

6.4. *The three-string vertex* $|V_3\rangle$

The three-string vertex, which is associated with the three-string overlap



can be computed in a very similar fashion to the two-string vertex above. The details of the calculation, however, are significantly more complicated. In this subsection we follow the original approach of Gross and Jevicki;[62] similar approaches were taken by other authors.[69,70] The method used by Gross and Jevicki is essentially the method used in (156) to write a delta function of two variables in oscillator form by imposing the constraints (159) on a general squeezed state. The 3-string vertex can also be computed by explicitly performing[71,72] the relevant Gaussian integrals.[a]

---

[a]Another approach to the cubic vertex has been explored extensively. By diagonalizing the Neumann matrices, the star product takes the form of a continuous Moyal

From the general structure of the overlap conditions it is clear that, like the two-string vertex, the three-string vertex takes the form of a squeezed state:

$$|V_3\rangle = \kappa \int d^{26}p^{(1)} d^{26}p^{(2)} d^{26}p^{(3)}$$
$$\times \exp\Big(-\frac{1}{2}\sum_{r,s=1}^{3} [a_m^{(r)} V_{mn}^{rs} a_n^{(s)} + 2a_m^{(r)} V_{m0}^{rs} p^{(s)} + p^{(r)} V_{00}^{rs} p^{(s)} + c_m^{(r)} X_{mn}^{rs} b_n^{(s)}]\Big)\,,$$
$$\times \delta(p^{(1)} + p^{(2)} + p^{(3)}) c_0^{(1)} c_0^{(2)} c_0^{(3)} \Big(|0; p^{(1)}\rangle \otimes |0; p^{(2)}\rangle \otimes |0; p^{(3)}\rangle\Big) \tag{171}$$

where $\kappa = \mathcal{N} = K^3 = 3^{9/2}/2^6$, and where the Neumann coefficients V_{mn}^{rs} and X_{mn}^{rs} are constants. Writing the momentum basis states in oscillator form

$$|p\rangle = \frac{1}{(\pi)^{1/4}} \exp\left[-\frac{1}{2}p^2 + \sqrt{2}a_0^\dagger p - \frac{1}{2}(a_0^\dagger)^2\right] |0\rangle, \tag{172}$$

we can write matter part of the 3-string vertex as

$$|V_3\rangle = \tag{173}$$
$$\left(\frac{2\,\pi^{1/4}}{\sqrt{3}(1+V_{00})}\right)^{26} \exp\Big(-\frac{1}{2}\sum_{r,s\leq 3}\sum_{m,n\geq 0} V_{mn}^{\prime rs}(a_m^{(r)\dagger}\cdot a_n^{(s)\dagger})\Big) \left(|0\rangle \otimes |0\rangle \otimes |0\rangle\right)$$

where $V_{00} = V_{00}^{rr}$ and

$$\begin{aligned}
V_{mn}^{\prime rs} &= V_{mn}^{rs} - \frac{1}{1+V_{00}} \sum_t V_{m0}^{rt} V_{0n}^{ts} \\
V_{m0}^{\prime rs} &= V_{0m}^{\prime sr} = \frac{\sqrt{2}}{1+V_{00}} V_{m0}^{rs} \\
V_{00}^{\prime rs} &= \frac{2}{3(1+V_{00})} + \delta^{rs}\left(1 - 2/(1+V_{00})\right)\,.
\end{aligned} \tag{174}$$

We now want to determine the coefficients $V_{mn}^{\prime rs}$ by using overlap conditions analogous to (159). It is convenient to use a $\mathbf{Z}_3$ Fourier decomposition of the three string modes $x^{(i)}(\sigma)$

$$\begin{aligned}
Q &= \frac{1}{\sqrt{3}}(x^{(1)} + \omega x^{(2)} + \omega^2 x^{(3)}) \\
Q^{(3)} &= \frac{1}{\sqrt{3}}(x^{(1)} + x^{(2)} + x^{(3)})
\end{aligned} \tag{175}$$

---

product.[7,73] This simplifies the vertex but complicates the propagator. For a recent discussion of this work, applications of this approach, and further references, see the review of Bars.[9]

where $\omega = \exp(2\pi i/3)$. The definitions (175) can be used to define $Q(\sigma)$ in terms of $x^{(i)}(\sigma)$ as well as to define Q_n in terms of $x_n^{(i)}$ (and similarly for $Q^{(3)}$); henceforth by Q we denote the collection of full-string modes Q_n (and similarly for $Q^{(3)}$). We can relate the full-string modes $Q, Q^{(3)}$ to half-string modes $L, R, L^{(3)}, R^{(3)}$ through the equations (166). In terms of these variables, the overlap conditions are

$$\begin{aligned} L - \omega R &= 0, \\ L^{(3)} - R^{(3)} &= 0 . \end{aligned} \tag{176}$$

In terms of the even and odd full-string modes (which, using the same notation as in section 6.2, we denote by $Q_{\text{e,o}}$) these conditions are expressed as

$$Q_o - i\sqrt{3} X_{oe} Q_e = 0 , \tag{177}$$

and

$$Q_o^{(3)} = 0 . \tag{178}$$

Multiplying by X_{eo}, (177) can be rewritten as

$$\frac{i}{\sqrt{3}} X_{eo} Q_o + Q_e = 0 . \tag{179}$$

This can be combined with (177) and written in the simpler form

$$(1 - Y)Q = 0, \tag{180}$$

where

$$Y = -\frac{1}{2}C + \frac{\sqrt{3}}{2}X . \tag{181}$$

Note that $Y^2 = 1$. Similarly, we can write (178) as

$$(1 - C)Q^{(3)} = 0 . \tag{182}$$

Equations (180) and (182) are the essential overlap equations satisfied by the three-string vertex. Writing the three-string vertex as a squeezed state in terms of oscillators $A, A^{(3)}$ related to the string oscillators $a^{(i)}$ through the analogue of (175), we then have

$$|V_3\rangle \sim \exp\left(-A^\dagger U \bar{A}^\dagger - \frac{1}{2}(A^{(3)})^\dagger C (A^{(3)})^\dagger \right) , \tag{183}$$

where U satisfies the overlap constraint (180). Recall that the string modes are proportional to

$$x \sim E(a - a^\dagger) \tag{184}$$

where

$$E_{mn} = \delta_{mn}\frac{1}{\sqrt{m}}, \quad m \neq 0, \quad E_{00} = 1/\sqrt{2}\,. \tag{185}$$

Thus, from (180) and (183) we see that U must satisfy the overlap constraint

$$(1 - Y)E(1 + U) = 0\,. \tag{186}$$

As we discussed in the last part of subsection 6.1, associated with this constraint there is an analogous constraint on the derivatives in the perpendicular direction. Since $Y^2 = 1$, we have

$$(1 + Y)(1 - Y) = (1 - Y)(1 + Y) = 0\,. \tag{187}$$

Since derivatives with respect to the x modes go as

$$\partial \sim E^{-1}(a + a^\dagger), \tag{188}$$

we have the additional overlap constraint on U

$$(1 + Y)E^{-1}(1 - U) = 0\,. \tag{189}$$

Equations (186) and (189) determine U completely, giving

$$U = (2 - EYE^{-1} + E^{-1}YE)[EYE^{-1} + E^{-1}YE]^{-1}. \tag{190}$$

Unfortunately, the matrix combination in brackets is difficult to explicitly invert. This does, however, give a closed form expression for the three-string vertex (173), where

$$\begin{aligned} V'^{rr} &= \frac{1}{3}(C + U + \bar{U}) \\ V'^{r,r\pm 1} &= \frac{1}{6}(2C - U - \bar{U}) \pm \frac{i\sqrt{3}}{6}(U - \bar{U})\,. \end{aligned} \tag{191}$$

While (190) is difficult to directly compute, given a formula for U one can check that the formula is correct by checking the overlap conditions (186) and (189). Expressions for V and X and hence for U and V' were computed[62] by essentially the method used in the previous section. Their results for V and X are given as follows[b]. Define A_n, B_n for $n \geq 0$ through

$$\left(\frac{1+ix}{1-ix}\right)^{1/3} = \sum_{n\,\text{even}} A_n x^n + i \sum_{m\,\text{odd}} A_m x^m \tag{192}$$
$$\left(\frac{1+ix}{1-ix}\right)^{2/3} = \sum_{n\,\text{even}} B_n x^n + i \sum_{m\,\text{odd}} B_m x^m .$$

These coefficients can be used to define 6-string Neumann coefficients $N_{nm}^{r,\pm s}$ through

$$N_{nm}^{r,\pm r} = \begin{cases} \frac{1}{3(n\pm m)}(-1)^n(A_n B_m \pm B_n A_m)\,, & m+n\,\text{even},\ m \neq n \\ 0\,, & m+n\,\text{odd} \end{cases}$$

$$N_{nm}^{r,\pm(r+\sigma)} = \begin{cases} \frac{1}{6(n\pm \sigma m)}(-1)^{n+1}(A_n B_m \pm \sigma B_n A_m)\,, & m+n\,\text{even},\ m \neq n \\ \sigma \frac{\sqrt{3}}{6(n\pm\sigma m)}(A_n B_m \mp \sigma B_n A_m)\,, & m+n\,\text{odd} \end{cases} \tag{193}$$

where in $N^{r,\pm(r+\sigma)}$, $\sigma = \pm 1$, and $r+\sigma$ is taken modulo 3 to be between 1 and 3. The 3-string matter Neumann coefficients V_{nm}^{rs} are then given by

$$\begin{aligned} V_{nm}^{rs} &= -\sqrt{mn}(N_{nm}^{r,s} + N_{nm}^{r,-s}), \quad m \neq n, \text{ and } m,n \neq 0 \\ V_{nn}^{rr} &= -\frac{1}{3}\left[2\sum_{k=0}^{n}(-1)^{n-k}A_k^2 - (-1)^n - A_n^2\right], \quad n \neq 0 \\ V_{nn}^{r,r+\sigma} &= \frac{1}{2}\left[(-1)^n - V_{nn}^{rr}\right], \quad n \neq 0 \\ V_{0n}^{rs} &= -\sqrt{2n}\left(N_{0n}^{r,s} + N_{0n}^{r,-s}\right), \quad n \neq 0 \\ V_{00}^{rr} &= \ln(27/16) \end{aligned} \tag{194}$$

[b]Note that in some references, signs and various factors in κ and the Neumann coefficients may be slightly different. In some papers, the cubic term in the action is taken to have an overall factor of $g/6$ instead of $g/3$; this choice of normalization gives a 3-tachyon amplitude of g instead of $2g$, and gives a different value for κ. Often, the sign in the exponential of (171) is taken to be positive, which changes the signs of the coefficients V_{nm}^{rs}, X_{nm}^{rs}. When the matter Neumann coefficients are defined with respect to the oscillator modes α_n rather than a_n, the matter Neumann coefficients V_{nm}^{rs}, V_{n0}^{rs} must be divided by $\sqrt{nm}$ and $\sqrt{n}$. This is the case for the coefficients N_{nm}^{rs} computed in (140), which are related to the V's through $N_{nm}^{rs} = V_{nm}^{rs}/\sqrt{nm}$. Finally, when α' is taken to be $1/2$, an extra factor of $1/\sqrt{2}$ appears for each 0 subscript in the matter Neumann coefficients.

The ghost Neumann coefficients $X^{rs}_{mn}, m \geq 0, n > 0$ are given by

$$\begin{aligned}
X^{rr}_{mn} &= \left(-N^{r,r}_{nm} + N^{r,-r}_{nm}\right), \quad n \neq m \\
X^{r(r\pm 1)}_{mn} &= m\left(\pm N^{r,r\mp 1}_{nm} \mp N^{r,-(r\mp 1)}_{nm}\right), \quad n \neq m \\
X^{rr}_{nn} &= \frac{1}{3}\left[-(-1)^n - A_n^2 + 2\sum_{k=0}^{n}(-1)^{n-k}A_k^2 - 2(-1)^n A_n B_n\right] \\
X^{r(r\pm 1)}_{nn} &= -\frac{1}{2}(-1)^n - \frac{1}{2}X^{rr}_{nn}
\end{aligned} \tag{195}$$

These expressions for the matter and ghost Neumann coefficients were computed by Gross and Jevicki,[62] and include minor corrections published later.[74] It was shown that the resulting matter matrices U indeed satisfy the overlap conditions (186) and (189). This shows that the conformal field theory method and the oscillator method give the same results for the matter part of the three-string vertex. The same is true for the ghost part of the vertex, although we will not go into the details of this discussion here.

Before leaving the three-string vertex, it is worth noting that the Neumann coefficients have a number of simple symmetries. There is a cyclic symmetry under $r \to r+1, s \to s+1$, which corresponds to the obvious geometric symmetry of rotating the vertex. The coefficients are also symmetric under the exchange $r \leftrightarrow s, n \leftrightarrow m$. Finally, there is a twist symmetry which, as discussed in section 4.3, is associated with reflection of the strings

$$\begin{aligned}
V^{rs}_{nm} &= (-1)^{n+m} V^{sr}_{nm} \\
X^{rs}_{nm} &= (-1)^{n+m} X^{sr}_{nm} .
\end{aligned} \tag{196}$$

This symmetry follows from the fact that half-strings carrying odd modes pick up a minus sign under reflection. Since each string carrying an odd mode gets two changes of sign, from the two ends of the string, it is straightforward to see that this symmetry guarantees that the three-vertex is invariant under reflection, and therefore satisfies condition (76).

6.5. *Calculating the SFT action*

Given the action

$$S = -\frac{1}{2}\langle V_2|\Psi, Q\Psi\rangle - \frac{g}{3}\langle V_3|\Psi, \Psi, \Psi\rangle , \tag{197}$$

and the explicit formulae (170, 171) for the two- and three-string vertices, we can in principle calculate the string field action term by term for each of the fields in the string field expansion

$$\Psi = \int d^{26}p \; \left[\phi(p)\;|0_1;p\rangle + A_\mu(p)\;\alpha^\mu_{-1}|0_1;p\rangle + \chi(p)b_{-1}c_0|0_1;p\rangle \right.$$
$$\left. +B_{\mu\nu}(p)\alpha^\mu_{-1}\alpha^\nu_{-1}|0_1;p\rangle + \cdots\right]. \qquad (198)$$

Since the resulting action has an enormous gauge invariance given by (54), it is often helpful to fix the gauge before computing the action. A particularly useful gauge choice is the Feynman-Siegel gauge

$$b_0|\Psi\rangle = 0. \qquad (199)$$

This is a good gauge choice locally, fixing the linear gauge transformations $\delta|\Psi\rangle = Q|\Lambda\rangle$. This gauge choice is not, however, globally valid; we will return to this point in subsection 7.4. In this gauge, all fields in the string field expansion which are associated with states that have an antighost zero-mode c_0 are taken to vanish. For example, the field $\chi(p)$ in (198) vanishes. In Feynman-Siegel gauge, the BRST operator takes the simple form

$$Q = c_0L_0 = c_0(N + p^2 - 1) \qquad (200)$$

where N is the total (matter + ghost) oscillator number.

Using (200), it is straightforward to write the quadratic terms in the string field action. They are

$$\frac{1}{2}\langle V_2|\Psi, Q\Psi\rangle = \int d^{26}p\;\left\{\phi(-p)\left[\frac{p^2-1}{2}\right]\phi(p) + A_\mu(-p)\left[\frac{p^2}{2}\right]A^\mu(p) + \cdots\right\}. \qquad (201)$$

The cubic part of the action can also be computed term by term, although the terms are somewhat more complicated. The leading terms in the cubic action are given by

$$\frac{1}{3}\langle V_3|\Psi,\Psi,\Psi\rangle = \int d^{26}pd^{26}q\;\frac{\kappa g}{3}\;e^{(\ln 16/27)(p^2+q^2+p\cdot q)} \qquad (202)$$
$$\times\left\{\phi(-p)\phi(-q)\phi(p+q) + \frac{16}{9}A^\mu(-p)A_\mu(-q)\phi(p+q)\right.$$
$$\left. -\frac{8}{9}(p^\mu + 2q^\mu)(2p^\nu + q^\nu)A^\mu(-p)A_\nu(-q)\phi(p+q) + \cdots\right\}.$$

In computing the ϕ^3 term we have used

$$V_{00}^{rs} = \delta^{rs}\ln(\frac{27}{16}). \qquad (203)$$

The $A^2\phi$ term uses

$$V_{11}^{rs} = -\frac{16}{27}, \quad r \neq s, \tag{204}$$

while the $(A \cdot p)^2\phi$ term uses

$$V_{10}^{12} = -V_{10}^{13} = -\frac{2\sqrt{2}}{3\sqrt{3}}. \tag{205}$$

The most striking feature of this action is that for a generic set of three fields, there is a *nonlocal* cubic interaction term that contains an exponential of a quadratic form in the momenta. This means that the target space formulation of string theory has a dramatically different character from a standard quantum field theory. From the point of view of quantum field theory, string field theory seems to contain an infinite number of nonrenormalizable interactions. Just like the simpler case of noncommutative field theories, however, the magic of string theory seems to combine this infinite set of interactions into a sensible model. It has been shown that all on-shell amplitudes computed from the string field theory action we have discussed here precisely reproduce the amplitudes given by the usual conformal field theory approach, including the correct measure on moduli space.[75,76,77] Note, though, that the bosonic open theory becomes problematic at the quantum level because of the closed string tachyon, whose instability is not yet understood. For the purposes of these lectures, we will restrict our attention to the classical bosonic open string action. Open superstring field theory should be better behaved since the closed string sector has no tachyon. There has been significant progress in understanding tachyon condensation in superstring field theory,[37,78] even though superstring field theory is less developed than bosonic string field theory.

7. Evidence for the Sen conjectures

In this section we review the evidence from Witten's OSFT for Sen's conjectures. Subsection 7.1 contains a derivation of the formula for the tension of a bosonic D-brane. In subsection 7.2, a general discussion is given of symmetries in the string field theory action and resulting constraints on the set of string fields which take nonzero values in the tachyon vacuum. Subsection 7.3 contains a summary of existing results for the determination of the stable vacuum in Witten's OSFT (Sen's first conjecture), including some results which appeared after these lectures were originally given in 2001. In subsection 7.4 we discuss the Feynman-Siegel gauge choice and its limitations. Subsection 7.5 summarizes results on lower-dimensional D-branes as solitons in OSFT (Sen's second conjecture). Subsection 7.6 discusses the general problem of finding all open string backgrounds within OSFT. Sen's third conjecture is discussed in the following section 8, which is completely devoted to a discussion of the physics in the stable vacuum (vacuum string field theory).

7.1. *Tension of bosonic Dp-branes*

In this section we learn how to relate the open string coupling constant of string field theory to the mass of the D-brane described by the open string field theory. The material presented in this subsection is an unpublished result due to Ashoke Sen,[79] who cited the result in the paper[39] on the subject of universality. Subsequently, the closely related computation for the superstring was explained in detail.[80] For an alternative check of the result, see Appendix A of the paper by Okawa.[81]

In general, an open string field theory is formulated using a BCFT (boundary conformal field theory) which describes some D-brane (or a configuration of D-branes). In order to describe a D-brane with finite mass, we consider a compactification of p spatial coordinates and wrap a Dp-brane around along these dimensions. The string field theory associated with this D-brane is written as before:

$$S(\Phi) = -\frac{1}{g^2}\left[\frac{1}{2}\langle\,\Phi\,,\,Q\,\Phi\rangle + \frac{1}{3}\langle\,\Phi\,,\,\Phi * \Phi\rangle\right]. \tag{206}$$

The D-brane in question is perceived by the effective $(25-p)$-dimensional observer as a point particle. The BCFT includes a Neumann field X^0, a set of Dirichlet fields X^i, with $i = 1, \ldots, 25-p$ and some set of Neumann fields X^a, with $a = 25-p+1, \ldots, 25$ that describe the internal sector of the BCFT. The string field theory effectively describes an infinite collection of fields $\phi_i(t, x^a)$. These fields do not depend on $x^1, \ldots, x^{25-p}$ because the corresponding string coordinates are Dirichlet. Since the coordinates x^a are compact, the fields $\phi_i(t, x^a)$ can be expanded in Fourier modes. These are a collection of degrees of freedom that are just time dependent. The string field theory action then reduces to an integral over time of a time-dependent Lagrangian density.

We will set up the string field theory in such a way that all dimensions (including time) are compactified on circles of unit circumference. In this case, the mass M of the Dp-brane coincides with the tension of the Dp-

brane. The claim is that

$$M = \frac{1}{2\pi^2 g^2} . \tag{207}$$

In this formula and in the following, we set $\alpha' = 1$. In these units the string tension is $T_0 = 1/(2\pi)$. When we consider the string field theory of a D25-brane, (207) gives

$$T_{25} = \frac{1}{2\pi^2 g^2} . \tag{208}$$

We begin our study by considering some special momentum states of the BCFT:

$$|k_0\rangle \equiv e^{(ik_0 X^0(0))}|0\rangle . \tag{209}$$

Moreover, we will normalize these states by declaring

$$\langle k_0|\, c_{-1}c_0c_1\, |k_0'\rangle = \delta_{k_0,k_0'} , \tag{210}$$

consistent with the discussion below (98). Since the time direction has been made compact via $t \sim t+1$, the time component k_0 of the momentum is quantized: $k_0 = 2\pi n$, with n integer, and we can use a Kronecker delta in the above inner product.

We will consider the computation of the brane mass in three steps.

Step 1: We consider time-dependent displacements of the D-brane. We will write down a string field that describes such a displacement and evaluate the kinetic term of the string action. This will make it clear how we can hope to calculate the brane mass.

Let X^i be one of the Dirichlet directions for the D-brane and assume that $x^i = 0$ is the original position of the brane. Consider now a displacement field $\phi^i(t)$ that is expected to be proportional to a coordinate displacement $x^i(t)$. We expand the field $\phi^i(t)$ as:

$$\phi^i(t) = \sum_{k_0} e^{ik_0 t}\, \phi^i(k_0) , \tag{211}$$

and we use the Fourier components $\phi^i(k_0)$ to assemble the corresponding string field:

$$|\Phi\rangle = \sum_{k_0} \phi^i(k_0)\, c_1 \alpha^i_{-1}\, |k_0\rangle . \tag{212}$$

As you can see, the string field is built using states of the massless scalar field that represents translations of the D-brane. For this string field, the kinetic term $S_2(\Phi)$ of the string action is given by

$$S_2(\Phi) = -\frac{1}{g^2} \sum_{k_0, k_0'} \phi^i(k_0)\phi^i(k_0')\langle -k_0'|c_{-1}\alpha_1^i\, c_0 L_0\, c_1 \alpha_{-1}^i|k_0\rangle\,. \tag{213}$$

Since $L_0 = p^2 + \ldots$ where the terms indicated by dots vanish in the present case, $L_0 = -k_0^2$ in (213) and

$$S_2(\Phi) = \frac{1}{2g^2} \sum_{k_0} \phi^i(-k_0)\, k_0^2 \phi^i(k_0)\,. \tag{214}$$

Let us now rewrite this string action in terms of the field $\phi^i(t)$ introduced in (211). A short computation gives

$$\int_0^1 dt\, \partial_t \phi^i\, \partial_t \phi^i = \sum_{k_0} \phi^i(-k_0)\, k_0^2 \phi^i(k_0)\,. \tag{215}$$

Comparing with (214) we find that

$$S_2(\Phi) = \frac{1}{2g^2} \int_0^1 dt\, \partial_t \phi^i\, \partial_t \phi^i\,. \tag{216}$$

As we mentioned earlier, the field $\phi^i(t)$ is expected to be proportional to the position $x^i(t)$ of the brane (at least for small, slowly varying displacements), so we can rewrite the above action as

$$S_2(\Phi) = \frac{1}{g^2}\left(\frac{\delta\phi^i}{\delta x^i}\right)^2 \int_0^1 dt\, \frac{1}{2}\partial_t x^i\, \partial_t x^i \tag{217}$$

where the derivatives are evaluated at zero displacement. Since $\partial_t x^i$ is the velocity of the D-brane, the above action represents the contribution from the (non-relativistic) kinetic energy of a D-brane that has a mass M given by [c]

$$M = \frac{1}{g^2}\left(\frac{\delta\phi^i}{\delta x^i}\right)^2. \tag{218}$$

---

[c]Note that at this point, it is possible to take a shortcut to get the D-brane mass directly using the fact that SFT at tree level reproduces Yang-Mills theory,[82] with[43,82] $g_{\mathrm{YM}} = g/\sqrt{2}$, where the Yang-Mills field appears in the string field expansion as $A_\mu(k)\alpha_{-1}^\mu|0_1;k\rangle$, and where an additional factor of 2π arises from the T-duality relation from section 2.4, $X \to 2\pi\alpha' A$. Thus, replacing $A^i \to X^i/2\pi$ in the Yang-Mills action we have $1/2g_{\mathrm{YM}}^2 F_{0i}F^{0i} \to 1/2g^2(\partial_0 x^i/\sqrt{2}\pi)^2$, so $M = 1/2\pi^2 g^2$. Although perhaps more transparent, this is essentially the same calculation as the original argument of Sen presented here, which we include in full as it sheds light on the structure of the theory.

Step 2. To find out how $\delta\phi^i$ is related to a true displacement δx^i, we add a reference D-brane a distance b away from our original brane, in the direction x^i. We will then consider a string stretched between the branes. We will use the string field action to compute the change in the mass of such string when our D-brane is displaced by some $\delta\phi^i$. Since the string tension is known, we will be able to calculate the value of the physical displacement δx^i.

Given a string of length L, its mass includes a contribution $T_0 L = L/(2\pi)$, and the corresponding contribution to the mass-squared is $L^2/(4\pi^2)$. If the original stretched string has length b and its length is then changed to $b + \delta x^i$, the change δm^2 of the mass-squared is

$$\delta m^2 = \frac{1}{4\pi^2}\Big((b+\delta x^i)^2 - b^2\Big) \simeq \frac{1}{2\pi^2}\, b\, \delta x^i\,. \tag{219}$$

Let us now consider a time independent displacement, that is, a configuration with $k_0 = 0$ (see (211)). We thus set $\delta\phi^i \equiv \delta\phi^i(k_0 = 0)$ and $\delta\phi^i(k_0 \neq 0) = 0$. The string field associated to this displacement is obtained using (212):

$$|\delta\Phi\rangle = \delta\phi^i\, c_1 \alpha^i_{-1}\, |0\rangle\,. \tag{220}$$

We want to learn the effect of this string field perturbation on the masses of stretched strings. To do so, we introduce a complex field η. The fields η and η^* represent the string that stretches from our brane (brane one) to the reference brane (brane two) and the string that stretches from the reference brane to our brane, respectively. The string field that describes these states is written as:

$$|\psi\rangle = \eta\, c_1 |k_0, b\rangle \otimes \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} + \eta^*\, c_1 | -k_0, -b\rangle \otimes \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}\,. \tag{221}$$

The matrices included here are Chan-Paton matrices $a_{\alpha\beta}$, with $\alpha, \beta = 1, 2$. A value of one for a given $a_{\alpha\beta}$, with zero for all other entries, is used to represent a string that stretches from brane α to brane β. When we have parallel D-branes, the string field is matrix-valued. The string action includes a trace operation Tr that applies to the matrices, and the star product includes matrix multiplication. The state $|k_0, b\rangle$ represents the ground state of a string with momentum k_0 that stretches a distance b in the x^i direction. It is necessary for our analysis to determine the CFT vertex operator that corresponds to this stretched string. We claim that the operator is

$$|k_0, b\rangle \quad \longleftrightarrow \quad e^{ik_0 X^0}\, e^{i\frac{b}{2\pi}(X^i_L - X^i_R)}\,. \tag{222}$$

The k_0 dependence of the operator is already familiar from (209). The field multiplying b is formed from the left-moving and right-moving parts of the open string coordinate X^i, evaluated at the string endpoint. This coordinate X^i satisfies Dirichlet boundary conditions, so at the boundary $X_L^i = -X_R^i$, and we can replace $X_L^i - X_R^i$ by $2X_L^i$. We also have the operator products and stress tensor

$$\partial X_L^i(x)\,\partial X_L^i(y) \sim -\frac{1}{2}\frac{1}{(x-y)^2}\,, \qquad T_{X^i} = -\partial X_L^i \partial X_L^i\,. \tag{223}$$

These relations allow us to compute the conformal dimension of an exponential. One readily finds that $\exp(i\alpha X_L^i)$ has conformal dimension $\alpha^2/4$. It follows that

$$\text{dimension}\,\left(e^{i\frac{b}{\pi}X_L^i}\right) = \left(\frac{b}{2\pi}\right)^2. \tag{224}$$

Since conformal dimension is the value of L_0, which, in turn, determines the mass-squared, this result confirms that the operator in (222) has correctly reproduced the mass-squared of the stretched string. For future use, the vertex operator can be written as

$$e^{ik_0X^0}\,e^{i\frac{b}{\pi}X_L^i}\,. \tag{225}$$

The evaluation of the kinetic term for the field in (221) is relatively straightforward. The only terms that survive are the off-diagonal ones, coupling η and η^*. There are two such terms, and their contributions are identical. The product of the two matrices give a matrix of trace one, and the overlap $\langle -k_0, -b|k_0, b\rangle$ is also equal to one. We then find

$$g^2\,S_2(\eta,\eta^*) = -\frac{1}{2}\cdot 2\,\eta^*\Big(-k_0^2 + \frac{b^2}{(2\pi)^2}\Big)\eta = \eta^*\Big(k_0^2 - \frac{b^2}{(2\pi)^2}\Big)\eta\,. \tag{226}$$

In the setup with two branes, the fluctuation (220) that represents the displacement of our brane is fully represented by

$$|\delta\Phi\rangle = \delta\phi^i\,c_1\alpha_{-1}^i\,|0\rangle \otimes \begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}\,. \tag{227}$$

With the chosen normalization for X^i, the vertex operator associated with $\alpha_{-1}^i|0\rangle$ is $i\sqrt{2}\partial X_L$.

Step 3. We must now include the effects of the interactions to see how the fluctuation (227) affects the mass of the stretched string. Since the mass can be read from equation (226), we will find a term proportional to $\eta^*\eta$ that arises from the interaction and modifies the value of the mass.

The interaction term takes the form

$$g^2 S_3(\Phi) = -\frac{1}{3}\langle\Phi,\Phi,\Phi\rangle\,, \tag{228}$$

and the string field is taken to be $|\Phi\rangle = |\psi\rangle + |\delta\Phi\rangle$, in order to see the effect of the fluctuation on the stretched string. We are looking for the terms of the form $\eta^*\eta\,\delta\phi^i$, so we have three different operators to insert at three different punctures. There are a total of six possible arrangements, that can be divided into two groups of three arrangements each. In each of these groups the cyclic ordering of the operators is the same. The Chan-Paton matrices imply that one cyclic ordering contributes while the other does not. Indeed,

$$\begin{pmatrix}0&1\\0&0\end{pmatrix}\begin{pmatrix}0&0\\1&0\end{pmatrix}\begin{pmatrix}1&0\\0&0\end{pmatrix}=\begin{pmatrix}1&0\\0&0\end{pmatrix} \tag{229}$$

is a matrix of unit trace, while

$$\begin{pmatrix}0&0\\1&0\end{pmatrix}\begin{pmatrix}0&1\\0&0\end{pmatrix}\begin{pmatrix}1&0\\0&0\end{pmatrix}=\begin{pmatrix}0&0\\0&0\end{pmatrix} \tag{230}$$

is a matrix of vanishing trace. We conclude that the Chan-Paton matrices contribute a factor of $+3$. The operators to be inserted can be chosen to be physical (dimension zero) so we need not worry about local coordinates at the punctures. Using punctures at ∞, -1, and 0 we then find:

$$g^2 S_3(\Phi) = -\eta^*\eta\delta\phi^i\Big\langle e^{-ik_0X^0 - i\frac{b}{\pi}X^i_L}c(\infty)\,\sqrt{2}\,i\,\partial X_L c(-1)\,e^{ik_0X^0+i\frac{b}{\pi}X^i_L}c(0)\Big\rangle\,. \tag{231}$$

Since the vertex operators are on-shell, and the ghost insertions are placed at standard positions, the whole correlator gives a factor of one, except for the contraction between the $\partial X_L(-1)$ and the finitely located $\exp(i\frac{b}{\pi}X^i_L(0))$:

$$g^2 S_3 = -\eta^*\eta\delta\phi^i\,\sqrt{2}\,i\,\frac{ib}{\pi}\Big(-\frac{1}{2}\Big)\frac{1}{(-1-0)} = \frac{b}{\sqrt{2}\,\pi}\eta^*\eta\delta\phi^i\,. \tag{232}$$

Combining this result with (226) we find

$$g^2\,(S_2+S_3) = \eta^*\Big(k_0^2 - \frac{b^2}{(2\pi)^2} + \frac{b}{\sqrt{2}}\frac{\delta\phi^i}{\pi}\Big)\eta\,. \tag{233}$$

The last term in parenthesis corresponds to a change in m^2. So, comparing with (219) we obtain

$$\frac{1}{2\pi^2}b\,\delta x^i = -\frac{b}{\sqrt{2}}\frac{\delta\phi^i}{\pi} \quad\rightarrow\quad \frac{\delta\phi^i}{\delta x^i} = -\frac{1}{\sqrt{2}\,\pi}\,. \tag{234}$$

This is the needed relation between the field $\delta\phi^i$ that represents a displacement of the brane and the resulting displacement δx^i. The mass of the brane now follows directly from (218):

$$M = \frac{1}{g^2}\Big(\frac{1}{\sqrt{2}\,\pi}\Big)^2 = \frac{1}{2\pi^2 g^2}\,. \tag{235}$$

This is the result we wanted to establish.

7.2. *Constraints and symmetries*

It may appear that *a priori* all scalar fields in the spectrum of open strings could acquire a vacuum expectation value in the tachyonic vacuum. Nevertheless, there are a set of considerations that imply that only a subset of these scalar fields acquire expectation values. In this section we explore these ideas. They are subdivided into the following:

(1) Universality conditions.
(2) Twist conditions.
(3) Gauge fixing conditions.
(4) $SU(1,1)$ conditions.

Among these conditions, the third one, which concerns gauge fixing, is on a somewhat different footing. The other three conditions apply because of a simple general argument which we discuss first.

Consider a subdivision of all the scalar fields into two disjoint set of fields. The first set contains the fields $t_0, t_1, t_2, \ldots$ and the second set contains the fields $u_0, u_1, u_2, \ldots$. Let us denote by t_i the elements of the first set and by u_a the elements of the second set. Suppose the string field action $S(t_i, u_a)$ is such that there are no terms that are linear in u_a. We then claim that it is consistent to search for a solution of the equations of motion that assumes $u_a = 0$ for all a. The reason is easy to explain. If all terms with u fields contain at least two of them, the equations of motion for the u fields are composed of terms all of which contain at least one u field. As a result, $u_a = 0$ satisfies these equations of motion. In our analysis we will try to construct a set $\{t_i\}$ with the smallest possible number of fields, so that none of the remaining fields couples linearly in the action. The tachyon field, of course, is one of the elements of the set $\{t_i\}$.

Let's begin by explaining how (1) works. For this we split the state space of the BCFT into three groups. In each of these groups, the ghost part of the states is the same: it includes all states of ghost number one.

The nontrivial part of the argument uses the matter part of the conformal field theory. We write

$$\mathcal{H} = \mathcal{H}_1 \oplus \mathcal{H}_2 \oplus \mathcal{H}_3 \,, \tag{236}$$

where $\mathcal{H}_1, \mathcal{H}_2$, and $\mathcal{H}_3$ will be disjoint vector subspaces of $\mathcal{H}$ (their intersections give the zero vector). We also write

$$\mathcal{H}_i = \mathcal{M}_i \otimes |\mathcal{G}\rangle \,, \quad i = 1, 2, 3 \,. \tag{237}$$

where $|\mathcal{G}\rangle$ denotes the general state of ghost number one in the ghost conformal field theory. The spaces $\mathcal{M}_1, \mathcal{M}_2$, and $\mathcal{M}_3$ are disjoint subspaces of the matter CFT whose union gives the total matter CFT state space. The $\mathcal{M}$ subspaces are defined as:

$$\begin{aligned}
&\mathcal{M}_1: \text{ primary } \;\; |0\rangle \text{ and descendents,}\\
&\mathcal{M}_2: \text{ primaries } |k_0 \neq 0\rangle \text{ and descendents,}\\
&\mathcal{M}_3: \text{ primaries } |k_0 = 0\rangle \text{ different from } |0\rangle \text{ and descendents.}
\end{aligned} \tag{238}$$

In the above, primary means Virasoro primary in the matter sector, and descendent means Virasoro descendent. The union of the spaces give the full CFT because for unitary matter CFT's (which we are assuming our CFT is) the state space can be broken into primaries and their descendents. Any matter primary belongs to one of the three spaces above. It should also be clear that the primaries in $\mathcal{M}_3$ have positive conformal dimension.

We now claim that the fields in $\mathcal{H}_2$ and $\mathcal{H}_3$ need not acquire expectation values (they are u fields); the tachyon condensate is all in $\mathcal{H}_1$. We are therefore defining the t fields to be precisely the fields in $\mathcal{H}_1$. To prove that this is valid, we first note that a field in $\mathcal{H}_2$ cannot appear linearly in a term where all other fields are t fields (*i.e.*, fields in $\mathcal{H}_1$). The reason is simply momentum conservation.

A little more work is necessary to show that the fields in $\mathcal{H}_3$ cannot couple linearly to the fields in $\mathcal{H}_1$. Let us begin with the kinetic term. Since the BRST operator is composed of terms that include ghost oscillators and matter Virasoro operators, it maps each $\mathcal{H}_i$ space into itself. The primaries in $\mathcal{H}_1$ and $\mathcal{H}_3$ are BPZ orthogonal, so any two states in the descendent towers are also orthogonal (this is proven by using the BPZ conjugation properties of Virasoro operators to move them from one state to the other until some state is annihilated or the whole expression reduces to the BPZ inner product of the primaries). For the interaction term a similar argument holds. First note that the three string vertex does not couple two matter primaries from $\mathcal{H}_1$ to a matter primary from $\mathcal{H}_3$. This is because in the

CFT matter correlator the primaries from $\mathcal{H}_1$ appear as identity operators, so the whole correlator is proportional to the one-point function of the primary in $\mathcal{H}_3$, which vanishes because the state has non-zero dimension. The Virasoro conservation laws on the vertex then imply that the coupling of any two states in $\mathcal{H}_1$ to a state in $\mathcal{H}_3$ must vanish. This completes our proof.

The space $\mathcal{H}_1$ is universal. It does not depend on the details of the matter conformal field theory, except for the existence of a zero-momentum SL(2,R) ground state. The space can be written as

$$\mathcal{H}_1 \equiv \text{Span}\Big\{ L^m_{-j_1} \ldots L^m_{-j_p}\, b_{-k_1} \ldots b_{-k_q}\, c_{-l_1} \ldots c_{-l_r}\, |0\rangle \Big\} \tag{239}$$

where

$$j_1 \geq j_2 \geq \ldots \geq j_p\,, \quad j_i \geq 2\,, \quad k_i \geq 2\,, \quad l_i \geq -1\,, \quad \text{and} \quad l - q = 1\,. \tag{240}$$

The first inequality ensures that the descendents are built unambiguously, the second inequality is needed because $L_{-1}|0\rangle = 0$. The third and fourth inequalities are familiar, and the last equality ensures that the ghost number of the state is one.

Let us now explain how twist properties allow us to restrict $\mathcal{H}_1$ further. The claim is that we can restrict ourselves to the twist even subspace of $\mathcal{H}_1$. Heuristically, this follows from the fact that the two- and three-string vertices are invariant under reflection, so all terms linear in twist fields would pick up a change of sign and therefore vanish. The twist-even space, of course, contains the zero momentum tachyon $c_1|0\rangle$ (recall that $|0\rangle$ is twist odd, and $\Omega c_{-n}\Omega^{-1} = (-)^n c_{-n}$). The first two properties in (75) ensure that the kinetic term in the string action does not couple a twist odd field to a twist even field. We also studied the twist properties of the three string vertex. In fact, in an exercise, we considered a twist even field A_+ and a twist odd field A_- ($\Omega A_\pm = \pm A$) both of which were Grassmann odd (like the string field is). You then showed that $\langle A_+, A_+, A_-\rangle = 0$ (see (79)). Consider now a general string field $\Phi \in \mathcal{H}$ and split it into twist even and twist odd parts $\Phi = \Phi_+ + \Phi_-$. When the interaction vertex is evaluated, the terms linear in Φ_- are of the form $\langle \Phi_+, \Phi_+, \Phi_-\rangle$ (any other similar looking term is related to this by cyclicity). So terms linear on twist odd fields vanish. This proves that we can indeed constrain $\mathcal{H}_1$ further.

The twist eigenvalue of a state is given as $\Omega = (-1)^N$, where N is the number eigenvalue of the state, defined with $N = 0$ for the zero momentum tachyon. In terms of level, states at odd levels are twist odd, and states at even levels are twist even. So, the twist condition allows us to restrict ourselves to the states of $\mathcal{H}_1$ at even levels.

We now turn to the gauge fixing condition. This gauge fixing condition, the Feynman-Siegel gauge condition $b_0|\Phi\rangle = 0$, restricts further the space $\mathcal{H}_1$. We will discuss the global validity of the Siegel gauge later, but here we discuss its clear validity at the linearized level and within the subspace $\mathcal{H}_1$ already restricted to states at even level. First, we show that the gauge condition can be reached starting from fields that do not satisfy it. Let $|\Phi\rangle$ be a field such that $b_0|\Phi\rangle \neq 0$. Since $|\Phi\rangle$ cannot be of level one, $L_0|\Phi\rangle \neq 0$. Then consider the following gauge equivalent state

$$|\widetilde{\Phi}\rangle = |\Phi\rangle - Q\frac{b_0}{L_0}|\Phi\rangle\,. \tag{241}$$

Using $\{b_0, Q\} = L_0$ one readily checks that $b_0|\widetilde{\Phi}\rangle = 0$, so the gauge can be reached. Moreover, we now show that no gauge transformation remains in this gauge. If there were, there would exist a non-zero string field in the gauge slice that happens to be pure gauge. Such field $|\Phi\rangle$ would then satisfy $b_0|\Phi\rangle = 0$, $L_0|\Phi\rangle \neq 0$, and $|\Phi\rangle = Q|\epsilon\rangle$. Since both b_0 and Q annihilate the state:

$$0 = b_0Q|\Phi\rangle + Qb_0|\Phi\rangle = \{b_0, Q\}|\Phi\rangle = L_0|\Phi\rangle\,, \tag{242}$$

in contradiction with the fact that the state does have non-zero dimension. The Siegel gauge is clearly a good gauge at the linearized level and within the twist truncated $\mathcal{H}_1$.

Let's now consider briefly the additional truncation that is allowed by $SU(1,1)$ symmetry (item (4) of our list). Once we work in the Siegel gauge, this further truncation is allowed. This truncation is only possible because of the particular form of the string vertex. It would not be allowed for arbitrarily defined star products. Let us recall how this symmetry arises in the cubic open string field theory.[83] In the Siegel gauge, the string field action reads

$$S \sim \frac{1}{2}\langle\phi|L_0|\phi\rangle + \frac{1}{3}\langle\phi|\langle\phi|\langle\phi\,|\,v_3\rangle\,. \tag{243}$$

The vertex coupling the three string fields is of the form

$$|v_3\rangle \sim \exp\left(E_{matt}\right)\, \exp\Big(-\sum_{r,s=1}^{3}\sum_{n,m=1}^{\infty} c^r_{-n}\, X^{rs}_{nm}\, b^s_{-m}\Big)\, |0_1\rangle_{123}\,, \tag{244}$$

where we have focused on the ghost sector. The Neumann coefficients are known to satisfy[83]

$$X^{rs}_{nm} = \frac{n}{m} X^{sr}_{mn}\,, \quad n, m \geq 1\,. \tag{245}$$

This relation is not true for general three string vertices, but holds for the open string field theory vertex. Given equation (245), the argument of the exponential in $|v_3\rangle_{123}$ can be written as a sum of terms of the form (r, s, n, m, not summed)

$$X^{rs}_{nm}\, c^r_{-n} b^s_{-m} + X^{sr}_{mn}\, c^s_{-m} b^r_{-n} = \frac{1}{m} X^{sr}_{mn} \Big(n c^r_{-n} b^s_{-m} + m c^s_{-m} b^r_{-n} \Big)\,. \tag{246}$$

The term in parenthesis is invariant under the continuous transformations

$$\begin{aligned} b_{-n}(\theta) &= b_{-n} \cos\theta - n c_{-n} \sin\theta\,, \\ c_{-n}(\theta) &= c_{-n} \cos\theta + \frac{1}{n} b_{-n} \sin\theta\,. \end{aligned} \tag{247}$$

These transformations, valid for all $n \neq 0$, imply $\{c_n(\theta), b_m(\theta)\} = \delta_{n+m}$. One readily finds that they are generated by an operator $\mathcal{S}_1$:

$$b_{-n}(\theta) = e^{\theta \mathcal{S}_1}\, b_{-n} e^{-\theta \mathcal{S}_1}\,, \quad c_{-n}(\theta) = e^{\theta \mathcal{S}_1}\, c_{-n} e^{-\theta \mathcal{S}_1}\,, \tag{248}$$

where $\mathcal{S}_1$ is given by

$$\mathcal{S}_1 = \sum_{n=1}^{\infty} \Big(\frac{1}{n}\, b_{-n} b_n - n c_{-n} c_n \Big)\,. \tag{249}$$

Since the vacuum $|0_1\rangle$ is annihilated by $\mathcal{S}_1$, the vertex $|v_3\rangle$ is invariant under this $U(1)$ symmetry: $\exp\Big(\theta(\mathcal{S}_1^{(1)} + \mathcal{S}_1^{(2)} + \mathcal{S}_1^{(3)})\Big)|v_3\rangle_{123} = |v_3\rangle_{123}$. Equivalently,

$$\Big(\mathcal{S}_1^{(1)} + \mathcal{S}_1^{(2)} + \mathcal{S}_1^{(3)}\Big)|v_3\rangle_{123} = 0\,. \tag{250}$$

Since the vertex $|v_3\rangle_{123}$ is built from ghost bilinears of zero ghost number, we deduce that the ghost number operator $\mathcal{G}$

$$\mathcal{G} = \sum_{n=1}^{\infty} \Big(c_{-n} b_n - b_{-n} c_n \Big). \tag{251}$$

is also conserved:

$$\Big(\mathcal{G}^{(1)} + \mathcal{G}^{(2)} + \mathcal{G}^{(3)}\Big)|v_3\rangle_{123} = 0\,. \tag{252}$$

We can then form the commutator

$$[\mathcal{S}_1\,, \mathcal{G}\,] = 2\mathcal{S}_2\,, \quad \text{with} \quad \mathcal{S}_2 = \sum_{n=1}^{\infty} \Big(\frac{1}{n}\, b_{-n} b_n + n c_{-n} c_n \Big)\,. \tag{253}$$

The remaining commutators are readily computed:

$$[\mathcal{S}_2\,,\,\mathcal{G}\,] = 2\mathcal{S}_1\,, \quad [\mathcal{S}_1\,,\,\mathcal{S}_2\,] = -2\mathcal{G}\,. \tag{254}$$

These relations show that $\{\mathcal{S}_1, \mathcal{S}_2, \mathcal{G}\}$ generate the algebra of $SU(1,1)$. These generators are the same as those in the $SU(1,1)$ algebra in Siegel and Zwiebach.[24,d] Since both $\mathcal{S}_1$ and $\mathcal{G}$ are symmetries of the three string vertex, we also have

$$\left(\mathcal{S}_2^{(1)} + \mathcal{S}_2^{(2)} + \mathcal{S}_2^{(3)}\right)|v_3\rangle_{123} = 0\,. \tag{255}$$

In summary, the three string vertex is fully $SU(1,1)$ invariant.

The set of Fock space states built with the action of ghost and antighost oscillators on the vacuum $|0_1\rangle$ can be decomposed into finite dimensional irreducible representations of $SU(1,1)$. Note that (nc_{-n}, b_{-n}) transforms as a doublet. As usual, from the tensor product of two doublets one can obtain a nontrivial singlet; this is just

$$mb_{-n}c_{-m} + nb_{-m}c_{-n}\,. \tag{256}$$

It is now simple to argue that the twist even subspace of $\mathcal{H}_1$ in the Siegel gauge can be further restricted to $SU(1,1)$ singlets. Since the kinetic operator L_0 commutes with the $SU(1,1)$ generators, the kinetic term cannot couple a non-singlet to a singlet. Indeed, consider such a term $\langle s|L_0|a\rangle$, where $\langle s|$ is a singlet and $|a\rangle$ is not a singlet. Given the structure of the representations (completely analogous to the finite dimensional unitary representations of $SU(2)$), it follows that there is a state $|b\rangle$ and an $SU(1,1)$ generator $\mathcal{J}$ such that $|a\rangle = \mathcal{J}|b\rangle$. Therefore $\langle s|L_0|a\rangle = \langle s|L_0\mathcal{J}|b\rangle = \langle s|\mathcal{J}L_0|b\rangle = 0$, where the last step gives zero because $\mathcal{J}$ annihilates the singlet (this requires $bpz(\mathcal{J}) = \pm\mathcal{J}$, which is true). It remains to show that the vertex cannot couple a non-singlet to two singlets. Indeed, with analogous notation we have

$$\begin{aligned} {}_1\langle s_1|_2\langle s_2|_3\langle a\,|v_3\rangle &= {}_1\langle s_1|_2\langle s_2|_3\langle b\,|\mathcal{J}^{(3)}|v_3\rangle \\ &= -{}_1\langle s_1|_2\langle s_2|_3\langle b\,|(\mathcal{J}^{(1)} + \mathcal{J}^{(2)})|\,v_3\rangle = 0\,, \end{aligned} \tag{257}$$

where we used the conservation of $\mathcal{J}$ on the vertex, and on the last step the $\mathcal{J}$ operators annihilate the singlets.

[d] Defining $X = (\mathcal{S}_2 - \mathcal{S}_1)/2$, $Y = (\mathcal{S}_2 + \mathcal{S}_1)/2$, and $H = \mathcal{G}$ we recover the conventional definition of the isomorphic (real) Lie algebra $sl(2,R)$, with brackets $[X,Y] = H$, $[H,X] = 2X$, $[H,Y] = -2Y$. Note that $T_+ = -2X$, where T_+ is the operator that multiplies b_0 in the BRST operator.

This completes our discussion of the various symmetries and conditions that can be used to constrain the subspace of the string state space that acquires vacuum expectation values in the tachyon vacuum.

7.3. *The nonperturbative vacuum*

Sen's first conjecture states that the string field theory action should lead to a nontrivial vacuum solution, with energy density

$$-T_{25} = -\frac{1}{2\pi^2 g^2}\,. \tag{258}$$

In this subsection we discuss evidence for the validity of this conjecture in Witten's OSFT. As mentioned in the introduction, this result holds exactly in BSFT.

The string field theory equation of motion is

$$Q\Psi + g\Psi \star \Psi = 0\,. \tag{259}$$

Despite much work over the last few years, there is still no analytic solution of this equation of motion[e]. There is, however, a systematic approximation scheme, known as level truncation, which can be used to solve this equation numerically.[57] The level (L, I) truncation of the full string field theory involves dropping all fields at level $N > L$, and disregarding any interaction term between three fields whose levels add up to a number that is greater than I. For example, the simplest truncation of the theory is the level (0, 0) truncation. This is the truncation which was used in section 5.3. Including only the zero-momentum component of the tachyon field, since we are looking for a Lorentz-invariant vacuum, the truncated theory is simply described by a potential for the tachyon zero-mode

$$V(\phi) = -\frac{1}{2}\phi^2 + g\bar{\kappa}\phi^3\,. \tag{260}$$

where $\bar{\kappa} = \kappa/3 = 3^{7/2}/2^6$. This cubic function, which was computed in (123) using the CFT approach, and in (202) using the oscillator approach, is graphed in Figure 7. As discussed in section 5.3, this potential has a local minimum at

$$\phi_0 = \frac{1}{3g\bar{\kappa}}\,, \tag{261}$$

[e] as of October, 2003

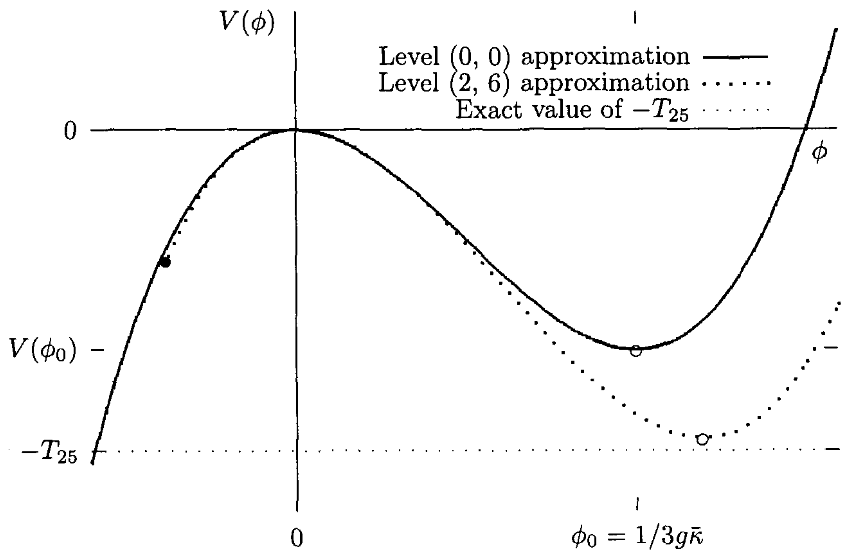


Figure 7. The effective tachyon potential in level (0, 0) and (2, 6) truncations. The open circles denote minima in each level truncation. The filled circle denotes a branch point where the level (2, 6) truncation approximation reaches the limit of Feynman-Siegel gauge validity.

and at this point the potential is

$$V(\phi_0) = -\frac{1}{54}\frac{1}{g^2\bar{\kappa}^2} = -\frac{2^{11}}{3^{10}}\frac{1}{g^2} \approx (0.68)\left(-\frac{1}{2\pi^2 g^2}\right). \tag{262}$$

Thus, simply including the tachyon zero-mode gives a nontrivial vacuum with 68% of the vacuum energy density predicted by Sen. This vacuum is denoted by an open circle in Figure 7.

At higher levels of truncation, there are a multitude of fields with various tensor structures. However, again assuming that we are looking for a vacuum which preserves Lorentz symmetry, we can restrict attention to the interactions between scalar fields at zero momentum. We will work in Feynman-Siegel gauge to simplify calculations; as shown in the previous subsection, this gauge is good at least in a local neighborhood of the point where all fields vanish. The situation is further simplified by the existence of the twist symmetry, which as mentioned in the previous subsection guarantees that no cubic vertex between (zero-momentum) scalar fields can connect three fields with a total level which is odd, and thus means that odd fields are not relevant to diagrams with only external tachyons at tree level. Therefore, we need only consider even-level scalar fields in looking for Lorentz-preserving solutions to the SFT equations of motion. With these simplifications, in a general level truncation the string field is simply expressed as a sum of a finite number of terms

$$\Psi_s = \sum_i \phi_i |s_i\rangle \tag{263}$$

where ϕ_i are the zero-modes of the scalar fields associated with even-level states $|s_i\rangle$. As discussed in the previous subsection, this set of scalar fields can be further restricted to be SU(1, 1) singlets in the universal subspace $\mathcal{H}_1$. For example, including fields up to level 2, we have

$$\Psi_s = \phi|0_1\rangle + B\,(\alpha_{-1}\cdot\alpha_{-1})|0_1\rangle + \beta\, b_{-1}c_{-1}|0_1\rangle\,. \tag{264}$$

In terms of the matter Virasoro generators, the state associated with the field B is

$$(\alpha_{-1}\cdot\alpha_{-1})|0_1\rangle = 2L_{-2}|0_1\rangle\,, \tag{265}$$

which lies in the universal subspace $\mathcal{H}_1$. The potential for all the scalars appearing in the level-truncated expansion (263) can be simply expressed as a cubic polynomial in the zero-modes of the scalar fields

$$V = \sum_{i,j} d_{ij}\phi_i\phi_j + g\bar{\kappa}\sum_{i,j,k} t_{ijk}\phi_i\phi_j\phi_k\,. \tag{266}$$

Using the expressions for the Neumann coefficients given in Section 5.3, the potential for all the scalar fields up to level L can be computed in a level (L, I) truncation. For example, the potential in the level $(2, 6)$ truncation is given by

$$\begin{aligned} V = {} & -\frac{1}{2}\phi^2 + 26B^2 - \frac{1}{2}\beta^2 \\ & +\bar{\kappa}g\left[\phi^3 - \frac{130}{9}\phi^2 B - \frac{11}{9}\phi^2\beta + \frac{30212}{243}\phi B^2 + \frac{2860}{243}\phi B\beta + \frac{19}{81}\phi\beta^2\right. \\ & \left. - \frac{2178904}{6561}B^3 - \frac{332332}{6561}B^2\beta - \frac{2470}{2187}B\beta^2 - \frac{1}{81}\beta^3\right] \end{aligned} \tag{267}$$

As an example of how these terms arise, consider the $\phi^2 B$ term. The coefficient in this term is given by

$$g\,\langle V_3|\left(|0_1\rangle\otimes|0_1\rangle\otimes\alpha_{-1}\cdot\alpha_{-1}|0_1\rangle\right) = -g\bar{\kappa}\,(3\cdot 26)\,V_{11}^{11} = -g\bar{\kappa}\frac{130}{9}\,, \tag{268}$$

where we have used $V_{11}^{11} = 5/27$.

In the level (2, 6) truncation of the theory, the nontrivial vacuum is found by simultaneously solving the three quadratic equations found by setting to zero the derivatives of the potential (267) with respect to ϕ, B, and β. There are a number of different solutions to these equations, but only one is in the vicinity of $\phi = 1/3g\bar{\kappa}$. The solution of interest is

$$\phi \approx 0.39766 \frac{1}{g\bar{\kappa}}, \quad B \approx 0.02045 \frac{1}{g\bar{\kappa}}, \quad \beta \approx -0.13897 \frac{1}{g\bar{\kappa}}. \tag{269}$$

Plugging these values into the potential gives

$$E_{(2,6)} = -0.95938\, T_{25}, \tag{270}$$

or 95.9% of the result predicted by Sen. This vacuum is denoted by an open circle in Figure 7.

It is a straightforward, computationally intensive project to generalize this calculation to higher levels of truncation. This calculation was carried out to level (4, 8) by Kostelecky and Samuel[57] many years ago. They noted that the vacuum seemed to be converging, but they lacked any physical picture to give meaning to this vacuum. Following Sen's conjectures, the level (4, 8) calculation was done again using somewhat different methods by Sen and Zwiebach,[38] who showed that the energy at this level is $-0.986\, T_{25}$. The calculation was automated by Moeller and Taylor,[84] who calculated up to level (10, 20), where there are 252 scalar fields, including all even-level scalar fields up to level 10; this computation was done using oscillators, without restriction to the universal subspace. Up to this level, the vacuum energy converges monotonically, as shown in Table 1. These numerical

level	$g\bar{\kappa}\langle\phi\rangle$	V/T_{25}
(0, 0)	0.3333	-0.68462
(2, 4)	0.3957	-0.94855
(2, 6)	0.3977	-0.95938
(4, 8)	0.4005	-0.98640
(4, 12)	0.4007	-0.98782
(6, 12)	0.4004	-0.99514
(6, 18)	0.4004	-0.99518
(8, 16)	0.3999	-0.99777
(8, 20)	0.3997	-0.99793
(10, 20)	0.3992	-0.99912

calculations indicate that level truncation of string field theory leads to a good systematic approximation scheme for computing the nonperturbative tachyon vacuum. It is also worth noting that in these computations, level $(L, 2L)$ and $(L, 3L)$ approximations give fairly similar values.

The preceding results were the best values for the vacuum energy at the time of these original lectures. More recently, Gaiotto and Rastelli reported further numerical results.[85,86] By programming in C++ instead of mathematica, and by computing using matter Virasoro operators rather than oscillators, so that only fields in the universal subspace $\mathcal{H}_1$ were included, they were able to extend the computation to level (18, 54). Their results are shown in Table 2 These results were rather surprising, indicating that while

level	V/T_{25}
(12, 24)	0.99979
(12, 36)	0.99982
(14, 28)	1.00016
(14, 42)	1.00017
(16, 32)	1.00037
(16, 48)	1.00038
(18, 36)	1.00049
(18, 54)	1.00049

the energy monotonically approaches $-T_{25}$ up to level 12, at level (14, 42) the energy drops below $-T_{25}$, and that the energy continues to decrease, reaching $-1.00049\,T_{25}$ at level (18, 54). We will discuss the resolution of this unexpected overshoot shortly.

First, however, it is interesting to consider the tachyon condensation problem from the point of view of the effective tachyon potential. If instead of trying to solve the quadratic equations for all N of the fields appearing in (266), we instead fix the tachyon field ϕ and solve the quadratic equations for the remaining $N-1$ fields, we can determine a effective potential $V(\phi)$ for the tachyon field. This has been done numerically up to level (16, 48).[84,86] At each level, the tachyon effective potential smoothly interpolates between the perturbative vacuum and the nonperturbative vacuum near $\phi = 0.4/g\bar{\kappa}$. For example, the tachyon effective potential at level (2, 6) is graphed in Figure 7. In all level truncations other than (0, 0) and (2, 4) (at least up to level (10, 20)), the tachyon effective potential has two branch point singularities at which the continuous solution for the other fields breaks down; for the level (2, 6) truncation, these branch points occur at $\phi \approx -0.127/g\bar{\kappa}$ and $\phi \approx 2.293/g\bar{\kappa}$; the lower branch point is denoted by a solid circle in Figure 7. As a result of these branch points, the tachyon effective potential is only valid for a finite range of ϕ, ranging between approximately $-0.1/g\bar{\kappa}$ and $0.6/g\bar{\kappa}$. In Section 7.4 we review results which indicate that these branch points arise because the trajectory in field space associated with this potential encounters the boundary of the region of Feynman-Siegel gauge validity. It seems almost to be a fortunate accident that the nonperturbative vacuum lies within the region of validity of this gauge choice. It is worth mentioning again here that in the BSFT approach, the tachyon potential can be computed exactly.[10] In this formulation, there is no branch point in the effective potential, which is unbounded below for negative values of the tachyon. On the other hand, the nontrivial vacuum in the background-independent approach arises only as the tachyon field goes to infinity, so it is harder to study the physics of the stable vacuum from this point of view.

Another interesting perspective on the tachyon effective potential is found by performing a perturbative (but off-shell) computation of the coefficients in the tachyon effective potential in the level-truncated theory. This gives a power series expansion of the effective potential

$$\begin{aligned} V(\phi) &= \sum_{n=2}^{\infty} c_n (\bar{\kappa} g)^{n-2} \phi^n \\ &= -\frac{1}{2}\phi^2 + (\bar{\kappa} g)\phi^3 + c_4 (\bar{\kappa} g)^2 \phi^4 + c_5 (\bar{\kappa} g)^3 \phi^5 + \cdots \end{aligned} \tag{271}$$

The coefficients up to c_{60} have been computed in the level truncations up to (10, 20).[84] Because of the branch point singularity near $\phi = -0.1/g\bar{\kappa}$, this series has a radius of convergence much smaller than the value of ϕ at the nonperturbative vacuum. Thus, the energy at the stable vacuum lies outside the naive range of the potential defined by the perturbative expansion.

Now, let us return to the problem of the overshoot in energy below $-T_{25}$ found at level 14 by Gaiotto and Rastelli[f]. The most straightforward way of determining whether or not this represents a real problem for string field

[f] The material in the remainder of this section was developed only after the original TASI lectures in 2001, but is included because of its relevance to the main development in this section.

theory would be to simply continue the calculation to higher levels. Unfortunately, at present this is not tractable, as the difficulty of computation grows exponentially in the level. Thus, we must resort to more indirect methods. It was found empirically by Taylor[87] that the level L approximations of string field theory give on-shell and off-shell amplitudes with error of order $1/L$. This work and further evidence[82,88] indicates that amplitudes can be very accurately approximated by computing them in different level L truncations, and matching to a power series in $1/L$. Such an approach can be taken to determine highly accurate values for the coefficients c_n in (271). As noted above, the resulting power series has a finite radius of convergence, and the stable vacuum lies beyond this limit. There is a standard technique, however, known as the method of Padé approximants, which allows one to extrapolate a function beyond its naive radius of convergence, if the function is sufficiently well-behaved in the direction in which it is extrapolated. The idea of Padé approximants is to replace a power series having given coefficients for a fixed number of terms with a rational function with the same number of coefficients, choosing the coefficients of the rational function to give a power series which agrees with the fixed coefficients in the original power series. For example, consider the first three terms in the tachyon effective potential (271),

$$-\frac{1}{2}\phi^2 + \kappa g\phi^3 - \frac{34}{27}(\kappa g)\phi^4 \,. \tag{272}$$

This truncated expansion has no local minima, while the Padé approximant

$$P_1^3(\phi) = \frac{-\frac{1}{2}\phi^2 + \frac{10}{27}\kappa g\phi^3}{1 + \frac{34}{27}\kappa g\phi} \tag{273}$$

does; this approximant thus represents a better description of the tachyon potential than the truncated expression (272). The advantage of Padé approximants is that they allow one to incorporate poles into approximations of a function with a desired local power series behavior. For a wide class of functions, successive Padé approximants converge exponentially quickly in the region where the function is smooth. Empirically, this seems to be the case for the tachyon effective potential. Thus, the energy minimum at any finite level of truncation can be determined to an arbitrary degree of accuracy from the leading coefficients in the potential. For example, the energy can be computed to 10 digits of accuracy by including approximately 40 coefficients c_n; this calculation is, however, highly sensitive to the accuracy of the coefficients.[89]

Combining Padé approximants with approximations to the coefficients c_n, computed by matching level-truncated results in a $1/L$ expansion, it is possible to predict not only the exact value of the energy at the stable minimum as $L \to \infty$, but also to predict the values of the approximate energy at intermediate values of L. Such a computation was performed using the level approximated values of c_n up to level $(10, 20)$.[90] By first using these values to predict the level-approximated values at higher levels, and then inserting these values into Padé approximants, the overshoot phenomenon found by Gaiotto and Rastelli was accurately reproduced. For example, compared to the value -1.0003678 found by these authors at level (16, 32), extrapolation from results at levels $(L, 2L)$ up to (10, 20) gives a predicted value of -1.0003773 at level (16, 32). Furthermore, the extrapolated values E_L of the energy at the stable minimum were found to decrease up to approximately level 26, and then to increase, approaching an asymptotic value as $L \to \infty$ of $E_\infty \approx -1$ with error $\sim 10^{-4}$. These results suggested that the energy at the minimum in the level-truncated theory takes the form shown in Figure 8.

In the calculation just described, there were two sources of error: 1) the coefficients c_n had some numerical inaccuracy, and 2) there is some error introduced in extrapolating from low levels of truncation.

This computation was improved by Gaiotto and Rastelli.[86] These authors used a different approach: instead of extrapolating the finite L results for the coefficients c_n, they extrapolated the nonperturbatively computed effective potential $V(\phi)$ at various values of ϕ. Because Padé approximants are so accurate, for exactly known values of c_n and $V(\phi)$ this approach is equivalent to the combined Padé-extrapolation in c_n approach, but generally this approach trades inaccuracy in c_n for inaccuracy in $V(\phi)$. In practice, it is much easier to compute the coefficients c_n exactly than the nonperturbative effective potential $V(\phi)$, which requires numerically solving a large system of quadratic equations. Gaiotto and Rastelli were able, however, to use their results on $V(\phi)$ at higher levels, which greatly increased the accuracy of their extrapolations. They found that while level $(L, 2L)$ and $(L, 3L)$ approximations tend to be very similar, extrapolations based on level $(L, 3L)$ truncations seem more robust. Using data up to level (16, 48) they found an extrapolated value of $E_\infty \approx -1.00003$, differing from -1 by an order of magnitude less than the value of the energy estimated at level 28, where the overshoot is predicted to be maximal. This gives compelling support to the conclusion that the level-truncated approximations to the energy indeed behave as shown in Figure 8, and approach the value

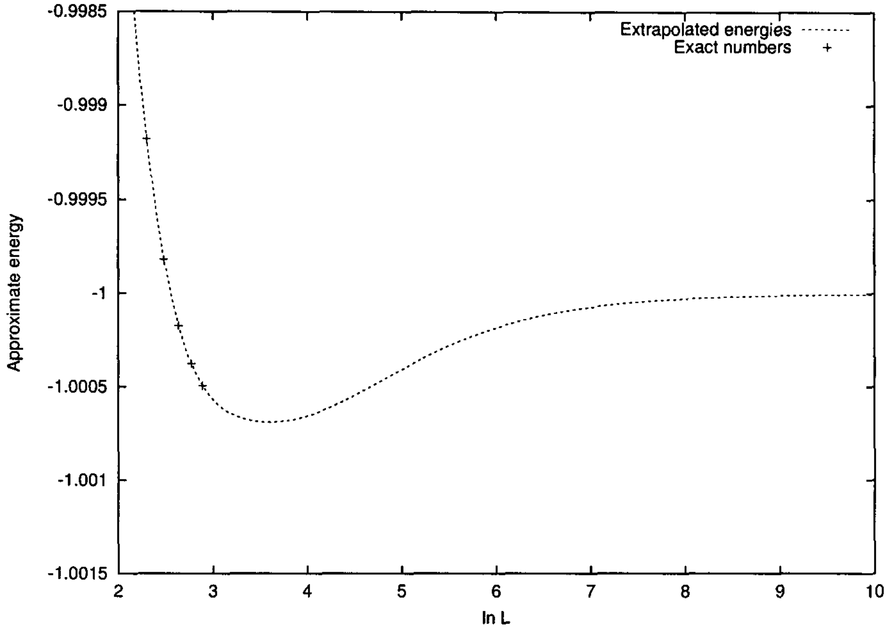


Figure 8. Expected approximations to the vacuum energy at different levels of truncation, extrapolated from data at lower levels of truncation.

predicted by Sen as $L \to \infty$.

7.4. *Gauge fixing*

In this subsection we discuss some aspects of the Feynman-Siegel gauge choice used in most explicit calculations in OSFT to date. Let us restrict attention to the zero momentum action for even-level scalar fields. This action is invariant under (54) with a general gauge parameter of the form

$$\Lambda = \sum \mu^a |s_a\rangle = \mu_1 b_{-2}|0_1\rangle + \cdots . \tag{274}$$

The ghost number zero states $|s_a\rangle$ are annihilated by b_0, so they do not contain c_0. The variation of a general zero-momentum scalar field takes the form

$$\delta\phi_i = D^{ia}\mu_a + g\bar{\kappa}T^{ija}\phi_j\mu_a . \tag{275}$$

At $\phi_i = 0$, we have the linear variation $\delta\phi_i = D^{ia}\mu_a$. Let ϕ_q denote fields associated with ghost number one states that contain a c_0. For example, at level two there is a field η associated with the state $c_0 b_{-2}|0_1\rangle$. At each level, the number of fields ϕ_q is clearly equal to the number of gauge parameters μ_a; the corresponding states are simply related by removing or replacing the c_0. From the formula for $Q_B = c_0 L_0 + \cdots$, it is easy to verify that D^{qa} is a linear one-to-one map at each level, so

$$\det D^{qa} \neq 0 \tag{276}$$

holds at each level. This is why the Feynman-Siegel gauge, which sets $\phi_q = 0$ at each level (and which limits us to gauge parameters associated with states without a c_0), is a good gauge choice near $\phi_i = 0$, as shown in subsection 7.2.

Let us now consider the gauge transformations at a general point in field space $\langle\phi_i\rangle$. We have

$$\delta\phi^i = M^{ia}\mu_a \tag{277}$$

where

$$M^{ia} = D^{ia} + g\bar{\kappa}T^{ija}\langle\phi_j\rangle\,. \tag{278}$$

Feynman-Siegel gauge breaks down whenever the determinant of this matrix vanishes

$$\det M^{qa} = 0\,. \tag{279}$$

This condition defines a region in field space within which Feynman-Siegel gauge is valid. At the boundary of this region, some gauge transformations give field variations which are tangent to the Feynman-Siegel gauge-fixed hypersurface. Some gauge orbits which cross the Feynman-Siegel gauge surface inside this region will cross again outside the region, giving a form of Gribov ambiguity. Furthermore, some gauge orbits never encounter the region of gauge validity. Thus, Feynman-Siegel gauge is really only locally valid.

We can study the region of Feynman-Siegel gauge validity in level truncation, using finite matrices M^{qa}. It is instructive to consider a simple example of the breakdown of this gauge choice. Consider dropping all fields other than the tachyon $\phi = \phi_1$ and the field $\eta = \phi_4$. The gauge transformation rules then become

$$\delta\phi = \mu g\bar{\kappa}\left[-\frac{16}{9}\phi + \frac{128}{81}\eta\right] \tag{280}$$
$$\delta\eta = -\mu + \mu g\bar{\kappa}\left[-\frac{224}{81}\phi + \frac{1792}{729}\eta\right].$$

In this simple model, M is a one-by-one matrix,

$$M = -\mu(1 + g\bar{\kappa}\frac{224}{81}\phi). \tag{281}$$

The gauge choice $\eta = 0$ breaks down when $\eta = \delta\eta = 0$ which occurs when

$$\phi = -\frac{1}{g\bar{\kappa}}\frac{81}{224}. \tag{282}$$

It is easy to see that smaller values of ϕ are gauge-equivalent to values of ϕ above this boundary value, while some gauge orbits never intersect the line $\eta = 0$.

The complete action including all even level (zero momentum) scalar fields and gauge invariances has been computed up to level (8, 16).[91] One result of this computation is that the Feynman-Siegel gauge boundary condition $\det M^{qa} = 0$ seems to be very stable near the origin as the level of truncation is increased. This gives some confidence that there is a well-defined finite region in field space where Feynman-Siegel gauge is valid, and that the boundary of this region can be arbitrarily well approximated by level-truncation calculations. Another interesting result which can be seen from these calculations is that (to the precision possible in the level-truncated analysis) the branch points in the tachyon effective potential arise precisely at those points where the trajectory in field space associated with the effective potential crosses the Feynman-Siegel gauge boundary. Thus, these branch points are gauge artifacts. As mentioned previously, the tachyon effective potential computed from boundary string field theory does not suffer from such branch point problems.

It would be very desirable, however, to have an approach which enables one to describe the full string field space, including configurations which do not have gauge representatives in the local region of Feynman-Siegel gauge validity. Other gauge choices can be made, but those which have been explored to date are only minor variations on the Feynman-Siegel gauge choice, and do not lead to qualitatively different results. One might have hoped to isolate the true vacuum without gauge fixing at all, given that level truncation breaks the gauge symmetry and thus allows a discrete set of solutions at any level. This approach, however, is not particularly promising: the solutions found at each level lie at very different places on the gauge orbit, and do not approach any natural limit. Nonetheless, it seems of paramount importance to find some method for exploring the full field space of the theory. Currently inaccessible regions of the field space may contain solutions that have not yet been found (see subsection 7.6).

7.5. *Lower-dimensional D-branes as solitons*

One aspect of the Sen conjectures (item (2) in the list of section 3.3), proposes that lower-dimensional D-branes can be viewed as solitons of the D25-brane string field theory. The solitons involve profiles for the tachyon field which arise because the tachyon potential is non-trivial. The tachyon solitons are lumps, as opposed to kinks, which appear in superstring field theory solitons.

In this section we will discuss the basic ideas required to test this conjecture. We will follow the approach of Möeller, Sen, and Zwiebach[92] (other attempts[93] do not use level expansion). In order to be able to use a level expansion we curl up one spatial coordinate x into a circle of radius R (the corresponding string coordinate is called X). We will work with $R > 1$. Along this direction, we will wrap a D1-brane. We will then consider the possibility that a certain process of tachyon condensation results in the D1-brane becoming a D0-brane. Our use of D1- and D0-branes is just a matter of notational ease. Additional D-brane dimensions could be included.

Recall that the mass of the D1-brane can be written in the form

$$M_{D1} = 2\pi R T_1 = \frac{1}{2\pi^2 g^2} \tag{283}$$

where g is the coupling constant of the open string field theory that describes the D1-brane:

$$S = -\frac{1}{g^2}\Big(\frac{1}{2}\langle\Phi, Q\Phi\rangle + \frac{1}{3}\langle\Phi, \Phi * \Phi\rangle\Big) \equiv -\frac{1}{g^2}\mathcal{V}(\Phi)\,. \tag{284}$$

A few remarks are in order. In the above string action we have included into the string coupling factor $(1/g^2)$ the volume $(2\pi R)$ of the compact circle where the D1-brane is wrapped. By doing so, we can still use a CFT overlap with unit normalization, and the right-hand side in (283) gives the total mass of the brane. The zero string field here is supposed to describe the vacuum with a D1-brane stretched around the circle. For time-independent string fields (the kind of fields we consider here), $\mathcal{V}(\Phi)$ is a potential. More precisely the potential energy P.E. associated with a string field is

$$\text{P.E.} = -S(\Phi) = \frac{1}{g^2}\mathcal{V}(\Phi) = (2\pi R T_1)\, 2\pi^2\, \mathcal{V}(\Phi), \tag{285}$$

where we used (283). This potential energy is really the potential energy of field configurations measured with respect to the D1-brane background. Therefore, the total energy E_{tot} of the configuration is obtained by adding the energy of the D1-brane to the above P.E. We find

$$E_{tot}(\Phi) = (2\pi R T_1)\Big(1 + 2\pi^2\, \mathcal{V}(\Phi)\Big). \tag{286}$$

Since we will use the level expansion to investigate if a D0-brane can be represented as a lump solution, it is reasonable to use the level expansion to calculate the mass of the D1-brane, as well. So, we re-express the energy of the D1-brane in (286) in terms of the string field potential at the vacuum. Let $\Phi = T_{vac}$ denote the string field of the D1-brane SFT that represents the tachyon vacuum. Then, we have $-1 = 2\pi^2 \mathcal{V}(T_{vac})$, and we can rewrite

$$E_{tot}(\Phi) = (2\pi R T_1)\Big(2\pi^2\, \mathcal{V}(\Phi) - 2\pi^2\, \mathcal{V}(T_{vac})\Big). \tag{287}$$

Indeed, this formula works correctly: when $\Phi = 0$ the total energy equals the mass of the D1-brane, and when $\Phi = T_{vac}$ the energy is zero (since the D1-brane has disappeared).

Let T_{lump} denote the lump (string field) solution, which is expected to represent the D0-brane in the field theory of the D1-brane. The energy of the lump solution is obtained from (287) for $\Phi = T_{lump}$:

$$E_{lump} = E_{tot}(T_{lump}) = (2\pi R T_1)\Big(2\pi^2\, \mathcal{V}(T_{lump}) - 2\pi^2\, \mathcal{V}(T_{vac})\Big). \tag{288}$$

The tensions T_0 and T_1 of the D0- and the D1-branes are related by $T_0 = 2\pi T_1$ (the D0-brane tension is the D0-brane energy). We can therefore form the ratio $r(R)$ of the lump energy and the D0-brane energy

$$r(R) = \frac{E_{lump}}{T_0} = R\Big(2\pi^2\, \mathcal{V}(T_{lump}) - 2\pi^2\, \mathcal{V}(T_{vac})\Big). \tag{289}$$

In the exact solution (or at infinite level), the ratio $r(R)$ should be equal to one. This is the content of the second tachyon conjecture. At any finite level $r(R)$ is some slowly varying function of R. Testing the conjecture for $R \to 1$ is quite difficult, and one must go to very high level in the computation. Testing the conjecture for R very large is also laborious, since many terms enter into any finite-level expansion. So, in practice, one chooses some reasonable value of R (the value $R = \sqrt{3}$ is convenient) and calculates to a fixed level.

Before reviewing some of the results obtained, let's do the simplest computation explicitly. We consider a tachyon field $T(x)$ which is expanded as

$$T(x) = t_0 + \sum_{n=1}^{\infty} t_n \cos(nx/R)\,. \tag{290}$$

The corresponding string field is written as

$$\begin{aligned} |T\rangle \;&= t_0 c_1|0\rangle + \sum_{n=1}^{\infty} \frac{1}{2} t_n \Big(e^{inX(0)/R} + e^{-inX(0)/R} \Big)\, c_1|0\rangle\,, \\ &= t_0 c_1|0\rangle + \sum_{n=1}^{\infty} \frac{1}{2}\, t_n \Big(c_1|n/R\rangle + c_1|-n/R\rangle \Big)\,. \end{aligned} \tag{291}$$

We now evaluate the string action, keeping t_0 and the first tachyon harmonic t_1:

$$|T\rangle = t_0 c_1|0\rangle + \frac{1}{2}\, t_1 \Big(c_1|1/R\rangle + c_1|-1/R\rangle \Big)\,. \tag{292}$$

Consider first the contribution of t_1 to the kinetic term

$$\begin{aligned} \frac{1}{2}\langle T, QT\rangle\Big|_{t_1} \;&= \frac{1}{2}\frac{t_1}{2}\cdot\frac{t_1}{2}\Big(\big(\langle 1/R| + \langle -1/R|\big) c_{-1}c_0 L_0 c_1 \big(|1/R\rangle + |-1/R\rangle\big) \Big) \\ &= \frac{1}{4} t_1^2 \Big(-1 + \frac{1}{R^2} \Big)\,. \end{aligned} \tag{293}$$

Note that, as mentioned earlier, the overlaps have unit normalization. Let us now calculate the terms that arise from the interaction. Because of momentum conservation there are no t_1^3 or $t_1 t_0^2$ terms. There is only a $t_1^2 t_0$ coupling, which is readily calculated as

$$\frac{1}{3}\cdot 3\cdot \frac{t_1}{2}\cdot\frac{t_1}{2} t_0 \cdot 2 \cdot \langle c_1 e^{iX/R}, c e^{-iX/R}, c\rangle = \frac{1}{2} t_0 t_1^2 K^{3-\frac{2}{R^2}}\,. \tag{294}$$

Let us explain the origin of the various factors. The first $1/3$ is the one that comes with the interaction term in the action. The factor of 3 is because there are three possible places to insert the operator associated with t_0. The factors of $t_1/2$ and t_0 come from the field expansion, and the factor of two arises because there are two ways in which the momentum can be conserved. The correlator has been evaluated in a way similar to the previous computation that led to (122). Indeed, the only difference is that the conformal dimension of two of the operators has been shifted from -1 to $-1+\frac{1}{R^2}$.

Collecting now our results and using the previously calculated potential for t_0 (123) we find

$$\mathcal{V}(t_0, t_1) = -\frac{1}{2} t_0^2 - \frac{1}{4}\Big(1 - \frac{1}{R^2}\Big) t_1^2 + \frac{1}{3} K^3 t_0^2 + \frac{1}{2} t_0 t_1^2 K^{3-\frac{2}{R^2}}\,. \tag{295}$$

The original tachyon is still there: it corresponds to the field t_0, which in the present expansion has no momentum. For $R > 1$, the field t_1 is also a tachyon. This field is present because of the instability to form a D1-brane. Indeed, for $R > 1$ the energy of the D1-brane is larger than the energy of the D0-brane, and the decay is possible. For $R < 1$, the D0-brane has more energy than the D1-brane. In this case, it is not clear if some high level computation can exhibit the D0-brane as a solution of the D1-brane field theory. We return to this problem in the next subsection.

Let's take $R = \sqrt{3}$. In this case the potential $\mathcal{V}(t_0, t_1)$ has a critical point which represents a lump: $t_0 \simeq 0.18$ and $t_1 = -0.34$. Of course there is also the conventional tachyon vacuum solution with $t_0 = 1/K^3$ and $t_1 = 0$. With these two solutions, one can readily compute the ratio $r(\sqrt{3})$ in (289). We find $r(\sqrt{3}) \simeq 0.774$ in this lowest order calculation. The result is certainly quite good. This computation is called a level (1/3; 2/3) computation since the highest level field t_1 has level $1/3 = 1/R^2$, and we kept terms in the potential up to level 2/3. A computation at level (2,4) gives $r \simeq 1.02$, and for level (3,6) one finds $r \simeq 0.994$. The convergence to the answer is quite spectacular. This computation includes the tachyon harmonics t_1, t_2, and t_3, as well as fields from the second level and their first harmonics. No higher level computations have been done for this problem. The computations are not completely universal since the Virasoro structure of the state space depends on the radius of the circle. For rational values of R one may find null states, so this is why we took R irrational. Even for R irrational, not all states can be written as Virasoro descendents of the vacuum $|0\rangle$. New primaries (and their descendents) are needed starting at level 4.

Since we are equipped with the tachyon harmonics, one is able to construct explicitly the tachyon profile for the lump solution which represents the D0-brane. As the level is increased, the profile appears to settle into a well-defined limit. That same profile appears to arise for various values of the radius R of the circle used for the computation. The profile is roughly of the form

$$T(x) \simeq a + b\, e^{-x^2/(2\sigma^2)}\,, \quad a \simeq 0.56, \quad b \simeq -0.83, \quad \sigma \simeq 1.52\,. \qquad (296)$$

The σ width of the lump is therefore about $1.5\sqrt{\alpha'}$. The significance (or gauge independence) of this width is not clear. Nevertheless, it is interesting that D-branes, which are defined by definite positions in CFT, appear as thick objects in SFT. Physical questions regarding D-branes are expected to have identical answers in the two approaches.

The above computations have been generalized to the case of lump solutions of codimension two. In this case, we can imagine a D2-brane wrapped on a torus T^2 which decays into a D0-brane. The results in the level expansion appear to confirm that the lump solutions do represent D0-branes. Less accurate results are obtained; the energy has only been estimated with about ten percent accuracy.

The above results have simple analogs in field theory.[94] Consider a simple scalar field theory in $p+1$ spatial dimensions, where we single out a coordinate x for special treatment:

$$S = \int dt d^p y dx \left\{ \frac{1}{2}\left(\frac{\partial \phi}{\partial t}\right)^2 - \frac{1}{2}|\nabla_y \phi|^2 - \frac{1}{2}\left(\frac{\partial \phi}{\partial x}\right)^2 - V(\phi) \right\}. \tag{297}$$

As you can readily verify, time-independent solitonic solutions $\phi(x)$, which depend only on the coordinate x, are obtained by solving the second-order ordinary differential equation

$$\frac{d^2\phi}{dx^2} = V'(\phi(x)). \tag{298}$$

This equation takes the form of the equation of motion of a unit mass particle in a one-dimensional potential $-V(x)$. As an example, we consider a theory with potential[95]

$$V(\phi) = \frac{1}{3}(\phi - 1)^2 \left(\phi + \frac{1}{2}\right). \tag{299}$$

The potential has a maximum at $\phi = 0$ and a local minimum at $\phi = 1$. At $\phi = 0$ the interpretation is that of a D$(p+1)$-brane with tension

$$T_{p+1} = V(\phi = 0) = \frac{1}{6}. \tag{300}$$

As a simple exercise, verify that

$$\phi(x) = 1 - \frac{3}{2}\operatorname{sech}^2(x/2), \tag{301}$$

is a lump solution for this potential.

Exercise: Show that the lump solution is an object with tension $T_p = 6/5$.

In string theory the ratio $\frac{1}{2\pi}\frac{T_p}{T_{p+1}}$ is equal to one. In this field theory model with a cubic potential, we find

$$\frac{1}{2\pi}\frac{T_p}{T_{p+1}} = \frac{1}{2\pi}\frac{6}{5}\cdot 6 = \frac{18}{5\pi} \simeq 1.146. \tag{302}$$

It is also a familiar result in soliton field theory that the spectrum of excitations that live on the world-volume of the lump solution $\bar{\phi}(x)$ is governed by a Schrödinger equation with a potential $V''(\bar{\phi}(x))$. The mass-squared values for the modes that live on the lump coincide with the Schrödinger energies.

There has been some interest in finding potentials that accurately describe the behavior of the tachyon. While the kinetic terms are not standard, the potential

$$V(\phi) = -\frac{1}{4}\phi^2 \ln \phi^2 \,, \quad \phi > 0 \,. \tag{303}$$

appears to be an exact effective tachyon potential. This potential was obtained[96] in an attempt to construct realistic tachyon potentials, and was later confirmed to appear in the BSFT approach to string field theory.[10] The tachyon vacuum is at $\phi = 0$, and surprisingly (but correctly!) the tachyon mass goes to infinity at this vacuum. This is consistent with the conjecture that perturbative open string degrees of freedom disappear at the tachyon vacuum.

Exercise. Show that $\bar{\phi}(x) = \exp(-x^2/4)$ is the lump solution for the potential (303) and the Schrödinger potential for fluctuations on the lump solution is $\frac{x^2}{4} - \frac{3}{2}$, a simple harmonic oscillator potential. Finally, confirm that the values of m^2 for the particles that live on the lump are $-1, 0, 1, 2, \ldots$. This is the expected string spectrum!

7.6. *Open string theory backgrounds*

We mentioned in the last subsection that when the radius R of a circle on which a D1-brane is compactified becomes small, it is not known how to represent a D0-brane in the string field theory on the D1-brane. When $R < 1$, the energy of the resulting D0-brane is larger than the energy of the original D1-brane. Thus, such a solution would have positive energy with respect to the original system. The difficulty of constructing such a D0-brane solution is an example of a more general, and we believe crucial, question for OSFT: Does OSFT, either through level truncation or some more sophisticated analytic approach, admit classical solutions which describe open string backgrounds with higher energy than the configuration with respect to which the theory is originally defined? If OSFT is to be a truly complete formulation of string theory, such solutions must be possible, since all open string backgrounds must be accessible to the theory.

Another problem of this type is to find, either analytically or numerically, a solution of the OSFT formulated with one D25-brane that describes *two* D25-branes. It should be just as feasible to go from a vacuum with one D-brane to a vacuum with two D-branes as it is to go from a vacuum with one D-brane to the empty vacuum. Despite some work on this problem,[97] there is as yet no evidence of a solution. Several approaches which have been tried include: *i*) following a positive mass field upward, looking for a stable point; this method seems to fail because of gauge-fixing problems—the effective potential often develops a singularity before reaching the energy $+T_{25}$, *ii*) following the intuition of the RSZ model (discussed in the following section) and constructing a gauge transform of the original D-brane solution which is $\star$—orthogonal to the original D-brane vacuum. It can be shown formally that such a state, when added to the original D-brane vacuum gives a new solution with the correct energy for a double D-brane; unfortunately, however, we have been unable to identify such a state numerically in level truncation.

While so far no progress has been made towards the construction of solutions with higher energy than the initial vacuum, it is also interesting to consider the marginal case. An example of such a situation is embodied is the problem of translating a single D-brane of less than maximal dimension in a transverse direction. It was shown by Sen and Zwiebach[98] (in a T-dual picture) that after moving a D-brane a finite distance of order of the string length in a transverse direction, the level-truncated string field theory equations develop a singularity. Thus, in level truncation it does not seem possible to move a D-brane a macroscopic distance in a transverse direction[g]. In this case, a toy model[95] suggests that the problem is that the infinitesimal marginal parameter for the brane translation ceases to parameterize the marginal trajectory in field space after a finite distance, just as the coordinate x ceases to parameterize the circle $x^2 + y^2 = 1$ near $x = 1$. Indeed, an explicit calculation[82] of the field redefinition needed to take the OSFT field A associated with the transverse motion to the correct marginal parameter a shows that this field redefinition has a subleading term

$$A = a + \alpha a^3 + \cdots , \tag{304}$$

where $\alpha < 0$. Thus, as a increases, eventually a point is reached where A begins to decrease. This shows that A is not a good parameter for marginal deformations of arbitrary size. It would be nice to have a clear understanding of how arbitrary marginal deformations are encoded in the theory.

[g]Although this can be done formally,[99] it is unclear how the formal solution relates to an explicit expression in the oscillator language.

To show that open string field theory is sufficiently general to address arbitrary questions involving different vacua, it is clearly necessary to show that the formalism is powerful enough to describe multiple brane vacua, the D0-brane lump on an arbitrary radius circle, and translated brane vacua. It is currently unclear whether the obstacles to finding these vacua are technical or conceptual. It may be that the level-truncation approach is not well-suited to finding these vacua, and a new approach is needed.

8. String field theory around the stable vacuum

The tachyon conjectures state that the classically stable vacuum is the closed string vacuum. This implies that there should be no open string excitations in this vacuum, given that the D-brane represented by the original OSFT has decayed and exists no more. Without a D-brane conventional perturbative open string states are not expected to exist. If any perturbative states exist in this vacuum, they should be closed string states, which are only expected to appear in the quantum open string field theory.

There are two natural questions concerning this conjecture. First, we ask: Can it be tested? For this, we can begin with the original OSFT on the background of a D25-brane, for example, and use the (numerical) solution Φ_0 for the tachyon vacuum to expand the classical OSFT around the tachyon vacuum and to calculate the spectrum. The conjecture requires that no physical states be encountered. Second, we ask: Is there a more natural formulation of open string theory around the tachyon vacuum, in which, for example, the background independence of the theory might be more manifest? The theory around the tachyon vacuum, is, no doubt, rather unusual. In the tachyon vacuum there are no apparent physical states, at least none that take any familiar form. Physical perturbative states can arise only from quantum effects or classically after the theory is shifted to a nontrivial background that represents some D-brane configuration.

The tachyon vacuum is a rather special vacuum: it is the end product of the decay of *any* D-brane configuration. Presumably, the theory at the tachyon vacuum is independent of the particular version of OSFT used to reach it upon tachyon condensation, in the sense that the string field theories associated with different D-brane configurations should be equivalent under field redefinition around the stable vacuum of each theory. If that is the case, there may exist a theory – which we can call *Vacuum String Field*

Theory, or VSFT – which formulates the physics of the tachyon vacuum directly, *without* using a D-brane background to reach the tachyon vacuum.

Presently, there is background dependence in the formulation of Witten's OSFT; some specific D-brane background must be chosen to define the theory, even though this D-brane configuration may be removed through tachyon condensation. As a result, even if the theory is in an abstract sense completely background independent, we are stuck with some particular choice of "coordinates" on the theory arising from the original choice of background, which may make physics in other backgrounds rather difficult to disentangle. The tachyon vacuum is also a specific background, but it is certainly a choice that is more canonical than one which picks one out of an infinite number of D-brane configurations. There are perhaps two canonical choices: an infinite number of space-filling D-branes, which has been motivated from the viewpoint of K-theory,[17] and a background with no D-branes whatsoever – the tachyon vacuum. In this section we investigate the second choice.

A strikingly simple formulation of VSFT was proposed by Rastelli, Sen, and Zwiebach (RSZ),[100] in which the BRST operator is taken to be purely contained in the ghost sector. In this theory, closed-form analytic solutions that represent D-branes can be found and take the form of projectors of the star-algebra. One shortcoming of this VSFT is that certain computations are singular and require regularization. It remains to be seen if a regular VSFT exists.

In subsection 8.1 we describe the form of the OSFT action when expanded around the classically stable tachyon vacuum. Subsection 8.2 describes evidence from Witten's OSFT that the open string degrees of freedom truly disappear from the theory in this vacuum. In 8.3 we introduce and discuss the RSZ model of VSFT. Subsection 8.4 describes an important class of states in the star algebra: slivers and projectors, which play a key role in constructing D-branes in the RSZ model, and which may also be useful in understanding solutions of the Witten theory. Finally, in 8.5 we discuss closed strings in OSFT.

8.1. *String field theory in the true vacuum*

We have seen that numerical results from level-truncated string field theory strongly suggest the existence of a classically stable vacuum solution Φ_0 to the string field theory equation of motion. The state Φ_0, while still unknown analytically, has been determined numerically to a high degree of precision. This state seems like a very well-behaved string field configuration. While there is no positive-definite inner product on the string field Fock space, the state Φ_0 certainly has finite norm under the natural inner product $\langle V_2|\Phi_0, c_0 L_0 \Phi_0\rangle$, and is even better behaved under the product $\langle V_2|\Phi_0, c_0\Phi_0\rangle$. Thus, it is natural to assume that Φ_0 defines a classically stable vacuum for the theory, around which we can expand the action to find a string field theory around the tachyon vacuum.

Let Φ_0 be the string field configuration describing the tachyon vacuum. This string field satisfies the classical field equation

$$Q\Phi_0 + \Phi_0 * \Phi_0 = 0\,. \tag{305}$$

If $\widetilde{\Phi} = \Phi - \Phi_0$ denotes the shifted open string field, then the cubic string field theory action (61) expanded around the tachyon vacuum has the form:

$$S(\Phi_0 + \widetilde{\Phi}) = S(\Phi_0) - \frac{1}{g^2}\left[\frac{1}{2}\langle\,\widetilde{\Phi}\,,\,\widetilde{Q}\,\widetilde{\Phi}\rangle + \frac{1}{3}\langle\,\widetilde{\Phi}\,,\,\widetilde{\Phi} * \widetilde{\Phi}\rangle\right]. \tag{306}$$

Here $S(\Phi_0)$ is a constant, which according to the energetics part of the tachyon conjectures equals the tension of the D-brane times its volume (as before, we assume that the time interval has unit length so that the action can be identified with the negative of the potential energy for static configurations). The kinetic operator $\widetilde{Q}$ is given in terms of Q and Φ_0 as:

$$\widetilde{Q}\widetilde{\Phi} = Q\widetilde{\Phi} + \Phi_0 * \widetilde{\Phi} + \widetilde{\Phi} * \Phi_0\,. \tag{307}$$

More generally, on arbitrary string fields one would define

$$\widetilde{Q}A = QA + \Phi_0 * A - (-1)^A A * \Phi_0\,. \tag{308}$$

The consistency of the new action (306) is guaranteed from the consistency of the action in (61). Since neither the inner product nor the star multiplication have changed, the identities in (63) still hold. One can also check that the identities in (62) hold when Q is replaced by $\widetilde{Q}$. Just as the original action is invariant under the gauge transformations (71), the new action is invariant under $\delta\widetilde{\Phi} = \widetilde{Q}\Lambda + \widetilde{\Phi} * \Lambda - \Lambda * \widetilde{\Phi}$ for any Grassmann-even ghost-number zero state Λ.

Since the energy density of the brane represents a positive cosmological constant, it is natural to add the constant $-M = -S(\Phi_0)$ to (61). This will cancel the $S(\Phi_0)$ term in (306), and will make manifest the expected zero energy density in the final vacuum without D-brane. For the analysis around this final vacuum it suffices therefore to study the action

$$S_0(\widetilde{\Phi}) \equiv -\frac{1}{g^2}\left[\frac{1}{2}\langle\,\widetilde{\Phi}\,,\,\widetilde{Q}\,\widetilde{\Phi}\rangle + \frac{1}{3}\langle\,\widetilde{\Phi}\,,\,\widetilde{\Phi} * \widetilde{\Phi}\rangle\right]. \tag{309}$$

This string field theory around the stable vacuum has precisely the same form as Witten's original cubic string field theory, only with a different BRST operator $\widetilde{Q}$, which so far is only determined numerically. While this is insufficient for a complete formulation, it suffices to test the conjecture that open string excitations disappear in the tachyon vacuum, as we will discuss in Section 8.2.

The numerical solution for Φ_0 provides a numerical definition of the string field theory around the tachyon vacuum. How could we do better? If we had a closed form solution Φ_0 available, the problem of formulating SFT around the tachyon vacuum would be significantly simplified. It is not clear, however, that the resulting formulation would be the best possible one. Previous experience with background deformations (small and large) in SFT indicates that even if we knew Φ_0 explicitly and constructed $S_0(\widetilde{\Phi})$ using eq.(309), this may not be the most convenient form of the action. Typically a nontrivial field redefinition is necessary to bring the shifted SFT action to the canonical form representing the new background.[101] In fact, in some cases, such as in the formulation of open SFT for D-branes with various values of magnetic fields, it is possible to formulate the various SFT's directly,[102,103] but the nontrivial classical solution relating theories with different magnetic fields are not known. This suggests that if a simple form exists for the SFT action around the tachyon vacuum it might be easier to guess it than to derive it.

In fact, this is exactly the approach to the formulation of vacuum string field theory (VSFT) taken by Rastelli, Sen, and Zwiebach (RSZ).[100] These authors postulate that at the tachyon vacuum the action takes the form

$$\mathcal{S}(\Phi) \equiv -K_0 \left[\frac{1}{2}\langle \Phi , \mathcal{Q}\Phi\rangle + \frac{1}{3}\langle \Phi , \Phi * \Phi\rangle \right], \qquad (310)$$

where the new kinetic operator $\mathcal{Q}$ is an operator build solely out of ghosts fields. If this gives a consistent theory at the tachyon vacuum, they argue, their choice of $\mathcal{Q}$ must be field redefinition equivalent to the $\widetilde{Q}$ that arises directly by shifting the original OSFT action with the tachyon solution Φ_0. We discuss the RSZ model in Section 8.3.

8.2. *Decoupling of open strings*

It may seem surprising to imagine that *all* the perturbative open string degrees of freedom will vanish at a particular point in field space, since these are all the apparent degrees of freedom available in the theory. This is not a familiar phenomenon from quantum field theory. To understand how the open strings can decouple, it may be helpful to begin by considering the simple example of the (0, 0) level-truncated theory. In this theory, the quadratic terms in the action become

$$-\int d^{26}p\,\phi(-p)\left[\frac{p^2-1}{2}+g\bar{\kappa}\left(\frac{16}{27}\right)^{p^2}\cdot 3\langle\phi\rangle\right]\phi(p)\,. \tag{311}$$

Taking $\langle\phi\rangle = 1/3\bar{\kappa}g$, we find that the quadratic term is a transcendental expression which does not vanish for any real value of p^2. Thus, this theory has no poles, and the tachyon has decoupled from the theory. Of course, this is not the full story, as there are still finite complex poles. It does, however suggest a mechanism by which the nonlocal parts of the action (encoded in the exponential of p^2) can remove physical poles.

To get the full story, it is necessary to continue the analysis to higher level. At level 2, there are 7 scalar fields, the tachyon and the 6 fields associated with the Fock space states

$$\begin{array}{rl} (\alpha_{-1}\cdot\alpha_{-1})|0_1,p\rangle & b_{-1}\cdot c_{-1}|0_1,p\rangle \\ c_0\cdot b_{-1}|0_1,p\rangle & (p\cdot\alpha_{-2})|0_1,p\rangle \\ (p\cdot\alpha_{-1})^2|0_1,p\rangle & (p\cdot\alpha_{-1})c_0b_1|0_1,p\rangle \end{array} \tag{312}$$

Note that in this analysis we cannot fix Feynman-Siegel gauge, as we only believe that this gauge is valid for the zero-modes of the scalar fields in the vacuum Ψ_0. An attempt at analyzing the spectrum of the theory in Feynman-Siegel gauge using level truncation has been made,[57] but gave no sensible results. Diagonalizing the quadratic term in the action on the full set of 7 fields of level ≤ 2, we find[104] that poles develop at $M^2 = 0.9$ and $M^2 = 2.0$ (in string units, where the tachyon has $M^2 = -1$). These poles correspond to states satisfying $\tilde{Q}\tilde{\Psi} = 0$. The question now is, are these states physical? If they are exact states, of the form $\tilde{\Psi} = \tilde{Q}\tilde{\Lambda}$, then they are simply gauge degrees of freedom. If not, however, then they are states in the cohomology of $\tilde{Q}$ and should be associated with physical degrees of freedom. Unfortunately, we cannot precisely determine whether the poles we find in level truncation are due to exact states, as the level-truncation procedure breaks the condition $\tilde{Q}^2 = 0$. Thus, we can only measure *approximately* whether a state is exact. A detailed analysis of this question was carried out by Ellwood and Taylor.[104] In their paper, all terms in the SFT action of the form $\phi_i\,\psi_j(p)\,\psi_k(-p)$ were determined, where ϕ_i is a scalar zero-mode, and $\psi_{j,k}$ are nonzero-momentum scalars. In addition, all gauge transformations involving at least one zero-momentum field were computed up to level (6, 12). At each level up to $L = 6$, the ghost number 1 states in the kernel Ker $\tilde{Q}^{(1)}_{(L,2L)}$ were computed. The extent to which each of these states lies in the exact subspace was measured using the formula

$$\% \text{ exactness} = \sum_i \frac{(s \cdot e_i)^2}{(s \cdot s)} \tag{313}$$

where $\{e_i\}$ are an orthonormal basis for Im $\tilde{Q}^{(0)}_{(L,2L)}$, the image of $\tilde{Q}$ acting on the space of ghost number 0 states in the appropriate level truncation. (Note that this measure involves a choice of inner product on the Fock space; several natural inner products were tried, giving roughly equivalent results). The result of this analysis was that up to the mass scale of the level truncation, $M^2 \leq L - 1$, all the states in the kernel of $\tilde{Q}^{(1)}$ were $\geq 99.9\%$ within the exact subspace, for $L \geq 4$. This result seems to give very strong evidence for Sen's third conjecture that there are no perturbative open string excitations around the stable classical vacuum Ψ_0. This analysis was only carried out for even level scalar fields; it would be nice to check that a similar result holds for odd-level fields and for tensor fields of arbitrary rank.

Another more abstract argument that there are no open string states in the stable vacuum was given by Ellwood, Feng, He and Moeller.[63] These authors argued that in the stable vacuum, the identity state $|I\rangle$ in the SFT star algebra, which satisfies $I \star A = A$ for a very general class of string fields A, seems to be an exact state,

$$|I\rangle = \tilde{Q}|\Lambda\rangle \,. \tag{314}$$

If indeed the identity is exact, then it follows immediately that the cohomology of $\tilde{Q}$ is empty, since $\tilde{Q}A = 0$ then implies that

$$A = (\tilde{Q}\Lambda) \star A = \tilde{Q}(\Lambda \star A) - \Lambda \star \tilde{Q}A = \tilde{Q}(\Lambda \star A) \,. \tag{315}$$

So to prove that the cohomology of $\tilde{Q}$ is trivial, it suffices to show that $\tilde{Q}|\Lambda\rangle = |I\rangle$. While there are some subtleties involved with the identity string field, Ellwood *et al.* found a very elegant expression for this field,

$$|I\rangle = \left(\cdots e^{\frac{1}{8}L_{-16}} e^{\frac{1}{4}L_{-8}} e^{\frac{1}{2}L_{-4}}\right) e^{L_{-2}}|0\rangle \,. \tag{316}$$

(Recall that $|0\rangle = b_{-1}|0_1\rangle$.) They then looked numerically for a state $|\Lambda\rangle$ satisfying (314). For example, truncating at level $L = 3$,

$$\begin{aligned} |I\rangle &= |0\rangle + L_{-2}|0\rangle + \cdots \\ &= |0\rangle - b_{-3}c_1|0\rangle - 2b_{-2}c_0|0\rangle + \frac{1}{2}(\alpha_{-1} \cdot \alpha_{-1})|0\rangle + \cdots \end{aligned} \tag{317}$$

while the only candidate for $|\Lambda\rangle$ is

$$|\Lambda\rangle = \alpha\, b_{-2}|0\rangle, \tag{318}$$

for some constant α. The authors of Ref. 63 showed that the state (317) is best approximated as exact when $\alpha \sim 1.12$; for this value, their measure of exactness becomes

$$\frac{\left|\tilde{Q}|\Lambda\rangle - |I\rangle\right|}{|I|} \rightarrow 0.17, \tag{319}$$

which the authors interpreted as a 17% deviation from exactness. Generalizing this analysis to higher levels, they found at levels 5, 7, and 9, a deviation from exactness of 11%, 4.5% and 3.5% respectively. At level 9, for example, the identity field has 118 components, and there are only 43 gauge parameters, so this is a highly nontrivial check on the exactness of the identity. Like the results of Ellwood and Taylor,[104] these results strongly support the conclusion that the cohomology of the theory is trivial in the stable vacuum. In this case, the result applies to fields of all spins and all ghost numbers.

Given that the Witten string field theory seems to have a classical solution with no perturbative open string excitations, in accordance with Sen's conjectures, it is quite interesting to ask what the physics of the string field theory in the stable vacuum should describe. One natural assumption might be that this theory should include closed string states in its quantum spectrum. Unfortunately, addressing this question requires performing calculations in the quantum theory around the stable vacuum. Such calculations are quite difficult (although progress in this direction has been made by Minahan in the p-adic version of the theory[105]). Even in the perturbative vacuum, it is difficult to systematically study closed strings in the quantum string from theory. We discuss this question again in the final subsection of this section.

8.3. *Pure ghost Vacuum String Field Theory*

Our discussion in Section 8.1 suggests that a VSFT may be formulated as a cubic string field theory, with some new choice $\mathcal{Q}$ for the kinetic operator. The choice of $\mathcal{Q}$ will be required to satisfy the following properties:

- The operator $\mathcal{Q}$ must be of ghost number one and must satisfy the conditions (62) that guarantee gauge invariance of the string action.
- The operator $\mathcal{Q}$ must have vanishing cohomology.

- The operator $\mathcal{Q}$ must be universal, namely, it must be possible to write without reference to the brane boundary conformal field theory.

The first condition is unavoidable; the theory must be gauge invariant if it is to be consistent. The second condition is reasonable, but perhaps stronger than needed: all we probably know is that there should be no cohomology at ghost number one, which is the ghost number at which physical states appear. The third constraint is the most stringent one. It implies that VSFT is an intrinsic theory that can be formulated without using an auxiliary D-brane.

The simplest possible choice is $\mathcal{Q} = 0$, which gives the purely cubic version of open string field theory.[106] Indeed, it has long been tempting to identify the tachyon vacuum with a theory where the kinetic operator vanishes because, lacking the kinetic term, the string field gauge symmetries are not spontaneously broken. Nevertheless, there are well-known complications with this identification. The string field shift $\bar{\Phi}$ that relates the cubic to the purely cubic OSFT appears to satisfy $Q\bar{\Phi} = 0$ as well as $\bar{\Phi} * \bar{\Phi} = 0$. The tachyon condensate definitely does not satisfy these two identities. We therefore search for nonzero $\mathcal{Q}$.

We can satisfy the three requirements by letting $\mathcal{Q}$ be constructed purely from ghost operators. In particular we claim that the ghost number one operators

$$\mathcal{C}_n \equiv c_n + (-)^n c_{-n}\,, \quad n = 0, 1, 2, \cdots \tag{320}$$

satisfy the properties

$$\begin{aligned}
&\mathcal{C}_n \mathcal{C}_n = 0, \\
&\mathcal{C}_n(A * B) = (\mathcal{C}_n A) * B + (-1)^A A * (\mathcal{C}_n B)\,, \\
&\langle \mathcal{C}_n A, B \rangle = -(-)^A \langle A, \mathcal{C}_n B \rangle\,.
\end{aligned} \tag{321}$$

The first property is manifest. The last property follows because under BPZ conjugation $c_n \to (-)^{n+1} c_{-n}$. The second property follows from the conservation laws[65]

$$\langle V_3 | (\mathcal{C}_n^{(1)} + \mathcal{C}_n^{(2)} + \mathcal{C}_n^{(3)}) = 0\,. \tag{322}$$

These conservation laws arise by consideration of integrals of the form $\int dz c(z) \varphi(z)$ where $\varphi(z)(dz)^2$ is a globally defined quadratic differential.

Each of the operators $\mathcal{C}_n$ has vanishing cohomology. To see this note that for each n the operator $\mathcal{B}_n = \frac{1}{2}(b_n + (-)^n b_{-n})$ satisfies $\{\mathcal{C}_n, \mathcal{B}_n\} = 1$.

It then follows that whenever $\mathcal{C}_n\psi = 0$, we have $\psi = \{\mathcal{C}_n, \mathcal{B}_n\}\psi = \mathcal{C}_n(\mathcal{B}_n\psi)$, showing that ψ is $\mathcal{C}_n$ trivial. Since they are built from ghost oscillators, all $\mathcal{C}_n$'s are manifestly universal.

It is clear from the structure of the consistency conditions that we can take $\mathcal{Q} = \sum_{n=0}^{\infty} a_n \mathcal{C}_n$, where the a_n's are constant coefficients. As we will see below, many properties of the RSZ theory follow simply from the fact that $\mathcal{Q}$ is pure ghost. But, there are some computations that may require a choice of $\mathcal{Q}$ (more on this later). The work of Hata and Kawano[107] gave the clue for the choice of $\mathcal{Q}$ taken by RSZ:

$$\begin{aligned}\mathcal{Q} &= \frac{1}{2i}(c(i) - \bar{c}(i)) = \frac{1}{2i}(c(i) - c(-i)) = \sum_{n=0}^{\infty}(-1)^n \mathcal{C}_{2n}\,,\\ &= c_0 - (c_2 + c_{-2}) + (c_4 + c_{-4}) - \cdots\,.\end{aligned} \tag{323}$$

Since the canonical zero-time open string in the complex z-plane is the half-circle $|z| = 1$ that lies on the upper half plane, the operator $c(i)$ represents a ghost insertion precisely at the open string midpoint. This is the most delicate point on the open string given that the three string interaction is a world-sheet with a curvature singularity at the point where the three string midpoints meet. The other operator $c(-i)$ is needed in order that $\mathcal{Q}$ is twist invariant (see the first equation in (75)). With this choice of $\mathcal{Q}$, the string field action is written as

$$S = -K_0\Big[\frac{1}{2}\langle\Phi, \mathcal{Q}\Phi\rangle + \frac{1}{3}\langle\Phi, \Phi * \Phi\rangle\Big]\,, \tag{324}$$

where the overall normalization K_0 turns out to be infinite. Although the constant K_0 can be absorbed into a rescaling of Ψ, this changes the normalization of $\mathcal{Q}$. We shall instead choose a convenient normalization of $\mathcal{Q}$ and keep the constant K_0 in the action as in eq.(324).

In this VSFT the ansatz was made that any Dp-brane solution takes the factorized form[74]

$$\Phi = \Phi_m \otimes \Phi_g\,, \tag{325}$$

where Φ_g denotes a state obtained by acting with the ghost oscillators on the SL(2,R) invariant vacuum of the ghost CFT, and Φ_m is a state obtained by acting with matter operators on the SL(2,R) invariant vacuum of the matter CFT. Let us denote by $*^g$ and $*^m$ the star product in the ghost and matter sector respectively. Eq.(306) then factorizes as

$$\mathcal{Q}\Phi_g = -\Phi_g *^g \Phi_g\,, \tag{326}$$

and

$$\Phi_m = \Phi_m *^m \Phi_m \,. \tag{327}$$

This last equation is particularly simple: it states that Φ_m is a projector (a projector P in an algebra with product $*$ is an element that satisfies $P * P = P$). The equation for Φ_g appears to be more complicated.

For any string field configuration Φ that satisfies the equation of motion, the action is given by

$$S = -\frac{K_0}{6}\langle \Phi, Q\Phi \rangle \,, \tag{328}$$

and with the ansatz (325) this becomes

$$S = -\frac{K_0}{6}\langle \Phi_g | Q\Phi_g \rangle \langle \Phi_m | \Phi_m \rangle \,, \tag{329}$$

Here the inner products are the BPZ ones for the separate matter and ghost conformal field theories. For any static solution, the action is equal to minus the potential energy. If we are describing a Dp-brane, the action is equal to minus the volume of the brane times the tension of the brane.

To proceed further it is assumed that the ghost part Φ_g is universal for all Dp-brane solutions. Under this assumption the ratio of energies associated with two different D-brane solutions with matter parts Φ'_m and Φ_m respectively, is given by:

$$\frac{E'}{E} = \frac{\langle \Phi'_m | \Phi'_m \rangle_m}{\langle \Phi_m | \Phi_m \rangle_m} \,. \tag{330}$$

Thus the ghost part drops out of this calculation. The inner products in the above right-hand side include brane volume factors, which once removed, give us brane tensions. Equation (330) has allowed some important tests of VSFT. If solutions Φ'_m and Φ_m are available, one can calculate the ratio of tensions of D-branes. Since the ratios are known, one has a test of VSFT. The solutions, as mentioned before, are projectors of the star algebra. The D25-brane solution, for example, can be represented by the sliver state $|\Xi\rangle$, which is the first example of a star-algebra projector that was discovered. The sliver state can be constructed for any conformal field theory (a brief discussion is given in the following subsection). Similarly, Dp-brane solutions can be obtained as modified slivers, and numerical verification that the correct ratio of tensions emerges was obtained.[74] Subsequently, and equipped with a better understanding of the star-algebra, Okuyama[108] was able to demonstrate analytically that the correct ratio of tensions emerges.

In a series of stimulating papers,[107,109] Hata, Kawano, and Moriyama, showed that the relationship $2\pi^2 g^2 T_{25} = 1$ between the D25-brane tension and the string coupling can be tested in VSFT without knowledge of the explicit form of the purely ghost $\mathcal{Q}$. In other words, the normalization of the action, the infinite constant K_0, does not feature in the computation. This is easy to see. The D-brane tension, which is proportional to the value of the action evaluated on the sliver solution, is linearly proportional to K_0. In order to calculate the string coupling, Hata and Kawano proposed to look for the tachyon state on the D-brane; this state should appear as a fluctuation around the sliver solution. With this tachyon state, the string coupling g can be obtained as the coupling for three on-shell tachyons. The effective action for the tachyon fluctuation t would take the form

$$K_0\Big(\alpha\,\frac{1}{2}\,t\,(\partial^2+1)t+\frac{1}{3}\,\beta\,t^3\Big)\,, \tag{331}$$

where α and β are calculable finite constants. The field rescaling $t = T/\sqrt{K_0\alpha}$ brings this action to canonical form

$$\frac{1}{2}\,T\,(\partial^2+1)T+\frac{1}{3}\,\frac{\beta}{\sqrt{K_0\alpha}}\,T^3\,, \tag{332}$$

and the string coupling can be read $g = \beta/\sqrt{K_0\alpha}$. Since $T_{25} \sim K_0$, the relation $2\pi^2 g^2 T_{25} = 1$ does not involve K_0. The original computations, however, did not work out, because the tachyon state had been incorrectly identified.[110] In a remarkable work,[81] Okawa gave a correct identification of the tachyon state and demonstrated that the relation between the string coupling and the brane tension works out correctly. Still both the string coupling and the brane tension are singular.

It is interesting to wonder what features of VSFT that depend on the particular choice of pure ghost operator $\mathcal{Q}$. It appears that a completely regular definition of the spectrum of strings around D-brane solutions may involve $\mathcal{Q}$. Indeed, Okawa has recently demonstrated that the knowledge of $\mathcal{Q}$ is necessary to produce VSFT solutions that give rise to a string coupling and brane tension both of which are finite.[111] The specific form of $\mathcal{Q}$ may also be needed for the calculation of closed string amplitudes using VSFT. It is clear, however, that the choice in (323) is rather special. We remarked earlier that the equation (326) for the ghost part of the solution is not just a projector equation. It turns out, however, that there is a twist of the ghost CFT of (b, c) in which the antighost becomes a field of dimension one and ghost becomes a field of dimension zero. The new CFT has central charge $c = -2$. If $\mathcal{Q}$ is given by (323), the solution of (326) is simply the sliver state of the twisted conformal field theory.[112]

We conclude this subsection with some comments on regularization and the singular aspects of VSFT. Arguments by Gross and Taylor,[68] and by Schnabl (unpublished) indicated that the brane tension associated with VSFT solutions is zero for any finite K_0. Numerical experiments confirm these arguments. A regulation scheme was developed by Gaiotto *et.al*[112] in which K_0 is replaced by $K_0(a)$, and the gauge-fixed kinetic operator of VSFT is made a-dependent in such a way that for for infinite a the pure ghost operator is recovered. The $K_0(a)$ divergence as $a \to \infty$ is determined from the requirement that the D-brane tension is correctly reproduced. The regulated theory appears to be well defined, but universality is lost in the regulation. On the other hand, the analysis of the regulated equations led to the discovery of another special projector of the star algebra: the butterfly state.[112,64,113]

We noted in section 8.1 that after a shift to the tachyon vacuum the open string field theory on a D25-brane becomes a cubic string field theory with kinetic operator $\widetilde{Q}$. This operator is not made solely of ghosts. We would expect, however, that the RSZ theory, if fully correct, is field redefinition equivalent to the theory with $\widetilde{Q}$. If we consider the action (309), a homogeneous field redefinition of the type

$$\widetilde{\Phi} = e^{K}\,\Phi\,, \tag{333}$$

has special properties if K is a ghost number zero Grassmann even operator that satisfies the following relations

$$\begin{aligned} &K(A * B) = (KA) * B + A * (KB)\,,\\ &\langle\, KA, B\,\rangle = -\langle A, KB\rangle\,. \end{aligned} \tag{334}$$

These properties guarantee that the form of the cubic term is unchanged, and that, after the field redefinition, the action takes the form

$$S(\Phi) = -\frac{1}{g^2}\left[\frac{1}{2}\langle\,\Phi\,,\,\widehat{Q}\,\Phi\rangle + \frac{1}{3}\langle\,\Phi\,,\,\Phi * \Phi\rangle\right], \tag{335}$$

where

$$\widehat{Q} = e^{-K}\,\widetilde{Q}\,e^{K}\,. \tag{336}$$

It is a good exercise to verify that equations (334) guarantee that $\widetilde{Q}$ satisfies the properties listed in (62). Therefore the new action is consistent.

The operator $\widetilde{Q}$ is, by construction, regular, while $\widehat{Q}$, which we want to be equal to the VSFT operator $\mathcal{Q}$, should be an infinite constant times a ghost insertion at the string midpoint (the infinite constant is necessary because g is finite). A large class of string reparameterizations that leave the open string midpoint invariant can be constructed with operators K that satisfy the relations (334). A reparameterization in which a finite part of the string is squeezed into an infinitesimal neighborhood of the string midpoint will turn a regular $\widetilde{Q}$ that contains a term linear in the ghost field, into an operator $\widehat{Q}$ whose leading term is precisely a divergent ghost insertion at the string midpoint.[112] This happens because the term linear in the ghost field is the term with an operator of lowest possible dimension, and a squeezing transformation, will transform this negative-dimension operator with an infinite factor. It is thus plausible that a singular squeezing transformation relates the string field theory around the tachyon vacuum to the RSZ theory.

8.4. *Slivers and projection operators*

From the point of view of the RSZ approach to VSFT just discussed, projection operators of the star algebra play a crucial role in the construction of solutions of the theory. Such projection operators may also be useful in understanding solutions in the original Witten theory. Quite a bit of work has been done on constructing and analyzing projectors in the star algebra since the RSZ model was originally proposed. Without going into the technical details, we now briefly review some of the important features of projectors.

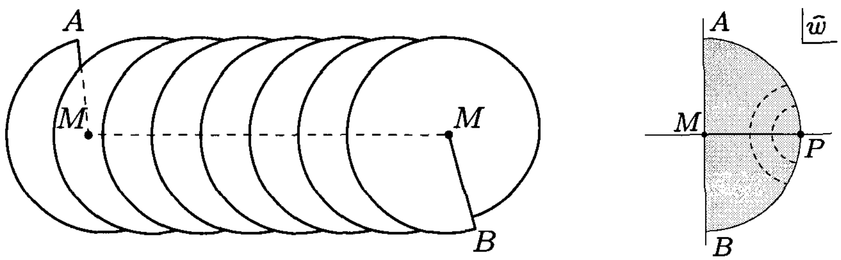


Figure 9. The sliver appears as a cone with infinite excess angle– namely, an infinite helix. The segments AM and BM represent the left-half and the right-half of the string. The local coordinate patch, represented by the shaded half disk shown to the right, must be glued in to form the complete surface.

The first matter projector which was explicitly constructed is the "sliver" state. This state was identified as a conformal field theory surface state by Rastelli and Zwiebach.[65] As such, there is a surface associated with the state: a disk with one puncture on the boundary and a specified local coordinate at the puncture. This conformal field theory picture gives a complete state; it includes both the matter and the ghost part of the state. Moreover, the state can be constructed for any conformal field theory:

$$|\Xi\rangle = \exp\Big(-\frac{1}{3}L_{-2} + \frac{1}{30}L_{-4} - \frac{11}{1890}L_{-6} + \frac{34}{467775}L_{-8} + \cdots\Big)|0\rangle\,. \quad (337)$$

The geometrical picture of the sliver state is shown in figure 9. The full punctured disk is the glued surface obtained by attaching the infinite helix and the coordinate patch, which carries the puncture P. There are many alternative pictures of the sliver.

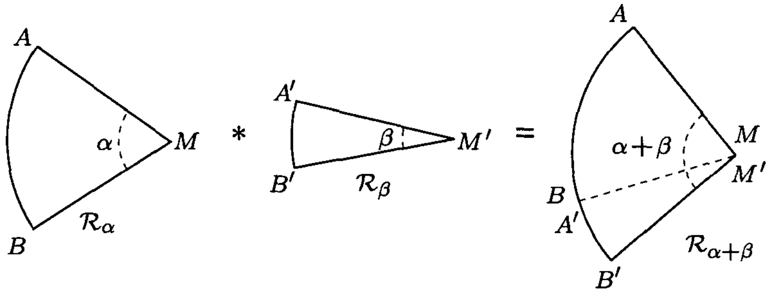


Figure 10. The star multiplication of a sector state with angle α to a sector state with angle β gives a sector state with angle $\alpha+\beta$. Sector states are just another presentation of wedge states.

To understand why the sliver state squares to itself one must have a picture of star multiplication for surface states. A full discussion[114] would take too long, but the rough idea is easily explained. The sliver state is essentially the limit $\lim_{n\to\infty}(|0\rangle)^n$, where multiplication is performed via the star product. A surface state in a BCFT can be viewed (by excising the coordinate patch) as a disk whose boundary has two parts: a part in which the boundary condition that defines the BCFT is imposed, and a part which represents an open string. To star-multiply two surface states, one glues the right-half of the string in the first surface to the left-half of the string in the second surface; the resulting surface is the surface that represents the star product. A particularly simple class of surface states are sector states or wedge states. One such state $\mathcal{R}_\alpha$ is shown to the left of figure 10. The BCFT boundary condition applies to the curved boundary of the sector. The radial segment AM is the left-half of the open string and the radial segment MB is the right-half of the open string. The sector state is defined by the angle α at the string midpoint M. In the figure we show the multiplication of $\mathcal{R}_\alpha$ and $\mathcal{R}_\beta$. The result is a sector state $\mathcal{R}_{\alpha+\beta}$ with total angle $\alpha + \beta$. The sliver state Ξ is the wedge state $\mathcal{R}_\infty$ with infinite angle. It is then clear that the star product of two slivers is still a wedge state of infinite angle, and thus also a sliver. The state obtained in the limit when the angle is equal to zero is in fact the identity state of the star algebra. It is manifestly clear that the product of any surface state with the identity gives the surface state. The identity state can also be written as an exponential of Virasoro operators acting on the vacuum. In fact, as mentioned in section 8.2, a very curious result was found:[63]

$$
\begin{aligned}
|\mathcal{I}\rangle &= \left(\prod_{n=2}^{\infty} \exp\left\{-\frac{2}{2^n} L_{-2^n}\right\}\right) e^{L_{-2}}|0\rangle \\
&= \ldots \exp(-\frac{2}{2^3} L_{-2^3}) \exp(-\frac{2}{2^2} L_{-2^2}) \exp(L_{-2})|0\rangle \,, \qquad (338)
\end{aligned}
$$

with the Virasoro operators of higher level stacking to the left. We thus confirm that the identity is also a Virasoro descendent of the vacuum.

In an independent construction, Kostelecky and Potting[115] constructed a state Ψ_m of the matter sector of the D25-brane BCFT that squared to itself (up to a proportionality constant). The construction used the oscillator language. This matter state takes the form of a squeezed state

$$
|\Psi_m\rangle = \mathcal{N} \exp\left[\frac{1}{2}\, a^\dagger \cdot S \cdot a^\dagger\right] |0\rangle \,. \qquad (339)
$$

By requiring that such a state satisfy the projection equation $\Psi \star \Psi = \Psi$, and by making some further assumptions about the nature of the state, an explicit formula for the matrix S was found in terms of the matrix X from (164).[115] Evidence quickly emerged that the state constructed by these authors is the matter sector of the sliver state, and a proof was given by Okuda.[116]

There are many other projectors that also have a simple picture as surface states.[64,113,117] In these projectors, the open string midpoint approaches (or even coincides with) the boundary of the surface where the boundary condition is applied. One particularly useful projector, which arises in the numerical solution of VSFT, is the so-called *butterfly* state $\mathcal{B}$. This is a very interesting state, whose picture is shown in Figure 11. When one glues two butterfly surfaces in the manner required by star-multiplication, the resulting surface does not appear to be, at first sight, another butterfly. Nevertheless, the resulting surface is conformally equivalent to a butterfly, and this is, in fact, all that is needed in order to have a projector. It has been demonstrated that the butterfly is the state that can be represented as the tensor product $|0\rangle \otimes |0\rangle$, where $|0\rangle$ is the vacuum of the half-string state space.[113] Generally, any state of the form $|a\rangle \otimes |a\rangle$ where $|a\rangle$ is the same state in the left and right half-string Fock spaces is a projector.[67] The butterfly has a remarkably simple expression as a Virasoro descendent of the vacuum

$$|\mathcal{B}\rangle = \exp\Big(-\frac{1}{2}L_{-2}\Big)|0\rangle\,. \tag{340}$$

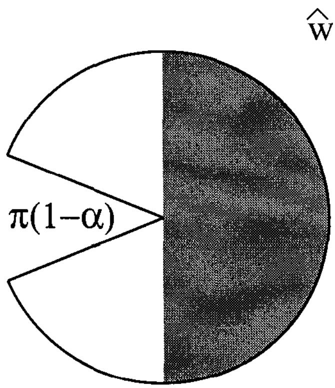


Figure 11. The butterfly state arises in the limit where $\alpha \to 1$ and the angle indicated in the figure vanishes.

Projectors have many properties which are reminiscent of D-branes. This relationship between projection operators and D-branes is familiar from noncommutative field theory, where projectors also play the role of D-brane solitons.[118] This connection becomes quite concrete in the presence of a background B field.[119,120] In the RSZ theory, states that describe an arbitrary but fixed configuration of D-branes are constructed by tensoring the matter projector for the appropriate BCFT with a fixed ghost state that satisfies the ghost equation of motion (326). Particular projectors like the sliver can be constructed which are localized in any number of space-time dimensions, corresponding to the codimension of a D-brane. Under gauge transformations, a rank one projector can be rotated into an orthogonal rank one projector, so that configurations containing multiple branes can be constructed as higher rank projectors formed from the sum of orthogonal rank one projectors.[121,67] This gives a very suggestive picture of how arbitrary D-brane configurations can be constructed in string field theory.

While this picture is quite compelling, however, there are some technical obstacles which make this still a somewhat incomplete story. In the RSZ model, singularities appear due to the separation of the matter and ghost sectors. Moreover, projectors are, in general, somewhat singular states. For example, the matrix S associated to the matter part of the sliver state has eigenvalues of ± 1 for any Dp-brane.[119,113] Such eigenvalues cause these states to be non-normalizable elements of the matter Fock space In the Dirichlet directions, this lack of normalizability occurs because the state is essentially localized to a point and is analogous to a delta function. In the Neumann directions, the singularity manifests as a "breaking" of the strings composing the D-brane, so that the functional describing the projector state is a product of a function of the string configurations on the left and right halves of the string, with no connection mediated through the midpoint. These geometric singularities seem to be generic features of the matter part of any projector, not just the sliver state.[64,113] The singular geometric features of projectors, which can be traced to the fact that the open string midpoint approaches the boundary, makes certain calculations in the RSZ theory somewhat complicated, as all singularities must be regulated. Singularities do not seem to appear in the Witten theory, where the BRST operator and the numerically calculated solutions seem to behave smoothly at the string midpoint. On the other hand, it may be that further study of the projectors will lead to analytic progress on the Witten theory, as discussed in a recent paper by Okawa.[122]

8.5. *Closed strings in open string theory*

We have discussed in earlier sections the fact that open string field theory, formulated on the background of a certain BCFT appears to capture many other open string backgrounds as solutions of the theory. Apart from its singular features, the RSZ theory admits any BCFT as a solution of the theory. One important question remains: Can closed string backgrounds be incorporated in open string field theory? The question can be answered both in the context of OSFT and in the context of the RSZ model. As we will discuss, there is very little concrete evidence as yet that this can be done in any of the two approaches. We therefore ask a simpler question: Can closed string states be seen in open string field theory? The answer here is yes, both in OSFT, and in VSFT (modulo the usual singularities), although so far this has been understood only in certain limited contexts.

As has been known since the earliest days of the subject, closed strings appear as poles in perturbative open string scattering amplitudes. This was demonstrated explicitly for Witten's theory by exhibiting the closed string poles arise in the one-loop 2-point function[123] (although in this calculation, spurious poles also appear which complicate the interpretation). More recently, in a similar calculation the closed string tadpole generated by the D-brane was identified in the one-loop open string 1-point function.[124] While in principle this type of argument can be used to construct all on-shell closed string amplitudes through factorization, it is much less clear how to think of asymptotic or off-shell closed string states in this context. If Witten's theory is well-defined as a quantum theory, it would follow from unitarity that the closed string states should also arise in some natural sense as asymptotic states of the quantum open string field theory. It is currently rather unclear, however, whether, and if so how, this might be realized. There are subtleties in the quantum formulation of the theory which have never completely been resolved,[59,124] although most of the problems of the quantum theory seem to be generated by the closed string tachyon, and may be absent in a supersymmetric theory. Both older SFT literature[125,126] and recent work[112,119,127,128] have suggested ways in which closed strings might be incorporated into the open string field theory, but a definitive resolution of this question is still not available.

In the RSZ model, one description of *on-shell* closed string states is reasonably natural[112,128,129,130] and scattering amplitudes have been computed.[131,132] For each on-shell closed string vertex operator V one can construct a gauge-invariant open string state $\mathcal{O}_V(\Phi)$, where Φ is the open string field, and the gauge invariance is the open string gauge invariance. The world-sheet picture of the state is that of an amputated semi-infinite strip whose edge represents the open strings, the two halves of which are glued and the closed string operator is inserted at the conical singularity. Given a set of gauge invariant operators associated with a set of on-shell closed string vertex operators, the RSZ correlator of the gauge invariant operators appears to give, up to proportionality factors that need regulation, the on-shell closed string amplitude on a surface *without* boundaries. This result uses a nontrivial and unusual decomposition of the moduli space of Riemann surfaces without boundaries.[112] The decomposition, is related to, but distinct from the one used in Witten's theory to cover the moduli space of Riemann surfaces that have at least one boundary. Other decompositions have been discussed by Drukker.[130]

If it were possible to encode *off-shell* physics naturally into open string field theory it would be reasonable to hope that closed string backgrounds could be changed by suitable expectation values of open string fields although this would presumably be a subtle effect in the quantum theory, and difficult to compute explicitly. Attaining a description of the full closed string landscape[133] using quantum OSFT is clearly an optimistic scenario, but it need not be farfetched; it may represent an extension of the AdS/CFT correspondence, in which the CFT side is changed from SYM into the full open string field theory. If, as it may be, it turns out to be that the closed string sector of the theory is encoded in a singular fashion in OSFT, one may be better off directly working with closed string field theory,[27] or with open/closed string field theory.[31] Because of the nonpolynomiality of these theories, it is not known at present if level expansion can be used to extract nonperturbative information. At any rate, it would be useful to have a clear picture of how far one can incorporate closed string physics from the open string point of view. Even if this cannot be realistically achieved in our current models of SFT, understanding the difficulties involved may help us in our search for a better formulation of the theory.

9. Conclusions

The work described in these lectures has brought the understanding of string field theory to a new level. We now have fairly conclusive evidence that open string field theory can successfully describe distinct vacua with very different geometrical properties, which are not related to one another through a marginal deformation. The resulting picture, in which a complicated set of degrees of freedom defined primarily through an algebraic structure, can produce different geometrical backgrounds as different solutions of the equations of motion, represents an important step beyond perturbative string theory. Such a framework is clearly necessary to discuss questions of a cosmological nature in string theory. For such questions, however, one must generalize from the work described here in which the theory describes distinct *open* string backgrounds, to a formalism where different *closed* string backgrounds also appear as solutions of the equations. Ideally, we would like to have a formulation of string/M-theory in which all the currently understood vacua can arise in terms of a single well-defined set of degrees of freedom.

It is not yet clear, however, how far it is possible go towards this goal using the current formulations of string field theory. It may be that the correct lesson to take from the work described here is simply that there *are* nonperturbative formulations in which distinct vacua can be brought together as solutions of a single classical theory, and that one should search for some deeper fundamental algebraic formulation where geometry, and even the dimension of space-time emerge from the fundamental degrees of freedom in the same way that D-brane geometry emerges from the degrees of freedom of Witten's open string field theory. A more conservative scenario, however, might be that we could perhaps use the current framework of string field theory, or some limited refinement thereof, to achieve this goal of providing a universal nonperturbative definition of string theory and M-theory. Following this latter scenario, we propose here a series of questions aimed at continuing the recent developments in open string field theory as far as possible towards this ultimate goal. It is not certain that this research program can be carried to its conclusion, but it will be very interesting to see how far open string field theory can go in reproducing important nonperturbative aspects of string theory.

There are, in our mind, two very important concrete problems related to Witten's string field theory that so far have resisted solution:

1) Finding an analytic description of the tachyonic vacuum. Despite several years of work on this problem, great success with numerical approximations, and some insight from the RSZ vacuum string field theory model, we still have no closed form expression for the string field Φ_0 which represents the tachyon vacuum in Witten's open string field theory. It seems almost unbelievable that there is not some elegant analytic solution to this problem. An analytic solution would almost certainly greatly enhance our understanding of this theory and would lead to other significant advances.
2) Finding certain open string backgrounds as solutions of open string field theory. As discussed in section 7.6, we do not know how to obtain a background with multiple D-branes starting with a background with one D-brane. Nor we know how to obtain the background which represents a D0-brane using the background of a D1-brane with lower energy. It is currently unclear whether the obstacles to finding these vacua are conceptual or technical.

There are other questions that are probably important to the future development of string field theory. These represent, in our opinion, subjects that merit investigation:

1) Is there a regular formulation of VSFT ? Such a version of the theory may have further similarities with BSFT and could turn out to be a complete and flexible formulation of open string field theory.
2) How do closed string backgrounds appear in open string field theory? While OSFT and VSFT appear to give somewhat singular/intractable descriptions of closed string physics, some better understood, or new, version of open string theory might provide a tractable description of closed string physics. Another possibility is that closed string fields are needed in addition to open string fields; this is the case in light-cone open string field theory and in covariant open/closed string field theory.
3) What are the new features of superstring field theory? The status of the tachyon conjectures for the superstring has been reviewed by Ohmori.[134] The large set of symmetries of superstring theory makes them, in many cases, more tractable than bosonic string theories. Nevertheless, as of yet, there is no clear sense in which superstring field theory is simpler than bosonic string field theory.[135] There are also significant conceptual problems that have not allowed a formulation of vacuum superstring field theory.[136]
4) How do we describe time-dependent tachyon dynamics ? String field theory gives clear and concrete evidence for the Sen conjectures. Although we have not studied this subject in the present review, there is much interest in the process by which the tachyon rolls from the unstable critical point down to the tachyon vacuum.[6] In fact, the early attempts to describe the rolling of the tachyon in Witten's string field theory[137,138] appear to be in contradiction with the results that follow from conformal field theory.

It is challenging to imagine a single set of degrees of freedom which could encode, in different phases, all the possible string backgrounds we are familiar with, including those associated with M-theory. In principle, a nonperturbative background-independent formulation of type II string theory should allow one to take the string coupling to infinity in such a way that the fundamental degrees of freedom of the theory remain at some finite point in the configuration space. This would lead to the vacuum associated with M-theory in flat space-time. It would be quite remarkable if this can be achieved in the framework of string field theory. Given the nontrivial relationship between string fields and low-energy effective degrees of freedom, such a result need not be farfetched. If this picture could be successfully implemented, it would give a very satisfying representation of the complicated network of dualities of string and M-theory in terms of a single underlying set of degrees of freedom.

Acknowledgments

We would like to thank all our collaborators: Nathan Berkovits, Erasmo Coletti, Michael Douglas, Ian Ellwood, David Gaiotto, David Gross, Hong Liu, Nicolas Moeller, Greg Moore, Leonardo Rastelli, Ashoke Sen, Jessie Shelton, and Ilya Sigalov, who helped us understand many of the issues described in these lectures. Thanks to the organizers and students of TASI 2001, who provided a stimulating environment for these lectures and much relevant feedback. WT would like to thank the Kavli Institute for Theoretical Physics for hospitality and support during the final stages of writing up these lectures. BZ would like to thank the Harvard University Physics Department for hospitality. This work was supported by the DOE through contract #DE-FC02-94ER40818.

References

1. K. Ohmori, "A review on tachyon condensation in open string field theories," `hep-th/0102085`.
2. P. J. De Smet, "Tachyon condensation: Calculations in string field theory," `hep-th/0109182`.
3. I. Y. Aref'eva, D. M. Belov, A. A. Giryavets, A. S. Koshelev and P. B. Medvedev, "Noncommutative field theories and (super)string field theories," `hep-th/0111208`.
4. L. Bonora, C. Maccaferri, D. Mamone and M. Salizzoni, "Topics in string field theory," `hep-th/0304270`.
5. W. Taylor, "Lectures on D-branes, tachyon condensation and string field theory", School on quantum gravity, Valdivia, Chile, January 2002; `hep-th/0301094`
6. A. Sen, "Rolling tachyon," JHEP **0204**, 048 (2002) `hep-th/0203211`. A. Sen, "Tachyon matter," JHEP **0207**, 065 (2002) `hep-th/0203265`. [For up-to-date references, see the long citation list for the two papers above].
7. I. Bars, "Map of Witten's * to Moyal's *," Phys. Lett. B **517**, 436 (2001) `hep-th/0106157`.
8. I. Bars and Y. Matsuo, "Associativity anomaly in string field theory," Phys. Rev. D **65**, 126006 (2002) `hep-th/0202030`; I. Bars and Y. Matsuo, "Computing in string field theory using the Moyal star product," Phys. Rev.

D **66**, 066003 (2002) hep-th/0204260; I. Bars, I. Kishimoto and Y. Matsuo, "String amplitudes from Moyal string field theory," Phys. Rev. D **67**, 066002 (2003) hep-th/0211131; I. Bars, I. Kishimoto and Y. Matsuo, "Fermionic ghosts in Moyal string field theory," JHEP **0307**, 027 (2003) hep-th/0304005.
9. I. Bars, "MSFT: Moyal star formulation of string field theory," hep-th/0211238.
10. A. A. Gerasimov and S. L. Shatashvili, "On exact tachyon potential in open string field theory," JHEP **0010**, 034 (2000), hep-th/0009103; D. Kutasov, M. Marino and G. Moore, "Some exact results on tachyon condensation in string field theory," JHEP **0010**, 045 (2000), hep-th/0009148; D. Ghoshal and A. Sen, "Normalisation of the background independent open string field JHEP **0011**, 021 (2000), hep-th/0009191; D. Kutasov, M. Marino and G. Moore, D. Kutasov, M. Marino and G. W. Moore, "Remarks on tachyon condensation in superstring field theory,", hep-th/0010108; S. Moriyama and S. Nakamura, "Descent relation of tachyon condensation from boundary string Phys. Lett. B **506**, 161 (2001), hep-th/0011002; A. A. Gerasimov and S. L. Shatashvili, "Stringy Higgs mechanism and the fate of open strings," JHEP **0101**, 019 (2001), hep-th/0011009; I. Y. Aref'eva, A. S. Koshelev, D. M. Belov and P. B. Medvedev, "Tachyon condensation in cubic superstring field theory," Nucl. Phys. B **638**, 3 (2002), hep-th/0011117; P. Kraus and F. Larsen, "Boundary string field theory of the D D-bar system," Phys. Rev. D **63**, 106004 (2001), hep-th/0012198; M. Alishahiha, H. Ita and Y. Oz, "On superconnections and the tachyon effective action," Phys. Lett. B **503**, 181 (2001), hep-th/0012222; G. Chalmers, "Open string decoupling and tachyon condensation," JHEP **0106**, 012 (2001), hep-th/0103056; M. Marino, "On the BV formulation of boundary superstring field theory," JHEP **0106**, 059 (2001), hep-th/0103089; V. Niarchos and N. Prezas, "Boundary superstring field theory," Nucl. Phys. B **619**, 51 (2001), hep-th/0103102; K. S. Viswanathan and Y. Yang, "Tachyon condensation and background independent superstring field theory," Phys. Rev. D **64**, 106007 (2001), hep-th/0104099; M. Alishahiha, "One-loop correction of the tachyon action in boundary superstring field theory," Phys. Lett. B **510**, 285 (2001), hep-th/0104164.
11. Hull, C. M., and P. K. Townsend, "Unity of superstring dualities," Nucl. Phys. **B 438**, 109, 1995; hep-th/9410167.
12. Witten, E. "String Theory Dynamics in Various Dimensions," *Nucl. Phys.* **B 443**; 85 1995; hep-th/9503124.
13. Townsend, P. K., "The eleven-dimensional supermembrane revisited," Phys. Lett. **B 350**, 184, 1995; hep-th/9501068.
14. Hořava, P. and E. Witten, "Heterotic and type I string dynamics from eleven dimensions," Nucl. Phys. **B 460**, 506, 1996; hep-th/9510209.
15. W. Taylor, "M(atrix) theory: matrix quantum mechanics as a fundamental theory," Rev. Mod. Phys. **73**, 419, 2001; hep-th/0101126
16. Aharony, O., S. S. Gubser, J. Maldacena, H. Ooguri, and Y. Oz, "Large N field theories, string theory and gravity," Phys. Rept. **323**, 183, 2000; hep-th/9905111.

17. E. Witten, "D-branes and K-theory," JHEP **9812**, 019 (1998) `hep-th/9810188`. P. Horava, "Type IIA D-branes, K-theory, and matrix theory," Adv. Theor. Math. Phys. **2**, 1373 (1999) `hep-th/9812135`. E. Witten, "Overview of K-theory applied to strings," Int. J. Mod. Phys. A **16**, 693 (2001) `hep-th/0007175`.
18. G. Veneziano, "Construction of a crossing-symmetric, Regge-behaved amplitude for linearly rising trajectories," Nuovo Cim. **57A** 190 (1968).
19. D. J. Gross, A. Neveu, J. Scherk and J. H. Schwarz, "Renormalization and unitarity in the dual resonance model," *Phys. Rev.* **D2** (1970) 697.
20. C. Lovelace, "Pomeron form factors and dual Regge cuts," *Phys. Lett.* **B34** (1971) 500.
21. M. R. Douglas and N. A. Nekrasov, "Noncommutative field theory," Rev. Mod. Phys. **73**, 977 (2001), `hep-th/0106048`.
22. M. Kato and K. Ogawa, "Covariant Quantization Of String Based On BRS Invariance," Nucl. Phys. B **212**, 443 (1983).
23. W. Siegel, Phys. Lett. B **149**, 162 (1984) [Phys. Lett. **151B**, 396 (1985)]. "Covariantly Second Quantized String. 2," Phys. Lett. B **149**, 157 (1984) [Phys. Lett. B **151**, 391 (1985)].
24. W. Siegel and B. Zwiebach, "Gauge String Fields," Nucl. Phys. B **263**, 105 (1986).
25. T. Banks and M. E. Peskin, "Gauge Invariance Of String Fields," Nucl. Phys. B **264**, 513 (1986).
26. E. Witten, "Non-commutative geometry and string field theory," *Nucl. Phys.* **B268** 253, (1986).
27. B. Zwiebach, "Closed string field theory: Quantum action and the B-V master equation," Nucl. Phys. B **390**, 33 (1993) `hep-th/9206084`.
28. M. Saadi and B. Zwiebach, Annals Phys. **192**, 213 (1989).
29. T. Kugo and K. Suehiro, "Nonpolynomial Closed String Field Theory: Action And Its Gauge Invariance," Nucl. Phys. B **337**, 434 (1990).
T. Kugo, H. Kunitomo and K. Suehiro, Phys. Lett. B **226**, 48 (1989).
30. M. Kaku and J. Lykken, "Modular Invariant Closed String Field Theory," Phys. Rev. D **38**, 3067 (1988). M. Kaku, "Geometric Derivation Of String Field Theory From First Principles: Closed Strings And Modular Invariance," Phys. Rev. D **38**, 3052 (1988).
31. B. Zwiebach, "Oriented open-closed string theory revisited," Annals Phys. **267**, 193 (1998) `hep-th/9705241`.
B. Zwiebach, "Quantum Open String Theory With Manifest Closed String Factorization," Phys. Lett. B **256**, 22 (1991).
32. E. Witten, "Interacting Field Theory Of Open Superstrings," Nucl. Phys. B **276**, 291 (1986).
33. C. Wendt, "Scattering amplitudes and contact interactions in Witten's superstring field theory," *Nucl. Phys.* **B314** (1989) 209.
34. J. Greensite and F. R. Klinkhamer, "Superstring amplitudes and contact interactions," *Phys. Lett.* **B304** (1988) 108.
35. I. Y. Arefeva, P. B. Medvedev and A. P. Zubarev, "New Representation For String Field Solves The Consistence Problem For Open Superstring Field," Nucl. Phys. B **341**, 464 (1990).

36. C. R. Preitschopf, C. B. Thorn and S. A. Yost, "Superstring Field Theory," Nucl. Phys. B **337**, 363 (1990).
37. N. Berkovits, "Super-Poincare invariant superstring field theory," *Nucl. Phys.* **B450** (1995), 90, hep-th/9503099; "Review of open superstring field theory," hep-th/0105230; "The Ramond sector of open superstring field theory," JHEP **0111**, 047 (2001), hep-th/0109100.
38. A. Sen and B. Zwiebach, "Tachyon condensation in string field theory," JHEP **0003**, 002 (2000), hep-th/9912249.
39. A. Sen, "Universality of the tachyon potential," JHEP **9912**, 027 (1999), hep-th/9911116.
40. J. Polchinski, "TASI Lectures on D-branes," hep-th/9611050.
41. W. Taylor, "Trieste lectures on D-branes, gauge theory and M(atrices)," hep-th/9801182.
42. M. J. Duff, R. R. Khuri and J. X. Lu, "String solitons," Phys. Rept. **259**, 213 (1995), hep-th/9412184.
43. J. Polchinski, "Dirichlet-Branes and Ramond-Ramond Charges," *Phys. Rev. Lett.* **75** (1995), 4724, hep-th/9510017.
44. R. G. Leigh, "Dirac-Born-Infeld action from Dirichlet sigma model," *Mod. Phys. Lett.* **A4** (1989), 2767.
45. E. Witten, "Bound States of Strings and p-Branes," *Nucl. Phys.* **B460** (1996), 335, hep-th/9510135.
46. A. A. Tseytlin, "Born-Infeld action, supersymmetry and string theory," hep-th/9908105; see also the following for some recent developments and further references: P. Koerber and A. Sevrin, "The non-abelian D-brane effective action through order alpha'**4," JHEP **0210**, 046 (2002), hep-th/0208044; S. Stieberger and T. R. Taylor, "Non-Abelian Born-Infeld action and type I - heterotic duality. II: Nonrenormalization theorems," Nucl. Phys. B **648**, 3 (2003), hep-th/0209064; D. T. Grasso, "Higher order contributions to the effective action of N = 4 super Yang-Mills," JHEP **0211**, 012 (2002), hep-th/0210146.
47. W. Taylor and M. Van Raamsdonk, "Multiple D0-branes in weakly curved backgrounds," *Phys. Lett.* **B558** (1999), 63, hep-th/9904095; "Multiple Dp-branes in weak background fields," *Nucl. Phys.* **B573**, 703 (2000), hep-th/9910052.
48. R. C. Myers, "Dielectric-branes," *JHEP* **9912** (1999), 022, hep-th/9910053.
49. Polchinski, J. *String theory* (Cambridge University Press, Cambridge, England, 1998).
50. W. Taylor, "D-brane Field Theory on Compact Spaces," *Phys. Lett.* **B394** (1997), 283; hep-th/9611042.
51. T. Banks and L. Susskind, "Brane - Antibrane Forces," hep-th/9511194.
52. A. Hashimoto and W. Taylor, "Fluctuation Spectra of Tilted and Intersecting D-branes from the Born-Infeld Action," *Nucl. Phys.* **B503** (1997), 193-219; hep-th/9703217.
53. E. Gava, K. S. Narain and M. H. Sarmadi, "On the bound states of p- and (p+2),-branes," Nucl. Phys. B **504**, 214 (1997), hep-th/9704006.
54. A. V. Morosov, *"Classical Decay of a Non-supersymmetric Configuration of Two D-branes"*, *Phys. Lett.* **B433** (1998), 291, hep-th/9803110.

55. K. Hashimoto and S. Nagaoka, *"Recombination of Intersecting D-branes by Local Tachyon Condensation,"* JHEP **0306**, 034 (2003), hep-th/0303204.
56. K. Bardakci and M. B. Halpern, "Explicit spontaneous breakdown in a dual model," *Phys. Rev.* **D10** (1974) 4230; K. Bardakci, "Spontaneous symmetry breaking in the standard dual string model," *Nucl. Phys.* **B133** (1978), 297.
57. V. A. Kostelecky and S. Samuel, "On a nonperturbative vacuum for the open bosonic string," *Nucl. Phys.* **B336** (1990), 263-296.
58. A. Leclair, M. E. Peskin and C. R. Preitschopf, "String field theory on the conformal plane (I)," *Nucl. Phys.* **B317** (1989), 411-463.
59. C. Thorn, "String field theory," *Phys. Rep.* **175** (1989), 1.
60. M. R. Gaberdiel and B. Zwiebach, "Tensor constructions of open string theories 1., 2.," *Nucl. Phys.* **B505** (1997), 569, hep-th/9705038; *Phys. Lett.* **B410** (1997), 151, hep-th/9707051.
61. Green, M. B., J. H. Schwarz, and E. Witten, *Superstring theory* (Cambridge University Press, Cambridge, England, 1987).
62. D. J. Gross and A. Jevicki, "Operator formulation of interacting string field theory (I), (II)," *Nucl. Phys.* **B283** (1987), 1; *Nucl. Phys.* **B287** (1987), 225.
63. I. Ellwood, B. Feng, Y. H. He and N. Moeller, "The identity string field and the tachyon vacuum," JHEP **0107**, 016 (2001), hep-th/0105024.
64. M. Schnabl, "Wedge states in string field theory," JHEP **0301**, 004 (2003), hep-th/0201095; "Anomalous reparametrizations and butterfly states in string field theory," Nucl. Phys. B **649**, 101 (2003), hep-th/0202139.
65. L. Rastelli and B. Zwiebach, "Tachyon potentials, star products and universality," JHEP **0109**, 038 (2001), hep-th/0006240.
66. J. Bordes, Chan H.-M., L. Nellen, Tsou S.-T., "Half-string oscillator approach to string field theory," *Nucl. Phys.* **B351** (1991) 441; A. Abdurrahman, F. Anton and J. Bordes, "Half string oscillator approach to string field theory (ghost sector 1, 2)," *Nucl. Phys.* **B397** (1993) 260; *Nucl. Phys.* **B411** (1994) 693; A. Abdurrahman and J. Bordes, "The relationship between the comma theory and Witten's string field theory," *Phys. Rev.* **D 58**:086003 (1998); L. Rastelli, A. Sen and B. Zwiebach, "Half-strings, projectors, and multiple D-branes in vacuum string field theory," hep-th/0105058; T. Kawano and K. Okuyama, "Open string fields as matrices," hep-th/0105129; K. Furuuchi and K. Okuyama, "Comma vertex and string field algebra," JHEP **0109**, 035 (2001) hep-th/0107101; N. Moeller, "Some exact results on the matter star-product in the half-string formalism," JHEP **0201**, 019 (2002) hep-th/0110204; T. G. Erler, "A fresh look at midpoint singularities in the algebra of string fields," hep-th/0304044; E. Fuchs, M. Kroyter and A. Marcus, "Continuous half-string representation of string field theory," hep-th/0307148.
67. D. J. Gross and W. Taylor, "Split string field theory I," JHEP **0108**, 009 (2001), hep-th/0105059.
68. D. J. Gross and W. Taylor, "Split string field theory II," JHEP **0108**, 010 (2001), hep-th/0106036.
69. E. Cremmer, A. Schwimmer and C. Thorn, "The vertex function in Witten's formulation of string field theory" *Phys. Lett.* **B179** 57 (1986).

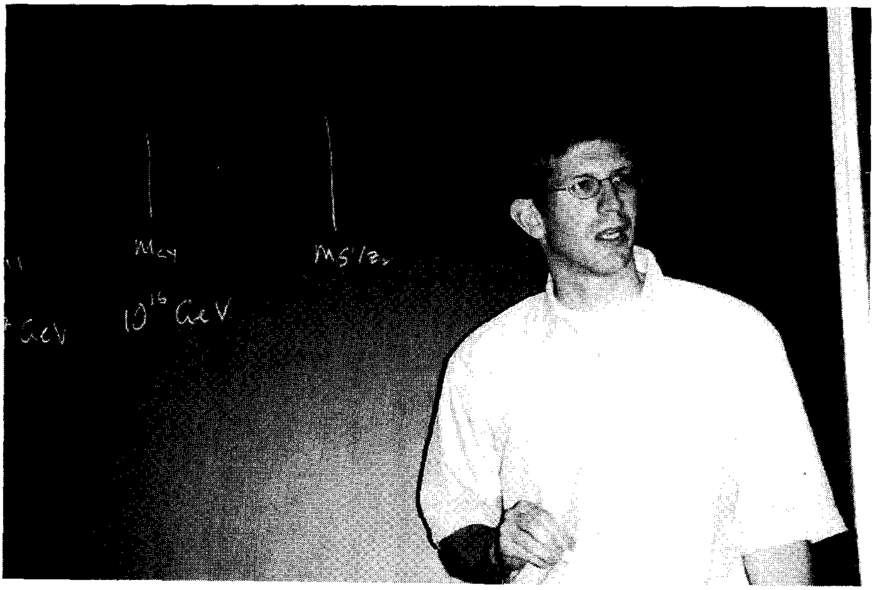
70. S. Samuel, "The physical and ghost vertices in Witten's string field theory," *Phys. Lett.* **B181** 255 (1986).
71. N. Ohta, "Covariant interacting string field theory in the Fock space representation," *Phys. Rev.* **D34** (1986), 3785; *Phys. Rev.* **D35** (1987), 2627 (E).
72. J. Shelton, unpublished.
73. L. Rastelli, A. Sen and B. Zwiebach, "Star algebra spectroscopy," JHEP **0203**, 029 (2002) hep-th/0111281; M. R. Douglas, H. Liu, G. Moore and B. Zwiebach, "Open string star as a continuous Moyal product," JHEP **0204**, 022 (2002) hep-th/0202087; B. Feng, Y. H. He and N. Moeller, "The spectrum of the Neumann matrix with zero modes," JHEP **0204**, 038 (2002) hep-th/0202176; B. Feng, Y. H. He and N. Moeller, "Zeeman spectroscopy of the star algebra," JHEP **0205**, 041 (2002) hep-th/0203175; B. Chen and F. L. Lin, "Star spectroscopy in the constant B field background," Nucl. Phys. B **637**, 199 (2002) hep-th/0203204; I. Y. Arefeva and A. A. Giryavets, "Open superstring star as a continuous Moyal product," JHEP **0212**, 074 (2002) hep-th/0204239; T. G. Erler, "Moyal formulation of Witten's star product in the fermionic ghost sector," hep-th/0205107; C. Maccaferri and D. Mamone, "Star democracy in open string field theory," JHEP **0309**, 049 (2003) hep-th/0306252.
74. L. Rastelli, A. Sen and B. Zwiebach, "Classical solutions in string field theory around the tachyon vacuum," Adv. Theor. Math. Phys. **5**, 393 (2002), hep-th/0102112.
75. S. B. Giddings and E. J. Martinec, "Conformal Geometry and String Field Theory," *Nucl. Phys.* **B278**, 91 (1986).
76. S. B. Giddings, E. J. Martinec and E. Witten, "Modular Invariance In String Field Theory," Phys. Lett. B **176**, 362 (1986).
77. B. Zwiebach, "A Proof That Witten's Open String Theory Gives A Single Cover Of Moduli Space," Commun. Math. Phys. **142**, 193 (1991).
78. I. Y. Aref'eva, A. S. Koshelev, D. M. Belov and P. B. Medvedev, "Tachyon condensation in cubic superstring field theory," Nucl. Phys. B **638**, 3 (2002), hep-th/0011117; G. Bandelloni and S. Lazzarini, "The geometry of W3 algebra: A twofold way for the rebirth of a symmetry," Nucl. Phys. B **594**, 477 (2001), hep-th/0011208.
79. A. Sen, unpublished.
80. N. Berkovits, A. Sen and B. Zwiebach, "Tachyon condensation in superstring field theory," Nucl. Phys. B **587**, 147 (2000), hep-th/0002211;
81. Y. Okawa, "Open string states and D-brane tension from vacuum string field theory," JHEP **0207**, 003 (2002), hep-th/0204012.
82. E. Coletti, I. Sigalov and W. Taylor, "Abelian and nonabelian vector field effective actions from string field theory," JHEP **0309**, 050 (2003) hep-th/0306041.
83. B. Zwiebach, "Trimming the tachyon string field with SU(1,1)," hep-th/0010190.
84. N. Moeller and W. Taylor, "Level truncation and the tachyon in open bosonic string field theory," Nucl. Phys. B **583**, 105 (2000), hep-th/0002237.

85. D. Gaiotto and L. Rastelli, "Progress in open string field theory," Presentation by L. Rastelli at Strings 2002, Cambridge, England; http://www.damtp.cam.ac.uk/strings02/avt/rastelli/.
86. D. Gaiotto and L. Rastelli, "Experimental string field theory," hep-th/0211012.
87. W. Taylor, "Perturbative diagrams in string field theory," hep-th/0207132.
88. M. Beccaria and C. Rampino, "Level truncation and the quartic tachyon coupling," JHEP **0310**, 047 (2003) hep-th/0308059.
89. J. Ellis, M. Karliner and W. Taylor, unpublished.
90. W. Taylor, "A Perturbative Analysis of Tachyon Condensation," hep-th/0208149.
91. I. Ellwood and W. Taylor, "Gauge invariance and tachyon condensation in open string field theory," hep-th/0105156.
92. N. Moeller, A. Sen and B. Zwiebach, "D-branes as tachyon lumps in string field theory," JHEP **0008**, 039 (2000), hep-th/0005036.
93. R. de Mello Koch, A. Jevicki, M. Mihailescu and R. Tatar, "Lumps and p-branes in open string field theory," Phys. Lett. B **482**, 249 (2000), hep-th/0003031.
94. J. A. Harvey and P. Kraus, "D-branes as unstable lumps in bosonic open string field theory," JHEP **0004**, 012 (2000), hep-th/0002117;
95. B. Zwiebach, "A solvable toy model for tachyon condensation in string field theory," JHEP **0009**, 028 (2000), hep-th/0008227.
96. J. A. Minahan and B. Zwiebach, "Field theory models for tachyon and gauge field string dynamics," JHEP **0009**, 029 (2000) hep-th/0008231.
97. I. Ellwood and W. Taylor, unpublished.
98. A. Sen and B. Zwiebach, "Large marginal deformations in string field theory," JHEP **0010**, 009 (2000), hep-th/0007153.
99. J. Kluson, "Marginal deformations in the open bosonic string field theory for N D0-branes," hep-th/0203089; "Exact solutions in open bosonic string field theory and marginal deformation in CFT," hep-th/0209255.
100. L. Rastelli, A. Sen and B. Zwiebach, "String field theory around the tachyon vacuum," Adv. Theor. Math. Phys. **5**, 353 (2002), hep-th/0012251.
101. A. Sen and B. Zwiebach, "A Proof of local background independence of classical closed string field theory," Nucl. Phys. B **414**, 649 (1994) hep-th/9307088.
102. T. Kawano and T. Takahashi, "Open string field theory on noncommutative space," Prog. Theor. Phys. **104**, 459 (2000) hep-th/9912274.
103. M. Fisk and M. Srednicki, "Magnetic string fields," Phys. Lett. B **313**, 308 (1989).
104. I. Ellwood and W. Taylor, "Open string field theory without open strings," Phys. Lett. B **512**, 181 (2001), hep-th/0103085.
105. J. A. Minahan, "Quantum corrections in p-adic string theory," hep-th/0105312.
106. G. T. Horowitz, J. Lykken, R. Rohm and A. Strominger, "A Purely Cubic Action For String Field Theory," Phys. Rev. Lett. **57**, 283 (1986).
107. H. Hata and T. Kawano, "Open string states around a classical solution in vacuum string field theory," JHEP **0111**, 038 (2001), hep-th/0108150.

108. K. Okuyama, "Ratio of tensions from vacuum string field theory," JHEP **0203**, 050 (2002) hep-th/0201136. K. Okuyama, "Ghost kinetic operator of vacuum string field theory," JHEP **0201**, 027 (2002), hep-th/0201015.
109. H. Hata and S. Moriyama, "Observables as twist anomaly in vacuum string field theory," JHEP **0201**, 042 (2002), hep-th/0111034.
110. L. Rastelli, A. Sen and B. Zwiebach, "A note on a proposal for the tachyon state in vacuum string field theory," JHEP **0202**, 034 (2002), hep-th/0111153. H. Hata, S. Moriyama and S. Teraguchi, "Exact results on twist anomaly," JHEP **0202**, 036 (2002), hep-th/0201177.
111. Y. Okawa, "Some exact computations on the twisted butterfly state in string field theory," hep-th/0310264.
112. D. Gaiotto, L. Rastelli, A. Sen and B. Zwiebach, "Ghost structure and closed strings in vacuum string field theory," hep-th/0111129.
113. D. Gaiotto, L. Rastelli, A. Sen and B. Zwiebach, "Star algebra projectors," JHEP **0204**, 060 (2002), hep-th/0202151.
114. L. Rastelli, A. Sen and B. Zwiebach, "Boundary CFT construction of D-branes in vacuum string field theory," JHEP **0111**, 045 (2001) hep-th/0105168.
115. V. A. Kostelecky and R. Potting, "Analytical construction of a nonperturbative vacuum for the open bosonic string," Phys. Rev. D **63**, 046007 (2001), hep-th/0008252.
116. T. Okuda, "The equality of solutions in vacuum string field theory," Nucl. Phys. B **641**, 393 (2002) hep-th/0201149.
117. E. Fuchs, M. Kroyter and A. Marcus, JHEP **0209**, 022 (2002) hep-th/0207001.
118. R. Gopakumar, S. Minwalla and A. Strominger, "Noncommutative solitons," JHEP **0005**, 020 (2000), hep-th/0003160.
119. G. Moore and W. Taylor, "The singular geometry of the sliver," JHEP **0201**, 004 (2002), hep-th/0111069.
120. L. Bonora, D. Mamone and M. Salizzoni, Nucl. Phys. B **630**, 163 (2002) hep-th/0201060; L. Bonora, D. Mamone and M. Salizzoni, JHEP **0204**, 020 (2002) hep-th/0203188; B. Chen and F. L. Lin, Phys. Rev. D **66**, 126001 (2002) hep-th/0204233; L. Bonora, D. Mamone and M. Salizzoni, JHEP **0301**, 013 (2003) hep-th/0207044.
121. L. Rastelli, A. Sen and B. Zwiebach, "Half strings, projectors, and multiple D-branes in vacuum string field theory," JHEP **0111**, 035 (2001), hep-th/0105058.
122. Y. Okawa, "Solving Witten's string field theory using the butterfly state," to appear.
123. D. Z. Freedman, S. B. Giddings, J. A. Shapiro and C. B. Thorn, "The Nonplanar One Loop Amplitude In Witten's String Field Theory," Nucl. Phys. B **298**, 253 (1988).
124. I. Ellwood, J. Shelton and W. Taylor, "Tadpoles and closed string backgrounds in open string field theory," JHEP **0307**, 059 (2003) hep-th/0304259.
125. A. Strominger, "Closed strings in open string field theory," *Phys. Rev. Lett.* **58** 629 (1987).

126. J. A. Shapiro and C. B. Thorn, "BRST invariant transitions between open and closed strings," *Phys. Rev.* **D36** 432 (1987); "closed string-open string transitions in Witten's string field theory," Phys. Lett. B **194**, 43 (1987).
127. A. A. Gerasimov and S. L. Shatashvili, "Stringy Higgs mechanism and the fate of open strings," JHEP **0101**, 019 (2001), `hep-th/0011009`. S. L. Shatashvili, "On field theory of open strings, tachyon condensation and closed strings," `hep-th/0105076`.
128. A. Hashimoto and N. Itzhaki, "Observables of string field theory," JHEP **0201**, 028 (2002), `hep-th/0111092`.
129. J. Ambjorn and R. A. Janik, "Interpolating between open and closed strings - a BSFT approach," Phys. Lett. B **538**, 189 (2002) `hep-th/0203185`.
130. N. Drukker, "Closed string amplitudes from gauge fixed string field theory," Phys. Rev. D **67**, 126004 (2003) `hep-th/0207266`.
131. M. Alishahiha and M. R. Garousi, "Gauge invariant operators and closed string scattering in open string field theory," Phys. Lett. B **536**, 129 (2002) `hep-th/0201249`.
132. T. Takahashi and S. Zeze, "Closed string amplitudes in open string field theory," JHEP **0308**, 020 (2003) `hep-th/0307173`.
133. L. Susskind, "The anthropic landscape of string theory," `hep-th/0302219`.
134. K. Ohmori, "Level-expansion analysis in NS superstring field theory revisited," `hep-th/0305103`.
135. P. De Smet and J. Raeymaekers, "Level four approximation to the tachyon potential in superstring field theory," JHEP **0005**, 051 (2000), `hep-th/0003220`; A. Iqbal and A. Naqvi, "Tachyon condensation on a non-BPS D-brane,", `hep-th/0004015`; P. De Smet and J. Raeymaekers, "The tachyon potential in Witten's superstring field theory," JHEP **0008**, 020 (2000), `hep-th/0004112`;
136. M. Marino and R. Schiappa, J. Math. Phys. **44**, 156 (2003) `hep-th/0112231`. J. Kluson, JHEP **0107**, 039 (2001) `hep-th/0106107`. I. Y. Arefeva, A. A. Giryavets and P. B. Medvedev, Phys. Lett. B **532**, 291 (2002) `hep-th/0112214`. K. Ohmori, JHEP **0204**, 059 (2002) `hep-th/0204138`. K. Ohmori, Nucl. Phys. B **648**, 94 (2003) `hep-th/0208009`; A. Koshelev, "Solutions of vacuum superstring field theory," `hep-th/0212055`.
137. N. Moeller and B. Zwiebach, "Dynamics with infinitely many time derivatives and rolling tachyons," JHEP **0210**, 034 (2002) `hep-th/0207107`.
138. M. Fujita and H. Hata, "Time dependent solution in cubic string field theory," JHEP **0305**, 043 (2003) `hep-th/0304163`.

This page intentionally left blank



DANIEL WALDRAM

This page intentionally left blank

INTRODUCTION TO MODEL BUILDING IN HETEROTIC M-THEORY

DANIEL WALDRAM
The Blackett Laboratory
Imperial College, London
Prince Consort Road, London SW7 2AZ, UK
E-mail: d.waldram@imperial.ac.uk

These lectures are an introduction to building semi-realistic models in heterotic M-theory and a continuation of those given by Burt Ovrut. First the phenomenological constraints on the compactification of supersymmetry, low-energy GUT groups and three families of charged matter are summarized. This involves a discussion of Calabi–Yau manifolds and, in particular, supersymmetric gauge field backgrounds as a special class of holomorphic bundles. We review the construction of supersymmetric bundles in the very simple example of a two-torus and show how this can be naturally understood in terms of the "Fourier–Mukai transform", which is really just the action of T-duality. We then review the structure of elliptically fibred Calabi–Yau manifolds X. Supersymmetric bundles can be constructed on X by extending the transform fibre by fibre. This is the spectral cover construction. Finally, we give an explicit example with GUT group $SU(5)$ and three families of matter. Although discussed in the context of M-theory, these constructions work equally well for the heterotic string and have applications to D-brane physics.

1. Introduction

These lectures are a continuation of those given by Burt Ovrut on heterotic M-theory.[1] The goal is to find M-theory braneworlds that have

> *explicit* four-dimensional models with $\mathcal{N} = 1$ supersymmetry, grand unified (GUT) gauge groups (*e.g.* $SU(5)$, $SO(10)$) and three families of matter.

Although the discussion will be in the context of the strongly coupled M-theory limit of the heterotic string,[2] almost all of the construction will be equally applicable to the weakly coupled limit. The only difference is the inclusion of M5-branes which allows a little more freedom in satisfying the model constraints. Thus much of the discussion about supersymmetry and three-family conditions is very familiar.[3]

One of the most difficult ingredients in building semi-realistic models is finding suitable gauge bundles for the E_8 gauge fields. This refers to the particular supersymmetric configuration of magnetic fields on the compact part of the spacetime. They must be chosen to leave a GUT gauge group unbroken as well as the right low-energy four-dimensional degrees of freedom. The conditions of supersymmetry translate into a very specific mathematical condition. While this provides an existence argument which allows one to get around actually solving non-linear differential equations, nonetheless it is still hard to find examples. The main new tool here will be to use the so-called "Fourier–Mukai" or "spectral cover" construction. This technique was first devised to describe the heterotic duals of F-theory compactifications.[4,5,6]

Necessarily there is a certain amount of mathematics which goes into describing the geometry of the compactification and of the supersymmetric gauge field configurations. These lectures aim to describe some of these details. The goal is to stress the physical meaning rather than be very precise about the mathematics, though hopefully also to give enough background to provide a starting point for performing the calculations. Some of this material is summarized in the appendices. Since supersymmetric geometries and gauge field configurations appear in many different contexts in string theory, many of these techniques have wider applicability, most directly in F-theory and D-branes configurations. In particular, as we will see, the Fourier–Mukai transform has a direct interpretation as the action of T-duality for wrapped D-branes.

It is also, perhaps, worth mentioning what we will not cover. The goal is to construct a simple GUT model so we will not be directly interested here in the other phenomenological details of the models, such as supersymmetry breaking, the Higgs sector or constraints from proton decay. How these issues can and have been addressed will be mentioned at the end of the lectures. In addition, we will also confine ourselves solely to heterotic M-theory braneworlds. It is worth pointing out that there are also very promising ways to construct semi-realistic braneworld models from intersecting D-brane models in type II or type I theories.[7]

The lectures will be divided as follows. In the next section we will review the heterotic M-theory set up as described by Burt Ovrut.[1] Section 3 discusses the constraints imposed by supersymmetry. Specifically we show that the geometry has the form $S^1/\mathbb{Z}_2 \times X$ where X is a Calabi–Yau manifold; that the gauge fields satisfy the Hermitian Yang–Mills equation, which is mathematically equivalent to a "semi-stable holonomorphic" vector bundle; and that the fivebranes wrap holomorphic curves in X. Section 4 discusses a topological constraint that arises from anomaly cancellation and the condition implied by the requirement that there are only three families of charged matter. Section 5 discusses how to construct suitable gauge bundles in the very simple example where X is a two-torus, and introduces the Fourier–Mukai transform and its interpretation in terms of T-duality. In section 6 we analyze the structure of elliptically fibred Calabi–Yau manifolds X. These have two-torus fibers and admit a fibrewise Fourier–Mukai transform. Section 7 constructs generic supersymmetric bundles on X and then gives an explicit example of a GUT model with $SU(5)$ gauge group. We end by discussing possible extensions and in particular how one can build models with standard-model $SU(3) \times SU(2) \times U(1)$ interactions.

2. Review of heterotic M-theory

We start by reviewing how one obtains braneworld models from the strongly coupled M-theory limit of the $E_8 \times E_8$ heterotic string.[8,9,10] Much of this material has already been discussed in Burt Ovrut's lectures,[1] but it is useful to repeat it here for definiteness.

2.1. *Hořava–Witten theory*

The description of the strongly coupled limit of the $E_8 \times E_8$ heterotic string was first discussed Hořava and Witten.[2] They argued that as the dilaton became large, as for the type IIA string, the ten-dimensional string theory grows an extra dimension. The full form of the new eleven-dimensional M-theory dual is not known. However, at low energies, there is an effective field theory description as an expansion in the eleven-dimensional Planck scale m_{11}. To leading order, the theory is simply supergravity.

In addition, the eleven-dimenisonal M-theory is compactified on, not a circle as in the type IIA case, but the orbifold interval $S^1/\mathbb{Z}_2$. This can be defined as a quotient of the real line $y \in \mathbb{R}$ by $y + 2\pi\rho \sim y$ (to define the circle S^1 of radius ρ) and $y \sim -y$ (for the $\mathbb{Z}_2$ quotient). The action of $\mathbb{Z}_2$ indentifies the two halves of the circle, namely $0 \leq y \leq \pi\rho$ with $0 \geq y \geq -\pi\rho$, and has two fixed points at $y = 0$ and $y = \pi\rho$. The resulting space is simply the interval $0 \leq y \leq \pi\rho$ with fixed points at either end, as shown in Fig. 1. Thus the full eleven-dimensional space M has two ten-dimensional boundaries ∂M_1 and ∂M_2 at the fixed points of the $S^1/\mathbb{Z}_2$ interval.

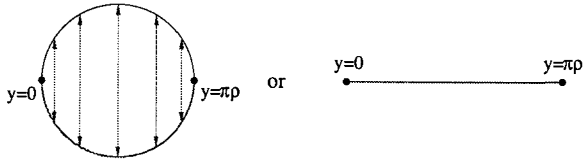


Figure 1. The $S^1/\mathbb{Z}_2$ orbifold

The supergravity multiplet for the low-energy bulk fields consists of a metric g_{IJ}, a three-form potential C_{IJK} with a field strength $G = \mathrm{d}C$ and a gravitino ψ_I. The low-energy action for the bosonic fields has the standard form

$$
\begin{aligned}
S_{\text{bulk}} = -\frac{1}{2\kappa^2} \int_M \sqrt{-g}\Big(R + \tfrac{1}{24}G^2 \\
+ \tfrac{\sqrt{2}}{1728}\epsilon^{I_1\dots I_{11}} C_{I_1I_2I_3} G_{I_4I_5I_6I_7} G_{I_8I_9I_{10}I_{11}}\Big),
\end{aligned}
\tag{1}
$$

where $\epsilon^{I_1\dots I_{11}}$ is the volume form and $\kappa^2 \sim m_{11}^{-9}$ is the gravitational coupling constant. This action is written in the "upstairs" formulation, where the integration is taken over the eleven-dimensional space $M_{10} \times S^1$, including the whole S^1 rather than just the interval $S^1/\mathbb{Z}_2$, and one then imposes symmetry conditions on the fields under the action of $\mathbb{Z}_2$. In particular, for the gravitinos, for the action to be invariant one requires

$$
\begin{aligned}
\psi_M(x^M, -y) &= \Gamma^{11}\psi_M(x^M, y), \\
\psi_{11}(x^M, -y) &= -\Gamma^{11}\psi_{11}(x^M, y),
\end{aligned}
\tag{2}
$$

where $y = x^{11}$ parametrises the S^1 direction and x^M the directions in M_{10}, while Γ^{11} is the gamma matrix in the circle direction. (Somewhat eccentrically our coordinates are thus $x^I = (x^M, x^{11})$ with M taking values in $0, \dots, 9$.) This relation has an important consequence: on the fixed planes $y = 0, \pi\rho$ the surviving components of the gravitino are chiral (*e.g.* $\psi_M(x^M, 0) = \Gamma^{11}\psi_M(x^M, 0)$). Chiral spinors in ten dimensions lead to gravitational amomalies, thus, as stands the pure bulk theory is not a consistent low-energy limit of a quantum theory. Hořava and Witten showed that there is a unique way to make the theory anomaly-free by, first, including additional fields restricted to the fixed planes, and, second, coupling these fields to the field strength G, in an analogy of the Green–Schwarz mechanism.

One finds that each fixed plane ∂M_i carries a ten-dimensional E_8 gauge supermultiplet with connection A_i with field strength F_i and gaugino χ^i.

The bosonic action reads

$$S_{\text{boundary}} = -\frac{1}{4\lambda^2}\int_{\partial M_1}\sqrt{-g}\left(\operatorname{tr}F_1^2 - \tfrac{1}{2}\operatorname{tr}R^2\right) \\ -\frac{1}{4\lambda^2}\int_{\partial M_2}\sqrt{-g}\left(\operatorname{tr}F_2^2 - \tfrac{1}{2}\operatorname{tr}R^2\right), \tag{3}$$

where g is the pull-back of the eleven-dimensional metric and R is the Rieman tensor viewed as a $SO(10)$ field strength for the spin connection. The traces are given by $\frac{1}{60}\operatorname{tr}_{\text{adjoint}}$ for E_8[a] and the trace in the vector representation of $SO(10)$ for R. The gauge coupling $\lambda^2 \sim m_{11}^{-6}$ is determined in terms of κ. Note that the fields on the fixed planes are rather like the additional twisted fields that appear stuck at orbifold fixed points in string theory. If we had the analogous fundamental formulation of the eleven-dimensional theory we would expect to be able to derive them directly. Since we do not, we simply infer their existence from anomaly cancellation.

The second correction is the coupling to G. The fixed planes provide magnetic sources for the Bianchi identity, so that

$$\mathrm{d}G = -k\Big\{2J_1\wedge\delta(y) + 2J_2\wedge\delta(y-\pi\rho) \\ + [\delta(D_1)+\delta(D_1')] + \cdots + [\delta(D_n)+\delta(D_n')]\Big\}, \tag{4}$$

where

$$J_i = \frac{1}{16\pi^2}\left(\operatorname{tr}F_i\wedge F_i - \tfrac{1}{2}\operatorname{tr}R\wedge R\right), \tag{5}$$

are the sources from the two fixed planes. Note that M-theory fivebranes are also magnetic sources for G. Consequently, to be as general as possible we have also included sources for a set of fivebranes localized on six-dimensional surfaces D_i in the eleven-dimensional bulk (defining five-form delta-functions $\delta(D_i)$). Since we are in the "upstairs" formulation, we must also include their mirror images under the $\mathbb{Z}_2$ transformation localized on D_i'. The constant $k \sim m_{11}^{-3}$ is again determined in terms of κ. Thus just as the boundary fields represent a correction to the action suppressed by m_{11}^{-3}, so the sources for G are corrections to $\mathrm{d}G = 0$ at the same order.

The fivebranes are also gravitation sources for the bulk fields. The bosonic fields on a single fivebrane are a set of scalars x^I describing its

[a]Note that this means[11] that for any (regular) $SU(n)$ sub-group of E_8, $\frac{1}{60}\operatorname{tr}_{\text{adjoint}}$ equals the trace in the $\mathbf{n}+\bar{\mathbf{n}}$ representation of $SU(n)$.

embedding $D_i \to M_{10} \times S^1/\mathbb{Z}_2$ and a self-dual threeform field strength h. If we set h to zero, the effective action is given by

$$\begin{aligned} S_{\text{fivebranes}} = &- T_5 \left(\int_{D_1} \sqrt{-h} + \int_{D_1'} \sqrt{-h} \right) - \dots \\ &- T_5 \left(\int_{D_n} \sqrt{-h} + \int_{D_n'} \sqrt{-h} \right) . \end{aligned} \tag{6}$$

where $h_{\alpha\beta} = \partial_\alpha x^I \partial_\beta x^J g_{IJ}$ is the pull-back of the bulk metric onto the relevant fivebrane. Again $T_5 \sim m_{11}^6$ is completely determined in terms of κ[b]. Thus, as anticipated, we see that the full action, and the expression for $\mathrm{d}G$, have an expansion in m_{11}

$$S = S_{\text{bulk}} + (S_{\text{boundary}} + S_{\text{fivebrane}}) + \dots . \tag{7}$$

The leading-order term gives the bulk action (1) together with $\mathrm{d}G = 0$. The fixed plane and fivebrane terms are then suppresssed by order m_{11}^{-3}, and the dots represent further higher-order corrections.

2.2. *Braneworld compactifications*

For a realistic model, we, of course, need a theory in four dimensions. The obvious way to do this, first discussed by Witten[8] is to further compactifiy Hořava–Witten theory on a compact six-dimensional manifold X. In addition, we require $\mathcal{N} = 1$ supersymmetry. As we will discuss this implies that X is a Calabi–Yau manifold (or "threefold" since the complex dimension of X is three). Thus our eleven-dimensional space becomes

$$M_{11} = \mathbb{R}^{3,1} \times S^1/\mathbb{Z}_2 \times X, \tag{8}$$

where $\mathbb{R}^{3,1}$ is four-dimensional Minkowski space.

In addition, we can turn on non-trivial gauge fields on the two boundary planes $\partial M_i = \mathbb{R}^{3,1} \times X$. To preserve Lorentz invariance on $\mathbb{R}^{3,1}$ we only allow magnetic fields $F_{AB}(x^A)$ pointing in the internal space, where x^A with $A = 5, \dots, 9$ are coordinates on X. Supersymmetry will again constraint the precise form of F_i. For the moment we will simply refer to these rather generally as "gauge-field configurations" V_i.

---

[b]Following the original work of Hořava and Witten,[2] there has been some debate as to the exact expressions for the constants λ, k and T_5 in terms of κ required to cancel the anomalies. The current consensus[12] is, for our conventions, $\lambda^6 = (4\pi)^5\kappa^4/12$, $k^6 = 9\pi^2\kappa^4/2$, $T_5^3 = \pi/2\kappa^4$.

Finally, we can also include M-fivebranes. Lorentz invariance implies that the six-dimensional worldvolumes must be of the form

$$D_i = \mathbb{R}^{3,1} \times W_i, \tag{9}$$

where W_i is a two-dimensional cycle within $X \times S^1/\mathbb{Z}_2$. As we will see, supersymmetry then implies that W_i in fact lies entirely in X. That is, the fivebranes cannot extend along $S^1/\mathbb{Z}_2$, but rather the whole worldvolume lies at the same point y_i in y. In addition W_i must be a particular type of two-cycle, namely a "holomorphic curve", where curve refers to the fact that its *complex* dimension is one.

The choice of fivebranes and gauge fields cannot be completely arbitrary since we have to be able to solve the Bianchi identity (4) for G. Given we only have magnetic sources, the flux G must lie entirely within the internal seven-dimensional space. On a compact space like $S^1/\mathbb{Z}_2 \times X$, one cannot have a net magnetic charge since there is nowhere for the flux to escape. In other words, if we integrate the sources on any closed five-cycle C_5 in $S^1/\mathbb{Z}_2 \times X$ we must get zero since they equal $\int_{C_5} \mathrm{d}G = 0$. Given the fivebranes lie a fixed points y_i in $S^1/\mathbb{Z}_2$, the fivebrane sources can be written as $\delta(D_i) = \delta(y - y_i) \wedge \delta(W_i)$. Choosing $C_5 = S^1 \times C_4$ in the upstairs picture, we then have that $\int_{C_4} [J_1 + J_2 + \sum_i \delta(W_i)] = 0$ for any closed four-cycle C_4 in X. Equivalently we have

$$\frac{1}{16\pi^2}\operatorname{tr} F_1 \wedge F_1 + \frac{1}{16\pi^2}\operatorname{tr} F_2 \wedge F_2 + \sum_i \delta(W_i) - \frac{1}{16\pi^2}\operatorname{tr} R \wedge R = \mathrm{d}H, \tag{10}$$

for some three form H on X. In other words, the right hand side must be *trivial in cohomology* (see Appendix C). Essentially the positive magnetic charges from V_i and the fivebranes W_i must balance the negative charges from the curvature term $\operatorname{tr} R \wedge R$. Since the magnetic sources in Bianchi identity are required for anomaly cancellation, this requirement is sometimes referred to as the "anomaly cancellation condition".

Such compactifications lead to four-dimensional models, with gauge and gravitational interactions.[13] In general, the non-trivial internal gauge fields V_i, lying in some subgroup $G_i \subset E_8$ break the gauge symmetry to the group H_i, generated by those generators which commute with G_i and hence leave V_i invariant. In this way, concentrating say on V_1, one can arrange that the four-dimensional model preserves a suitable Grand Unified gauge group

H_1. In particular we have

$$
\begin{array}{cccc}
E_8 \supset & SU(3) & \times & E_6, \\
& SU(4) & \times & SO(10), \\
& SU(5) & \times & SU(5). \\
& V_1 \text{ gauge group, } G_1 & & \text{commutant, } H_1
\end{array}
\qquad (11)
$$

One can even go further. A standard way[14] to break the GUT group down to something close to the standard model is to turn on a *Wilson line* on a non-simply connnected X. In particular, one takes a non-trivial gauge field $A_A(x^A)$, aligned in the GUT group H_1, such that, although the corresponding field strength F is zero, A is not gauge equivalent to zero. (See Appendix A.) This then breaks gauge group further, to only those transformations which also leave the Wilson line invariant.

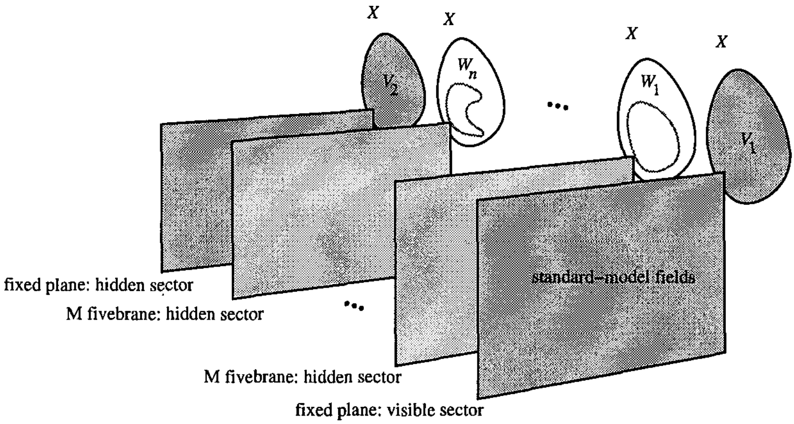


Figure 2. The heterotic M-theory braneworld

From the eleven-dimensional action (7) it is then straightforward to calculate the effective four-dimensional gravitational and GUT couplings in terms of κ, the size ρ of the orbifold interval and the volume V of the Calabi–Yau threefold. In addition, the energy scale associated to the Wilson line is set by the inverse size $V^{-1/6}$ of Calabi–Yau threefold. Thus we can identify $V^{-1/6}$ with the GUT scale. As Witten[8] and others[15] have noted, matching this and the couplings to their known values, one finds that $\rho^{-1} < V^{1/6} < m_{11}$. Thus there is a regime $\rho^{-1} < E < V^{-1/6}$ where the Calabi–Yau manifold is small but the orbifold interval is still relatively large, and the universe appears five-dimensional with $M_5 = \mathbb{R}^{3,1} \times S^1/\mathbb{Z}_2$.

In this limit, each of the fixed planes and the M-fivebranes span parallel $\mathbb{R}^{3,1}$ planes separated in the orbifold interval, and the spacetime looks like an array of domain walls or threebranes as shown in Fig. 2. Since the gauge fields are localized on the fixed planes, the standard model interactions are also confined to one the fixed planes. We are realizing the universe as a braneworld.[9]

It is important to note that this choice of background or "vacuum" given by X, the gauge-field configurations V_i and the fivebranes W_i should really be thought of as a perturbative expansion m_{11}. To leading order we simply fix the geometry of the bulk as $\mathbb{R}^{3,1} \times S^1/\mathbb{Z}_2 \times X$, together with zero flux $G = 0$. Now the curvature of the Calabi–Yau threefold X, means that to next order, suppressed by $V^{-1/3}m_{11}^3$, there is a contribution to dG from the boundaries as given in (4) as well as to the stress-energy tensor. The gauge-fields V_i and fivebranes W_i also contribute to this same order. These effects mean that G is no longer zero and the geometry is distorted. Nonetheless, as shown in by Witten,[8] supersymmetry can still be preserved. If we then had the next-order corrections to the M-theory effective action (7) we could then continue this process. In what follows, the details of the expansion will not concern us directly, only the existence of supersymetric vacua. The generic leading corrections and their effect on the low-energy effective action are discussed in detail in Burt Ovrut's lectures.[1]

As we have discussed, the preserved gauge symmetry is determined by the choice of V_i. In addition, the light degrees of freedom, namely the massless zero-modes of the eleven-dimensional fields, will also depend on the choice of brackground V_i, X and W_i. In particular, they determine the spectrum of charged matter on the braneworld. Our problem now is to choose three quantities, consistent with $\mathcal{N} = 1$ supersymmetry,

- a **Calabi–Yau threefold** X,
- **gauge-field configurations** V_1 and V_2 on the two fixed planes,
- a set of **fivebranes** wrapped on holomorphic curves within X, denoted by $W = W_1 + \cdots + W_n$,

so that the spectrum and interactions on the fixed-plane braneworld match, or are close to, those of the standard model.

We will next discuss each of these ingredients in turn, and formulate a set of conditions for realizing realistic supersymmetric models. For simplicity we require only unbroken GUT interactions with three families of matter on the braneworld.[16,17] Models with Wilson lines breaking the GUT group to $SU(3) \times SU(2) \times U(1)$ have been constructed,[18,19] but, while using the same basic techniques, are considerably more involved.

3. Conditions of supersymmetry

In this section we review how supersymmetry leads to various constraints on the ingredients which enter the braneworld model. We first consider the six-dimensional manifold X, then the gauge field configurations V_i. Here the constraints are exactly as for heterotic string models.[11] Finally we turn to the fivebranes W_i.

3.1. *Ingredient 1: the Calabi–Yau threefold, X*

The structure of the six-dimensional compact space X is determined by the requirement of $\mathcal{N} = 1$ supersymmmetry. Recall that, under supersymmetry transformations parametrised by ϵ, we always have

$$\begin{aligned} \delta_\epsilon(\text{bosons}) &= \text{fermions}, \\ \delta_\epsilon(\text{fermions}) &= \text{bosons}. \end{aligned} \tag{12}$$

If these variations vanish for certain ϵ then we have a background which preserves some supersymmetry. The cases we are interested in are purely bosonic so $\delta_\epsilon(\text{bosons}) = 0$ for all ϵ. For the bulk geometry the leading-order M-theory effective action is bulk eleven-dimensional supergravity. The only fermion is the gravitino ψ_I. Thus, for a preserved supersymmetry we require

$$\delta_\epsilon \psi_I = \nabla_I \epsilon + \tfrac{\sqrt{2}}{28} \left(\Gamma_{IJKLM} - 8 g_{IJ} \Gamma_{KLM}\right) G^{JKLM} \epsilon = 0, \tag{13}$$

where ϵ is a real 32-component spinor, Γ^I are the eleven-dimensional gamma matrices with $\{\Gamma^I, \Gamma^J\} = 2g^{IJ}$ and ∇_I is the covariant derivative,

$$\nabla_I \epsilon \equiv \left(\partial_I + \tfrac{1}{4} \omega_I{}^{JK} \Gamma_{JK}\right) \epsilon, \tag{14}$$

where $\omega_I{}^{JK}$ is the spin connection. To be compatible with the $\mathbb{Z}_2$ symmetry (2) of ψ_I, we require

$$\epsilon(x^M, -y) = \Gamma^{11} \epsilon(x^M, y). \tag{15}$$

Our ansatz is that to leading order $G = 0$, and the space is a product $\mathbb{R}^{3,1} \times S^1/\mathbb{Z}_2 \times X$. Thus the condition for supersymmetry simplifies to the "Killing spinor" equation

$$\nabla_I \epsilon = (\partial_\mu \epsilon, \partial_y \epsilon, \nabla_A \epsilon) = 0. \tag{16}$$

where x^μ, y and x^A label directions in $\mathbb{R}^{3,1}$, $S^1/\mathbb{Z}_2$ and X respectively. We can also decompose the eleven-dimensional spinor ϵ under corresponding tangent-space groups $Spin(3,1) \times Spin(6) \subset Spin(10,1)$ as $\mathbf{32} = (\mathbf{2},\mathbf{4})_+ + (\bar{\mathbf{2}},\bar{\mathbf{4}})_+ + (\mathbf{2},\bar{\mathbf{4}})_- + (\bar{\mathbf{2}},\mathbf{4})_-$. Here $\mathbf{2}$ and $\mathbf{4}$ are complex, chiral $Spin(3,1)$ and $Spin(6)$ spinors respectively, and the subscript refers to the chirality $\Gamma^{11}\epsilon = \pm\epsilon$.

Clearly, the $\partial_\mu \epsilon = \partial_y \epsilon = 0$ conditions are easy to satisfy by simply taking ϵ independent of x^μ and y. From the $\mathbb{Z}_2$ symmetry (15) of ϵ, this then implies that

$$\Gamma^{11}\epsilon = \epsilon. \tag{17}$$

Since $\operatorname{tr}\Gamma^{11} = 0$ and $(\Gamma^{11})^2 = \mathrm{id}$, we see that only half the components of ϵ survive and half the supersymmetry is broken. In terms of the decomposition of ϵ given above we have

$$\epsilon = \zeta \otimes \eta(x^A) + \bar{\zeta} \otimes \bar{\eta}(x^A), \tag{18}$$

where ζ is a constant chiral $SO(3,1)$ spinor and η a chiral $SO(6)$ spinor. For preserved supersymmetry the condition $\nabla_A \epsilon = 0$ implies that η must satisfy the six-dimensional Killing spinor equation on X

$$\nabla_A \eta \equiv \left(\partial_A + \tfrac{1}{4}\omega_A{}^{BC}\gamma_{BC}\right)\eta = 0, \tag{19}$$

where γ^A are six-dimensional gamma matrices and in the index A labels coordinates on X. From (18), we see that for each solution of this equation we have a preserved four-dimensional supersymmetry ζ. Thus for $\mathcal{N} = 1$ supersymmetry in four-dimensions, we need (19) to have a single solution.

Note that this type of problem appears in a large class of compactifications. Consider any supergravity theory (viewed perhaps as the low-energy limit of string or M-theory), compactified on a space $\mathbb{R}^{p,1} \times X$, with only the metric non-trivial. The condition for supersymmetry will then always reduce to solving the Killing spinor equation $\nabla\eta = 0$ on X where ∇ is the Levi–Citiva connection on X[c]. The existence of a solution means that the parallel transport of ϵ around closed loops in X is trivial. This implies that X has special holonomy $\mathrm{Hol}(X)$ (see Appendix A and Appendix B).

In the case in hand, $\mathrm{Hol}(X)$ is the subgroup of $Spin(6)$ which leaves a single spinor η invariant. Recall that $Spin(6) \cong SU(4)$ with chiral spinors transforming in the fundamental $\mathbf{4}$ representation. It is then clear that

[c]In fact, even with other bosonic fields in the supergravity multiplet, such as form fields, excited, the problem is still of this form, though with a generalized connection, for instance, including torsion.

the maximal subgroup which preserves one direction (given by η) in this four-dimensional space is $SU(3)$. We have that

$$\nabla\eta = 0 \quad \Leftrightarrow \quad \mathrm{Hol}(X) \subseteq SU(3), \tag{20}$$

Obviously, finding explicit metrics which satisfy the differential Killing spinor condition (19) or equivalently the holonomy condition (20), is very difficult. However, there is a very important and useful existence result conjectured by Calabi and and proved by Yau. It states

> **Calabi–Yau theorem:** If X is a compact, complex, Kähler manifold with vanishing first Chern class and Kähler form ω', then there exists a unique Kähler metric on X with $SU(n)$ holonomy, and with Kähler form ω in the same cohomology class as ω'.

Such manifolds are consequently known as Calabi–Yau n-folds. We will decode these conditions a little in a moment, but there are two main points. First, many complex Kähler manifolds are known and the Chern class condition is topological, so it is relatively easy to find examples; no actual differential equations need to be solved. Secondly, even though the theorem only states the existence of a $SU(3)$-holonomy metric, the fact that the space is complex and has definite class of Kähler form is enough to calculate most of the properties needed in the compactification.

Let us unpack these conditions a bit. First recall that a *complex manifold* is one which we admits a set of complex coordinate patches $z \in \mathbb{C}^n$ with holomorphic transition functions $z' = f(z)$ on the overlap of two patches with coordinates z and z'. Essentially this allows one to define holomorphic functions over the whole manifold. It means that the exterior derivative operator splits into holomorphic and anti-holomorphic parts, $\mathrm{d} = \partial + \bar{\partial}$, and all tensor objects decompose globally into $GL(n, \mathbb{C})$ representations with holomorphic and anti-holomorphic indices. We label these indices a and $\bar{a}$ and refer a (p, q)-form as one with p holomorphic and q antiholomorphic indices. (As noted in Appendix B, an equivalent definition is a real manifold with an additional "complex structure" $J^A{}_B$ with vanishing Nijenhuis tensor.)

A metric g on a complex manifold is *hermitian* if it is compatible with the complex structure. This means $g_{CD} J^C{}_A J^D{}_B = g_{AB}$, or equivalently we can write

$$\mathrm{d}s^2 = g_{a\bar{b}} \mathrm{d}z^a \mathrm{d}z^{\bar{b}}. \tag{21}$$

Given a hermitian metric we can define the fundamental $(1,1)$-form $\omega_{AB} =$

$g_{AC}J^C{}_B$. In holomorphic indices we have

$$\omega = -\mathrm{i}g_{a\bar{b}}\mathrm{d}z^a \wedge \mathrm{d}z^{\bar{b}}. \tag{22}$$

A *Kähler manifold* is one which admits a hermitian metric where the fundamental form is closed, The metric is known as a Kähler metric and the fundamental form as the Kähler form. The holonomy defined by a Kähler metric is in $U(n)$ (see Appendix B). Thus we have

$$\text{Kähler metric} \quad \Leftrightarrow \quad \mathrm{d}\omega = 0 \quad \Leftrightarrow \quad \mathrm{Hol}(X) \subseteq U(n) \tag{23}$$

As we review in Appendix H, the first Chern class $\mathrm{c}_1(TX)$ is a topological invariant of the tangent bundle TX of a complex manifold. It can be defined as the cohomology class of the Ricci form $\mathcal{R}_{AB} = -\frac{1}{2}R_{AB}{}^{CD}\omega_{CD}$ where R is the Riemann tensor. It is independent of the choice of Kähler metric on X. Thus the fact that $\mathrm{c}_1(TX)$ vanishes for a Calabi–Yau manifold implies for any Kähler metric on X that $\mathcal{R} = \mathrm{d}\alpha$ for some α. Given the Calabi–Yau theorem, this is the only requirement for X to admit a Kähler metric of $SU(n)$ holonomy. In fact, there is a unique such metric for each cohomology class of Kähler form and fixed complex structure on X.

It is perhaps useful to mention some very simple cases of complex, Kähler and Calabi–Yau manifolds

Riemann surfaces: On a two-dimensional manifold, the generic holonomy is $SO(2) \cong U(1)$, so every metric is Kähler. Note in this case the Calabi–Yau condition requires that the holonomy is trivial, and the only supersymmetric Riemann surface is the torus T^2.

Complex projective space $\mathbb{C}P^n$: These manifolds are defined simply be projecting flat $\mathbb{C}^{n+1}$ with the identification

$$(z_1, \ldots, z_{n+1}) \sim (\lambda z_1, \ldots, \lambda z_{n+1}) \quad \text{for } \lambda \in \mathbb{C} - \{0\}.$$

They naturally inherit a complex structure from $\mathbb{C}^{n+1}$. The simplest case is $\mathbb{C}P^1 = S^2$. However, they are not Calabi–Yau.

Intersections in $\mathbb{C}P^n$: Consider a function on $\mathbb{C}P^n$, defined in terms of homogeneous polynomial

$$P(\lambda z_1, \ldots, \lambda z_{n+1}) = \lambda^t P(z_1, \ldots, z_{n+1}).$$

One can then define a subspace of $\mathbb{C}P^n$ by

$$P_1 = P_2 = \cdots = P_q = 0 \quad \text{for some set of polynomials } \{P_i\}.$$

> Such a space is known as a *algebraic variety* and again inherits a complex structure from $\mathbb{C}P^n$. This is the basic idea of algebraic geometry: one replaces an abstract complex manifold with a variety in projective space, which can be analyzed purely algebraically. Picking $\{P_i\}$ carefully can give a Calabi–Yau manifold. (The classic example[11] is the quintic hypersurface in $\mathbb{C}P^4$.)

Before finishing this section it is useful to note some properties of Calabi–Yau manifolds. First we note that any special holonomy manifold is Ricci flat (see Appendix B). Thus satisfying the supersymmetry conditions imply that we also satisfy the, leading order, supergravity equations of motion.

Second consider the Killing spinor η. We can use the complex structure to define holomorphic gamma-matrices which satisfy the algebra of fermionic creation and annihilation operators, with $\gamma^{a\dagger} = \gamma^{\bar{a}}$,

$$\begin{aligned} \{\gamma^a, \gamma^b\} &= \{\gamma^{\bar{a}}, \gamma^{\bar{b}}\} = 0, \\ \{\gamma^a, \gamma^{\bar{b}}\} &= 2\delta^{a\bar{b}}. \end{aligned} \tag{24}$$

Thus, following a standard procedure,[11] we can build a general spinor ψ out of anti-symmetric products of $\gamma^{\bar{a}}$ acting on a vacuum state χ satisfying $\gamma^a \chi = 0$. We have

$$\psi = \lambda\chi + \lambda_{\bar{a}}\gamma^{\bar{a}}\chi + \cdots + \lambda_{\bar{a}_1\ldots\bar{a}_n}\gamma^{\bar{a}_1\ldots\bar{a}_n}\chi. \tag{25}$$

Recall that the Killing spinor χ is invariant under the holonomy group $SU(n) \subset Spin(2n)$. This implies that the $SU(n)$ generators $\phi^0_{a\bar{b}}\gamma^{a\bar{b}}$, where $g^{a\bar{b}}\phi^0_{a\bar{b}} = 0$, annihilate η. From Eq. (25), this means that, either $\eta \propto \chi$ or $\eta \propto \gamma^{\bar{a}_1\cdots\bar{a}_n}\chi$, the two cases being related by complex conjugation. For $SU(3)$ we will fix on the former case, that is $\gamma^a\eta = 0$, which in turn fixes the chirality of η.

The components in the expansion (25) are $(0,q)$-forms. It is easy to see that the action of the Dirac operator $i\gamma^A\nabla_A$ corresponds $\bar{\partial} + \bar{\partial}*$ acting on the set of forms $\{\lambda_{a_1\ldots a_p}\}$. This implies that a solution to the Dirac equation $i\gamma^A\nabla_A\psi = 0$ corresponds to a $\bar{\partial}^2$ harmonic $(0,q)$-form, that is an element of the Dolbeault cohomology (see Appendix C). By considering $\int_X \bar{\psi}(\gamma^A\nabla_A)^2\psi$, one can show that there is a solution to the Dirac equation if and only if $\nabla_A\psi = 0$. From the expansion we know that there are only two such solutions on a Calabi–Yau manifold, namely η and $\bar{\eta}$ corresponding to the $(0,0)$-form and the $(0,n)$-form. Thus we see that Hodge numbers $h^{0,q}$ are all zero except for $h^{0,0} = h^{0,n} = 1$. Thus for a Calabi–Yau threefold, given $h^{p,q} = h^{q,p}$ and $h^{n-p,n-q} = h^{p,q}$ for any Kähler manifold, we have the "Hodge diamond"

$$\begin{array}{ccccccc} & & & 1 & & & \\ & & 0 & & 0 & & \\ & 0 & & h^{1,1} & & 0 & \\ 1 & & h^{2,1} & & h^{1,2} & & 1 \\ & 0 & & h^{1,1} & & 0 & \\ & & 0 & & 0 & & \\ & & & 1 & & & \end{array}$$

3.2. *Ingredient 2: gauge-field configurations, V_1 and V_2*

As we have discussed, to get a suitable low-energy GUT group we have to include non-trivial E_8 gauge fields on the orbifold fixed-planes. In particular, we need fields in some $SU(n)$ sub-group of E_8 on the Calabi–Yau part X of the fixed planes. We also have to ensure, however, that we preserve some $\mathcal{N} = 1$ supersymmetry in four dimensions. As before, this requires that the supersymmetry transformations of the fermion partners of the gauge fields, the gauginos, vanish. We have in general

$$\delta_\epsilon \chi = F_{\bar{I}\bar{J}} \Gamma^{\bar{I}\bar{J}} \epsilon = 0 \tag{26}$$

where $\bar{I}$ etc label coordinates on the fixed planes ∂M_i. To be compatible with the supersymmetry preserved by the bulk spacetime, ϵ must be of the form given in Eq. (18). Since the background has F non-trivial on X, these conditions then reduce to

$$F_{AB} \gamma^{AB} \eta = 0. \tag{27}$$

On a Calabi–Yau manifold, we can choose η such that $\gamma^a \eta = 0$ and that a general spinor can be expanded in terms of $\gamma^{\bar{a}_1 \cdots \bar{a}_k}$ acting on η as in Eq. (25). Thus expanding, we find that the content of Eq. (27) is that the parts of F_{AB} representing generators of rotations in $Spin(2n)$ not in $SU(n)$ must be set to zero, namely

$$\begin{aligned} F_{ab} = F_{\bar{a}\bar{b}} &= 0, \\ \mathrm{i} g^{a\bar{b}} F_{a\bar{b}} &= 0. \end{aligned} \tag{28}$$

These equations are familiar in mathematics and are known as the *Hermitian Yang–Mills* (HYM) equations. They are of particular interest because they have nice moduli spaces of solutions. In some sense they are the generalisation to arbitrary Kähler manifolds of the self-dual instanton equations on four-dimensional manifolds.

Explicitly solving the HYM equations is extremely difficult as they are non-linear partial differential equations on a curved manifold (and we don't even have an explicit metric in the case of Calabi–Yau manifolds!). What we need is some topological condition, analogous to that of the Calabi–Yau theorem which implies the existence of a solution, even if we cannot solve the equations explicitly. Fortunately there is just such a result in the mathematics literature, due to Donaldson[20] (in four dimensions) and Uhlenbeck and Yau[21] (in general dimension):

> **Donaldson–Uhlenbeck–Yau theorem:** there is a one-to-one correspondence between solutions of the HYM equations and *poly-stable*, *holomorphic* vector bundles V with $c_1(V) = 0$.

Given this identification we will often refer to the solutions of the HYM equations, V_i, simply as "bundles".

Let us look at what the various terms in the theorem mean. Recall that for a *complex* vector bundle E each fibre is a copy of $\mathbb{C}^n$ (see Appendix G). A general connection $D = \mathrm{d} + \mathrm{i}A$ on E, is then a complex one-form and which has gauge transformations in $GL(n, \mathbb{C})$.

We can add more structure by considering a hermitian metric h giving an inner product $\bar{\xi} \cdot \zeta \equiv h(\xi, \zeta)$ on sections ξ and ζ. This allows us to define hermitian conjugates. Given such a metric (which always exists) the transition functions ϕ_{UV} are, by construction, unitary, and we can view E as a $U(n)$ vector bundle. We can also then ask for real connections satisfying $A^\dagger = A$. Such connections then necessarily preserve the metric h, that is

$$\mathrm{d}(\bar{\xi} \cdot \zeta) = \overline{D\xi} \cdot \eta + \bar{\xi} \cdot D\zeta. \tag{29}$$

Also note that the reality condition is only preserved by unitary gauge transformations. In other words we can think of a $U(n)$ connection as a real, metric-compatible connection on a complex vector bundle with hermitian metric h.

Forgetting about the metric h for a moment, let us now specialize to the case where the base manifold X is complex. This means we can decompose a connection into holomorphic and antiholomorphic pieces. We write $D = \partial_{(A)} + \bar{\partial}_{(A)}$ where $\partial_{(A)a} = \partial_a + \mathrm{i}A_a$ and $\bar{\partial}_{(A)\bar{a}} = \partial_{\bar{a}} + \mathrm{i}A_{\bar{a}}$. We can then decompose the components of the curvature F as

$$\begin{aligned} F &= F_{ab} \oplus \qquad F_{a\bar{b}} \qquad \oplus F_{\bar{a}\bar{b}} \\ &= \partial^2_{(A)} \oplus \partial_{(A)}\bar{\partial}_{(A)} + \bar{\partial}_{(A)}\partial_{(A)} \oplus \bar{\partial}^2_{(A)} \end{aligned} \tag{30}$$

Recall that one of our conditions in the HYM equations was that $F_{\bar{a}\bar{b}} = \bar{\partial}^2_{(A)\bar{a}\bar{b}} = 0$. One immediately notes that this is an integrability condition for finding solutions $\bar{\partial}_{(A)}\zeta = 0$, that is, for finding global *holomorphic* sections of E.

The obvious way to have global holomorphic sections is to have a *holomorphic* vector bundle (see Appendix G). This is a complex vector bundle over a complex base X where the transitions functions ϕ_{UV} are holomorphic functions on X. Such objects are important because the total space of base and fibre then forms a complex manifold. We can then consider complex connections D, where we take $\bar{\partial}_{(A)} = \bar{\partial}$, that is $A_{\bar{a}} = 0$. Then, since the transition functions are holomorphic, a section which is a holomorphic function on any given patch U, satisfies, $\bar{\partial}_{(A)}\zeta = \bar{\partial}\zeta = 0$ everywhere. Such connections are called compatible with $\bar{\partial}$. In fact, this is the only way to have solutions to $\partial_{(A)}\zeta = 0$: if we have a connection satisfying $\bar{\partial}^2_{(A)} = 0$, then E must be a holomorphic vector bundle and we have

Newlander–Nirenberg theorem:

$$F_{\bar{a}\bar{b}} = 0 \quad \Leftrightarrow \quad E \text{ is a holomorphic bundle.}$$

Note in the simplest case where the fibre is just $\mathbb{C}$, E is called a line bundle.

Now suppose in addition we have a metric h on the fibres. A connection which preserves h, will then also have $F_{ab} = 0$, by conjugation. This is just a reflection that we can also find conjugate anti-holomorphic sections $\partial_{(A)}\bar{\zeta} = 0$. We have seen that holomorphic bundles have a natural class of complex connections which are compatible with $\bar{\partial}$, that is $\bar{\partial}_{(A)} = \bar{\partial}$. If we also require compatibility with the metric, so that Eq. (29) is satisfied, then A is unique, we have

Theorem: A holomorphic bundle E with hermitian metric h, determines a unique complex connection compatible with both the metric and the complex structure, that is a $D = \partial_{(A)} + \bar{\partial}_{(A)}$ with $\bar{\partial}_{(A)} = \bar{\partial}$ and $Dh = 0$. Explicitly we have

$$A_{\bar{a}} = 0, \qquad A_a = -\mathrm{i}\,\partial_a h^{-1}\, h.$$

Now we can always write the hermitian metric as $h = V^\dagger V$ for some matrix V in $GL(n, \mathbb{C})$ fixed up to unitary transformations. Making a $GL(n, \mathbb{C})$ gauge transformation by V we have the real connection, unique up to $U(n)$ transformations,

$$A_a = \mathrm{i}\,\partial_a V^\dagger\, {V^\dagger}^{-1}, \qquad A_{\bar{a}} = \mathrm{i}\,\bar{\partial}_{\bar{a}} V\, V^{-1}. \tag{31}$$

Thus we have two equivalent objects

(1) A $U(n)$ connection satisfying $F_{ab} = F_{\bar{a}\bar{b}} = 0$ up to $U(n)$ gauge transformations;
(2) A holomorphic vector bundle E with metric h.

Of course, what we really want is a $SU(n)$ bundle so that the trace part of F or equivalently of A vanishes. Given the unique connection compatible with the metric h and complex structure, this requires a solution to $-\operatorname{tr}\partial_a h^{-1}\, h = \partial_a \det h = 0$. In other words, the $U(1)$ determinant bundle must have a constant section. This is possible if and only if this $U(1)$ bundle is trivial. In other words, we require $\mathrm{c}_1(E) = 0$ and then there is a unique $SU(n)$ bundle given a metric h on E.

Finally we come to the last condition of supersymmetry, namely that, given a Kähler metric g on the base, $\mathrm{i}g^{a\bar{b}}F_{a\bar{b}} = 0$. If is natural to first generalize this condition slightly to include a trace part (which is necessarily zero for $SU(n)$ bundles). We take

$$\mathrm{i}g^{a\bar{b}}F_{a\bar{b}} = \lambda\mathrm{id}, \tag{32}$$

for some constant λ. (This, in fact, together with the holomorphic conditions $F_{ab} = F_{\bar{a}\bar{b}} = 0$ is the usual form of the HYM equations.) Note that when $\mathrm{c}_1(V) = 0$ we necessarily have $\lambda = 0$.

From the above discussion, we see that solving Eq. (32) requires finding a metric h on a holomorphic vector bundle satisfying a second-order non-linear differential equation. In general not all holomorphic bundles admit such a solutions. The content of the Donaldson–Uhlenbeck–Yau theorem is to characterize the conditions on E for there to be solutions. It states that there is (essentially) a unique solution for each bundle satisfying the condition that it is *poly-stable* (this is sometimes also called poly-stable). Stablity is the condition that

> **Stablity:** a holomorphic bundle E is stable if for any sub-bundle $E' \subset E$ with $\mathrm{rk}(E') < \mathrm{rk}(E)$ we have $\mu(E') < \mu(E)$

where $\mu(E)$ is the "slope",

$$\mu(E) \equiv \frac{\deg(E)}{\mathrm{rk}\,E} = \frac{1}{n(d-1)!}\int_X \mathrm{c}_1(E) \wedge \underbrace{\omega \wedge \cdots \wedge \omega}_{(n\,-\,1)\text{-times}}, \tag{33}$$

where $\mathrm{rk}(E) = n$ is the rank of the bundle and $\deg(E)$ is known as the degree. A bundle is then *poly-stable* if it is direct sum $E = E_1 \oplus \cdots \oplus E_k$ of stable bundles.

Again, as in the case of the Calabi–Yau theorem we don't have to construct connections explicitly, we simply need to find a poly-stable holomorphic bundle V, and then we know a solution exists. The only additional requirement is the topological condition $c_1(V) = 0$ which ensures that we have an $SU(n)$ rather than $U(n)$ connection.

3.3. *Ingredient 3: fivebranes,* W_i

Finally we turn to the conditions on how the fivebranes are wrapped if they are to preserve $\mathcal{N} = 1$ supersymmetry. This has also been discussed in Burt Ovrut's lectures.[1] First, we note that for a fivebrane to preserve four-dimensional Lorentz invariance, four dimensions of the fivebrane must be left uncompactified, the remaining directions wrapping a two-cycle W_i in the internal $S^1/\mathbb{Z}_2 \times X$ space. It is the conditions of supersymmetry[22,8] which restrict what two-cycles W_i are allowed.

To analyse this problem we start with the Green–Schwarz action[23] for an M5-brane worldvolume $\Sigma \hookrightarrow M$ embedded in an eleven-dimensional space M. There are three fields on Σ: x^I giving the coordinates of the embedding, a spacetime spinor θ and a self-dual world-volume three-form field strength $h = db$. The action is invariant under both spacetime supersymmetry transformations, parameterized by ϵ, and κ-symmetry transformations, which are local symmetries on the worldvolume parametrized by a spacetime spinor κ. As for the Green–Schwarz superstring, the κ-symmetries are gauge symmetries which can be used to remove half the fermionic degrees of freedom. In general the fermion fields θ transform as[24]

$$\delta\theta = \epsilon + P_+\kappa \tag{34}$$

where P_+ is a projection operator satisfying $P_+^2 = \mathrm{id}$. We also define $P_- = \mathrm{id} - P_+$. In the case where $h = 0$, the projection operators take the form

$$P_\pm = \frac{1}{2}\left(1 \pm \frac{1}{6!\sqrt{h}}\epsilon^{m_1\ldots m_6}\partial_{m_1}x^{I_1}\ldots\partial_{m_6}x^{I_6}\Gamma_{I_1\ldots I_6}\right), \tag{35}$$

where g is the determinant of the induced metric

$$h_{mn} = \partial_m x^I \partial_n x^J g_{IJ}. \tag{36}$$

For a classical background the brane configuration is purely bosonic and we need only consider $\delta\theta = 0$ to look for supersymmetric configurations. This implies, $\epsilon = -P_+\kappa$, or equivalently we need an ϵ such that

$$P_-\epsilon = 0. \tag{37}$$

To be compatible with the supersymmetry preserved by the Calabi–Yau manifold X and the HYM solution, the spinor ϵ must have the form given in Eq. (18), statisfying $\Gamma^{11}\epsilon = \epsilon$. Substituting into the expression (35) for the projector, we find the condition

$$\left[\left(\partial x^A \bar{\partial} x^B - \bar{\partial} x^A \partial x^B\right)\gamma_{AB} + \left(\partial x^A \bar{\partial} x^{11} - \bar{\partial} x^A \partial x^{11}\right)\gamma_A\right]\eta = \sqrt{h}\eta, \quad (38)$$

where we have introduce complex coordinates $\sigma = \sigma^5 + i\sigma^6$ and on the two-dimensional part of the fivebrane worldvolume wrapped on the cycle W_i in the internal space. (For the rest of the worldvolume spanning $\mathbb{R}^{3,1}$ we simply have $x^\mu = \sigma^\mu$.) The induced metric on W_i is $h_{\sigma\bar{\sigma}} = \partial x^\alpha \bar{\partial} x^\beta g_{\alpha\beta}$ where α and β run over the seven-dimensional internal $S^1/\mathbb{Z}_2 \times X$ space.

If we use the properties that $\gamma^a\eta = 0$ and that a general spinor can be expanded, as in Eg. (25), in terms of $\gamma^{\bar{a}_1\cdots\bar{a}_k}$ acting on η, we find that Eq. (38) implies that

$$\begin{aligned} \partial x^{\bar{a}} \bar{\partial} x^{\bar{b}} - \bar{\partial} x^{\bar{a}} \partial x^{\bar{b}} &= 0, \\ \partial x^{\bar{a}} \bar{\partial} x^{11} - \bar{\partial} x^{\bar{a}} \partial x^{11} &= 0, \\ \left(\partial x^a \bar{\partial} x^{\bar{b}} - \bar{\partial} x^a \partial x^{\bar{b}}\right) g_{a\bar{b}} &= \sqrt{h} \end{aligned} \quad (39)$$

These, in turn, imply

$$\begin{aligned} x^{11} &= y_i && \text{constant}, \\ x^a &= f^a(\sigma) && \text{holomorphic function}. \end{aligned} \quad (40)$$

and $h_{\sigma\bar{\sigma}} = \partial_\sigma x^a \bar{\partial}_{\bar{\sigma}} x^{\bar{b}} g_{a\bar{b}}$. In other words we have shown that to be supersymmetric the whole of the M5-brane must lie at a single point in the orbifold interval $S^1/\mathbb{Z}_2$ and wrap a *holomorphic* curve (two-cycle) in X. This means that the complex embedding functions x^a are holomorphic functions $f^a(\sigma)$ of complex coordinate σ on the curve. Equivalently, W_i can be specified by the zeros of a set of holomorphic functions on X (see Appendix E). As discussed in Appendix F, in common with all supersymmetric wrappings, these are calibrated cycles, in this case calibrated by the Kähler form ω.

As is discussed in more detail in Burt Ovrut's lectures,[1] the fields on the fivebranes can also lead to additional low-energy degrees of freedom in four dimensions.[10,25] Since the brane is localized in the orbifold interval these appear as additional hidden sector branes in five dimensions. Of particular interest is when the fivebrane curves overlap or intersect, in which case non-Abelian gauge groups can appear.

4. Effective curves, three families and summary of conditions

In the previous section we have discussed how the conditions of supersymmetry constrain X to be a Calabi–Yau manifold, the bundles V_i to be poly-stable and holomorphic and the fivebranes to wrap holomorphic curve W_i in X. If V_1 is the bundle on the wall corresponding to our own universe, then, for a GUT model in four-dimensions we see from Eq. (11) that the gauge group of V_1 must be $SU(3)$, $SU(4)$ or $SU(5)$. It turns out that there are two further constraints we must consider: the consequences of the anomaly cancellation condition (10) and the requirement that we see three families of matter in the low-energy theory.

4.1. *Anomaly condition and effective classes*

Recall that given X is compact, the combination of fivebrane magnetic charge in Eq. (10) must be exact. In other words it must be trivial in as an element in cohomology $H^4(X,\mathbb{R})$. As discussed in Appendix D, elements in cohomology $H^p(X,\mathbb{R})$ are Poincaré dual to elements in homology $H_p(X,\mathbb{R})$. Thus we can equally well think of the terms in Eq. (10) as representing two-cycles in X.

For fivebrane terms the delta functions $\delta(W_i)$ are Poincaré dual to homology class of the curves themselves, namely W_i. Both the gauge bundles V_i and the tangent bundle TX of the Calabi–Yau manifold are complex vector bundles. As a result, the sources quadratic in F_1, F_2 and R, are, in fact, none other than the corresponding second Chern classes (see Appendix H). Viewing these classes in terms for the Poincaré dual elements in $H_2(X,\mathbb{R})$, we see that the anomaly condition becomes

$$\mathrm{c}_2(V_1) + \mathrm{c}_2(V_2) + [W] = \mathrm{c}_2(TX), \tag{41}$$

where $[W]$ is the homology class of the sum of fivebrane cycles $W = W_1 + \cdots + W_n$.

One way to view this condition is to recall that $\mathrm{c}_2(V)$ counts the instanton number for a bundle in four-dimensions. In four-dimensions, an instanton is localized at a point, thus on six-dimensional X such a configuration will localize on a two-dimensional curve. Viewed in homology $\mathrm{c}_2(V)$ is the homology class of this curve. From this point of view the $\mathrm{c}_2(V_i)$ look like five-brane sources (coming from the gauge field configurations) localized in the fixed planes, while $\mathrm{c}_2(TX)$ is a sort of negative fivebrane source coming from the curvature of the manifold. This picture actually has meaning physically, since one can show that there are so-called small-instanton transitions where V_i becomes singular and a new fivebrane is produced which can then move away from the fixed-plane, while keeping the net charge on the right-hand-side of Eq. (41) fixed.[26]

It might appear that whatever the values of $c_2(V_i)$ and $c(TX)$ we can always satisfy the condition (41), by choosing the fivebranes suitably. However, the class $[W]$ is not completely general since it must correspond to an actual combination of fivebranes. Since we cannot have $\frac{3}{8}$ths of a fivebrane in the background, the first condition is that $[W]$ is actually an integral class, an element of $H_2(X, \mathbb{Z})$. However, this is not in itself a constraint since the Chern classes $c_2(V_i)$ and $c(TX)$ are also integral (see Appendix H).

More significant is that, first, the fivebranes $W = W_1 + \cdots + W_n$ must be some actual physical set of holomorphic curves in X, and second, that we cannot have anti-branes in the background without breaking supersymmetry. The fact that W_i are holomorphic curves means that W is a *analytic cycle*. This is reviewed in Appendix E. Generically the homology group of analytic cycles is a sub-group of $H_{2p}(X, \mathbb{Z})$ known as the algebraic holonomy group. For curves we have

$$[W] \in H_2(X, \mathbb{Z})_{\text{alg}} \subseteq H_2(X, \mathbb{Z}). \tag{42}$$

Generically, this is actually not yet a constraint since one can show, as discussed in Appendix H that for holomorphic bundles V_i and TX, all the Chern classes are actually *algebraic* (provided X is an algebraic variety).

In fact, because X is a Calabi–Yau manifold we have $h^{2,0} = 0$, then the *Lefschetz theorem* (see Appendix E) implies that

$$H_2(X, \mathbb{Z})_{\text{alg}} = H_2(X, \mathbb{Z}). \tag{43}$$

Thus we necessarily have that $c_2(V_i)$ and $c_2(TX)$, which generically are elements of $H_2(X, \mathbb{Z})$, are actually algebraic.

Finally, we might imagine we can have two different fivebranes wrapping a curve W_i with opposite orientation so that they contribute to the magnetic charge with opposite sign. However, in general this will break supersymmetry. This can be seen explicitly in the fivebrane supersymmetry conditions (39). One might imagine you can find a "anti-holomorphic" solution with $x^a = f^a(\bar{\sigma})$. However, the third condition involving $\sqrt{h}$ will then have the wrong sign and so not be satisfied. Thus the supersymmetric branes come with a particular orientation, inherited from the orientation of X, and so cannot contribute a negative source.

Mathematically this is encoded in the notion of an *effective* class as defined in Appendix E. It means that W can be expressed a the positive sum of actual holomorphic curves with an orientation inherited from X. This requirement finally imposes a restriction, since there in no reason that the difference $c_2(TX) - c_2(V_1) - c_2(V_2)$ though algebraic should be effective.

In summary, we see that we must satisfy the anomaly cancellation condition (41), as an equation in integer homology, subject to the condition that $[W]$ is an *effective class.*

4.2. *Three family condition*

The final condition is very physical: it is the requirement that we have only three families of charged matter in the four-dimensional GUT model, corresponding to the three families of quarks and leptons in the standard model.[27]

By construction the four-dimensional effective action is supersymmetric. The charged matter fields must thus appear in massless chiral multiplets. (The particles then get masses through the Higgs mechanism.) The only fields in the full eleven dimensional theory which are charged under the GUT gauge group are those in the super Yang–Mills theory on the observable wall. Thus the charged fermions must all arise from components of the gauginos on the observable fixed plane. The gauginos χ are ten-dimensional spinors and lie in the adjoint representation of E_8. In ten-dimensions they are massless, satisfying the equation of motion

$$\mathrm{i}\Gamma^{\bar{I}} D_{\bar{I}}\chi = \left(\mathrm{i}\Gamma^{\mu}\partial_{\mu} + \mathrm{i}\Gamma^{A} D_{A}\right)\chi = 0 \tag{44}$$

where we have split the Dirac operator into parts in $\mathbb{R}^{3,1}$ and the internal X. Note that D includes both the spin-connection and the gauge field connection.

If we decompose under $Spin(3,1) \otimes Spin(6)$, we have $\chi = \lambda \otimes \psi + \bar{\lambda} \otimes \bar{\psi}$ where λ and ψ are chiral spinors on $\mathbb{R}^{3,1}$ and X respectively. (We are using the decomposition of gamma matrices $\Gamma^{\mu} = \gamma^{\mu} \otimes \mathrm{id}$ and $\Gamma^{A} = \gamma_5 \otimes \gamma^{A}$, where $\gamma_5 = \mathrm{i}\gamma^0\gamma^1\gamma^2\gamma^3$.) Then for a massless four-dimensional spinor field, satisfying $\mathrm{i}\gamma^{\mu}\partial_{\mu}\lambda = 0$ we must have a "zero-mode" on X

$$\mathrm{i}\gamma^{A} D_{A}\psi = 0. \tag{45}$$

(Note, if instead we solve for a non-zero eigenvalue $\mathrm{i}\gamma^{A} D_{A}\psi = m\psi$ then λ has mass m. On compact X, the value of m is set by the inverse size of the manifold, which, as we saw in Sec. 2 is of order the GUT scale. This is far too heavy for these Kaluza–Klein modes to correspond to standard model fields.)

Now χ also has a decomposition on its gauge indices under the GUT group H. It is in the **248** adjoint of E_8, and we find

$$\begin{aligned}
\mathbf{248} &= (\mathbf{8},\mathbf{1}) + (\mathbf{1},\mathbf{78}) + (\mathbf{3},\mathbf{27}) + (\bar{\mathbf{3}},\overline{\mathbf{27}}) && SU(3)\otimes E_6, \\
&= (\mathbf{15},\mathbf{1}) + (\mathbf{1},\mathbf{45}) \\
&\quad + (\mathbf{4},\mathbf{16}) + (\bar{\mathbf{4}},\overline{\mathbf{16}}) + (\mathbf{6},\mathbf{10}) && SU(4)\otimes SO(10), \\
&= (\mathbf{24},\mathbf{1}) + (\mathbf{1},\mathbf{24}) \\
&\quad + (\mathbf{10},\mathbf{5}) + (\mathbf{5},\overline{\mathbf{10}}) + (\overline{\mathbf{10}},\bar{\mathbf{5}}) + (\bar{\mathbf{5}},\mathbf{10}) && SU(5)\otimes SU(5).
\end{aligned} \tag{46}$$

These representations all decouple in the zero-mode equation (45). Consider, for instance, the second term in each expansion which transforms in the adjoint of the GUT group (E_6, $SO(10)$ or $SU(5)$). Each also transforms as a scalar under the gauge bundle group $G = SU(n)$. Thus, in this case, the gauge connection A decouples from the zero-mode equation (45) on X and we can replace D by ∇. As a result, we always have a single solution (see Sec. 3.1), since we can take ψ proportional to the Killing spinor η. We see that we always have massless four-dimensional gaugino fields in the adjoint representation of the GUT group. (This must be the case since these are the supersymmetric partners of the GUT gauge fields.)

In GUT theories, to have the correct charges under the standard model gauge groups, the chiral charged matter must be in the **27** of E_6, the **16** of $SO(10)$ and the $\mathbf{10} + \bar{\mathbf{5}}$ of $SU(5)$. These all appear, together with their conjugate representations in the expansion (46). In particular, we note they always appear partnered with the fundamental representation of the gauge bundle group $G = SU(n)$. Thus to find what charged matter we have we need to find the zero modes of Eq. (45) with ψ in the fundamental $\mathbf{n}$ and $\bar{\mathbf{n}}$ representations of $SU(n)$.

The standard model, and hence GUT theories, with three families are chiral in the sense that, fixing the representation of the GUT group, say the **16** of $SO(10)$, we need three spinors of positive four-dimensional chirality and none with negative chirality. The way we have written Eq. (45) is slightly different. There we have fixed the six-dimensional chirality of the spinor ψ to be positive but allowed both **16** and $\overline{\mathbf{16}}$ representations (or equivalently $\mathbf{n}$ and $\bar{\mathbf{n}}$ representations of G). However, the **16** representation is just the conjugate of the $\overline{\mathbf{16}}$ representation, and conjugation also reverses chirality. Furthermore the six- and four-dimensional chiralities are correlated since the spinors are chriral in ten dimensions. Thus the conjugates of the $\bar{16}$ zero modes correspond to negative chirality matter in the **16** representation. This means we can formulate the three family condition as follows. Take a general spinor $\hat{\psi}_{(\mathbf{n})}$ of arbitrary chirality but which transforms in a definite representation, **n**, of $G = SU(n)$. In all cases of GUT group we need to know the number of position and negative chirality solutions to the equation

$$\mathrm{i}\gamma^A D_A \hat{\psi}_{(\mathbf{n})} = 0. \tag{47}$$

At first sight it would appear that we need three positive chirality solutions ($n_+ = 3$) and no negative chirality solutions ($n_- = 0$). In fact, we can be a little less ambitious. If we have a pair of solutions with opposite chiralities, we can always write an explicit mass term for them in the low-energy theory. Since such a mass term is possible, we might expect that at some scale such a term is generated. Thus the minimum requirement is simply that there is an net excess of positive chirality solutions. That is

$$N = n_+ - n_- = 3. \tag{48}$$

Fortunately, this number has a very nice mathematical interpretation. It is the *index* $\mathrm{index}_{\mathbf{n}}(\mathrm{i}\gamma^A D_A)$ of the operator $\mathrm{i}\gamma^A D_A$ in the representation **n**. We will not give a explanation of index theory here but it is discussed briefly in Green, Schwarz and Witten[11] and in considerably more detail in Eguchi, Gilkey and Hansen.[63] The relevant point is that the index is a topological invariant and is given by Chern classes. In particular, on a Calabi–Yau threefold, it depends only on the holomorphic gauge bundle, namely V_1 since we are on the observable fixed plane. We have

$$N = \mathrm{index}_{\mathbf{n}}(\mathrm{i}\gamma^A D_A) = \tfrac{1}{2}\,\mathrm{c}_3(V_1) = \frac{1}{48\pi^2}\int_X \mathrm{tr}\, F^{(1)} \wedge F^{(1)} \wedge F^{(1)}. \tag{49}$$

The condition for three families is simply a cohomological constraint on the third Chern class of V_1.

4.3. *Summary of conditions*

To summarize, we are considering vacuum states of M-theory with the following structure. The conditions of supersymmetry imply that we have to choose,

- a **Calabi–Yau manifold** X,
- a pair of **poly-stable holomorphic gauge bundles** V_i with fiber group $G_i \subseteq E_8$ over X,

- a set of **five-branes** in the vacuum, which are wrapped on **holomorphic two-cycles** within X and are parallel to the orbifold fixed planes.

Subject to the conditions:

(1) For a **GUT model** in four-dimensions we have $G_1 = SU(n)$ for $n = 3$, 4 or 5 corresponding to E_6, $SO(10)$ and $SU(5)$ GUT groups respectively. This requires

$$c_1(V_1) = 0. \tag{50}$$

(2) There is a **cohomological constraint** on X

$$c_2(V_1) + c_2(V_2) + [W] = c_2(TX), \tag{51}$$

where $c_2(V_i)$ and $c_2(TX)$ are the second Chern classes of the gauge bundle V_i and the tangent bundle TX respectively and $[W]$ is an **effective class** associated with the five-branes .

(3) To have **three families** of chiral matter in four-dimensions we require

$$c_3(V_1) = 6. \tag{52}$$

Backgrounds of this type have been referred to as *heterotic M-theory vacua.* For simplicity, we will assume

$$V_2 \text{ is trivial, write } V \text{ for } V_1$$

implying that the full E_8 symmetry is preserved on the hidden fixed plane.

Note that these conditions would be exactly the same if one was constructing models in the weakly coupled heterotic string limit. The only difference is that one could not naturally include the fivebranes, since these are non-perturbative from the string perspective. This removes some of the freedom in constructing models because Eq. (51) is harder to satisfy. Nonetheless, the real difficulty in finding examples of such vacua is in constructing the $SU(n)$ gauge bundle. This is the problem we will turn to in the following sections.

Before we do so, let us briefly recall the first and most straightforward approach to finding solutions: the standard embedding.[11] The point is that, it is easy to show that, since X Calabi–Yau, the spin-connection ω, viewed as a $SU(3)$ satisfies the HYM equations. Thus the existence of X itself naturally supplies us with suitable $SU(3)$ gauge bundle. In the "standard embedding" model one then chooses an $SU(3)$ subgroup of the E_8 gauge connection to be equal to the spin connection. From the description of GUT groups in Sec. 2, this implies that the preserved GUT group is E_6. The other E_8 gauge bundle V_2 is assumed to be trivial. The anomaly condition (41) is automatically satisfied since $c_2(V_1) = c_2(TX)$ identically. No fivebranes are present and so the model has a good heterotic string theory description. The three-family condition (49) then gets related to the geometry of X since $c_3(V_1) = c_3(TX) = \chi(X)$ where the Euler characteristic $\chi(X)$ is a topological invariant given by $\chi(X) = 2h^{2,1} - 2h^{1,1}$ where $h^{p,q}$ are the Hodge numbers.

5. The HYM equation on T^2, T-duality and the Fourier–Mukai transform

The main difficulty in building semi-realistic models is constructing the supersymmetric gauge bundle V. Our approach will be to follow the work of Friedman, Morgan and Witten,[4] Donagi[5] and Bershadsky *et al.*[6] and build bundles on elliptically fibred Calabi–Yau manifolds X. These are manifolds which are T^2 fibrations over a four-dimensional base. The point is that constructing suitable bundles on each T^2 fibre is relatively simple. Bundles are then built on the full space by repeating the construction on T^2 fibre by fibre.

In the following two sections we will turn to the geometry of elliptically fibred Calabi–Yau manifolds and the full construction of V. Here as a warm up exercise, we start by considering how to construct solutions to the HYM equations on T^2 and how these correspond to stable holomorphic bundles. We give a simple interpretation of the construction in terms of T-duality of D-branes. Finally, we formalize this action of T-duality on bundles, or as it will turn out more generally sheaves, in terms of the Fourier–Mukai transform.

5.1. *Solutions to the HYM equations*

Consider solutions to the HYM equations (28) for a $U(n)$ gauge field on a two-torus T^2. For convenience let us assume we have a square torus with coordinates $z = x + iy$ and the periodic identification $x + R_1 \sim x$ and $y + R_2 \sim y$. Clearly the holomorphic conditions $F_{zz} = F_{\bar{z}\bar{z}} = 0$ are satisfied identically, while $ig^{z\bar{z}}F_{z\bar{z}} = 0 \Leftrightarrow F_{z\bar{z}} = 0$. In other words on a two-dimensional space the HYM equations simply reduce to looking at flat connections. Since F vanishes, the only non-trivial configurations are Wilson lines, measuring the holonomy around non-trivial loops on T^2 (see Appendix A).

More formally, the holonomy of the Wilson lines is a map $\mathrm{Hol}(V) : \pi_1(T^2) \to U(n)$ associating an element of $U(n)$ to each homotopy class of loops. On a torus, there are two basic non-trivial loops γ_1 and γ_2, corresponding to looping around the x-direction and the y-direction. A general loop wraps n times around γ_1 and m times around γ_2. Thus the fundamental group is Abelian, $\pi_1(T^2) = \mathbb{Z} \oplus \mathbb{Z}$. This means $\mathrm{Hol}(V)$ must be map into an Abelian subgroup of $U(n)$, that is a subgroup of the "maximal torus" $U(1)^n$ built out of the diagonal matrices. In fact, permutating the elements of the diagonal matrices gives a conjugate element of $U(n)$, (this is the Weyl group) thus the holonomy is really specified by a set of n $U(1)$ Wilson lines up to permutations. This then characterizes the distinct solutions of the HYM equations.

What is the space of these Wilson lines? For a given $U(1)$ factor, we simply need to specify the values of the Wilson lines in $U(1)$ around γ_1 and γ_2. This means giving two phases $\mathrm{e}^{\mathrm{i}\alpha_1}$ and $\mathrm{e}^{\mathrm{i}\alpha_2}$. In other words we have to give a point q on a two-torus, $\tilde{T}^2 = S^1 \times S^1$. (Mathematically, for any Riemann surface S, this space of flat $U(1)$ Wilson lines is known as the Jacobian $J(S)$ of S and is always a complex torus.) For the full $U(n)$ case we then have

$$U(n) \text{ HYM solution on } T^2 \Leftrightarrow n \text{ unordered points } q_i \text{ on } \tilde{T}^2 \tag{53}$$

Note also that one point e on $\tilde{T}^2$ is always picked out, namely that corresponding to the connection with trivial holonomy. For a $SU(n)$ bundle we have the condition that the determinant of the matrix representing the holonomy must be zero. In other words all the phases in the diagonal matrix must sum to zero, which implies that the sum of all the points on $\tilde{T}^2$ must equal the zero point e

$$SU(n) \text{ solution:} \quad q_1 + \cdots + q_n = e \tag{54}$$

There is also a natural metric on space of flat $U(1)$ connections given by

$$g(\Delta A, \Delta A) = \frac{1}{\mathrm{vol}(T^2)} \int_{T^2} \Delta A \wedge * \Delta A, \tag{55}$$

where $\Delta A = A_1 - A_2$ is the difference of two flat connections and is gauge invariant. Explicitly we can always choose a gauge where A is constant $A = a\,\mathrm{d}x + b\,\mathrm{d}y$. The metric is then just flat $g = \Delta a^2 + \Delta b^2$. Locally we have $A = \mathrm{d}\chi$ with $\chi = ax + by$. Since for $U(1)$ gauge transformations $\chi \sim \chi + 2\pi$, we have $0 \leq a < 2\pi/R_1$ and $0 \leq b < 2\pi/R_2$ giving the lengths of the sides of $\tilde{T}^2$ using the metric (55). Thus we see that, in string theory language, $\tilde{T}^2$ is actually the *T-dual torus* to T^2. (For of review of T-duality see for instance Polchinski.[28])

5.2. *D-branes and T-duality*

The appearance of the T-dual torus has a very natural explanation in string theory in terms of D-branes. We are really just reproducing a very standard result[28,29] that the T-dual of n D2-branes wrapping T^2 is n D0-branes on $\tilde{T}^2$.

Recall that the field theory for a collection of n overlapping Dp-branes is given by the Dirac–Born–Infeld (DBI) action (see for instance Polchinski's book[28]),

$$S = -T_p \int_{M_{p+1}} \mathrm{e}^{-\phi} \operatorname{tr}\left[-\det\left(G_{mn} + 2\pi\alpha' F_{mn} + B_{mn}\right)\right]^{1/2} + S_{\mathrm{WZ}}, \qquad (56)$$

where T_p is the tension of the brane and we will turn to the Wess–Zumino term W_{WZ} in a moment. The action is essentially the volume of the n Dp-branes modified by the presence of a $U(n)$ gauge field F their worldvolume M_{p+1}. Fixing a particular configuration of the branes that minimizes the volume, the effective action for fluctuations is simply, to leading-order in α', $U(n)$ super-Yang–Mills (SYM) coupled to a set of scalars giving the transverse fluctuations in position of the brane.

As supersymmetric Green–Schwarz action,[30] the DBI theory has two fermionic symmetries, just like the M5-branes action in Sec. 3.3. One is the usual supersymmetry, the other the local κ-symmetry on the branes. Requiring the brane configuration is supersymmetric, in some background with a Killing spinor η then reduces to a projection condition[31] on η, analogous to (37). If we consider a D2-brane wrapped on T^2 in the supersymmetric space $\mathbb{R}^{7,1} \times T^2$, then we find that preserving supersymmetry implies that the SYM gauge field satisfies

$$F_{mn}\gamma^{mn}\epsilon = \lambda\epsilon, \qquad (57)$$

where λ is a constant, and γ^m are gamma matrices on the D2-brane. Note this is just the usual condition (26) for a supersymmetric SYM gauge field configuration, modulo the factor λ, which is a result of the κ-symmetry, which is realized as a non-linear supersymmetry in SYM. This implies on T^2

$$F_{ab} = F_{\bar{a}\bar{b}} = 0, \qquad ig^{a\bar{b}}F_{a\bar{b}} = \lambda\mathrm{id}. \qquad (58)$$

We get precisely the generalized HYM equations (32) for a $U(n)$ field.

More generally, for a set of Dp-branes wrapping a curve C on more general supersymmetric space, the conditions of supersymmetry, to leading order in α' imply[31]

(1) C is a *calibrated* cycle, for instance a holomorphic cycle on a Calabi–Yau manifold X,
(2) F satisfies the analog of the HYM equations. For a holomorphic cycle these are just (58).

If one keeps all orders in α', the HYM equations generalise,[31] with F in the second equation of(58), replaced by some polynomial in F, equal to F at leading order. Mathematically this suggests an interesting generalization of the HYM equations, and the existence of solutions corresponding to a modified notion of stability.

Returning to the T^2 example, we see that the conditions for a supersymmetric D2-brane match our original HYM equations (28), provided we take $\lambda = 0$, or equivalently to $\operatorname{tr} F = 0$. This has a natural interpretation in terms of D-branes. To see this we turn to the Wess–Zumino term. Using requirements of anomaly cancellation,[32] it has been argued that for n coincident D-branes we have

$$S_{\mathrm{WZ}} = T_p \int_{M_{p+1}} \operatorname{tr} \mathrm{e}^{2\pi\alpha' F} \sqrt{\frac{\hat{\mathrm{A}}(TM)}{\hat{\mathrm{A}}(NM)}}\, C, \tag{59}$$

where TM is the tangent bundle and NM is the normal bundle to the brane in the ten-dimensional spacetime and $\hat{\mathrm{A}}$ is a characteristic class (see Appendix H). Each term is a power series in forms

$$\begin{aligned} \operatorname{tr} \mathrm{e}^{2\pi\alpha' F} &= n + 2\pi\alpha' \operatorname{tr} F + \frac{1}{2}(2\pi\alpha')^2 \operatorname{tr} F \wedge F + \dots, \\ \hat{\mathrm{A}}(W) &= 1 + \frac{4\pi^2\alpha'^2}{24 \cdot 8\pi^2} \operatorname{tr} R \wedge R + \dots, \\ C &= C_1 + C_3 + \dots, \end{aligned} \tag{60}$$

where C_i are RR potentials (in type IIA in this case) and R is the field strength for a real bundle W. The idea is to expand the power series and take the top $p+1$-form and then integrate. Note the $\operatorname{tr} \mathrm{e}^{2\pi\alpha' F}$ terms are essentially the Chern character $\operatorname{ch}(V)$ for the $U(n)$ bundle. Thus we see that the Dp-brane is a source for RR flux with the sources written in terms of characteristic classes. In general the brane is a source of all RR fluxes corresponding to the Dp-branes and lower dimensional branes[d].

Again specializing to our very simple case of D2-branes, we have

$$S_{WZ} = T_2 \int_{\mathbb{R}\times T^2} nC_3 + 2\pi\alpha' \operatorname{tr} F \wedge C_1. \tag{61}$$

We see that $\operatorname{tr} F$ is a source of D0-brane charge, since it couples to C_1. This implies we can view flux $\operatorname{tr} F$ on the T^2 as counting the number of embedded D0-branes in our D2-branes. Note that in the more general case given in Eq. (59), the terms corresponding to $\mathrm{ch}_k(V)$ (which are powers of F) are source of D$(p-2k)$-brane charge. But so also are terms from the Â-polynomials. Thus curved background geometry also generates an effective D-brane charge.

From (58), we see that $\lambda = 0$ implies $\operatorname{tr} F = 0$ and so we have a supersymmetric collection of n D2-branes with no D0-branes. Now it is a standard result[28] that under T-duality exchanges Dirichlet and Neumman boundary conditions. Thus dualizing in both directions of the torus any D2-branes transform to D0-branes and vice versa. Thus $\lambda = 0$ translates into the condition that the T-dual configuration contains only D0-branes on the dual torus $\tilde{T}^2$.

The T-duality symmetry means that the moduli space of n supersymmetric D2-branes on T^2 (with $\operatorname{tr} F = 0$) must match the moduli space of n D0-branes on $\tilde{T}^2$. But this is exactly what we found: the space of solutions to the HYM was precisely n points on $\tilde{T}^2$.

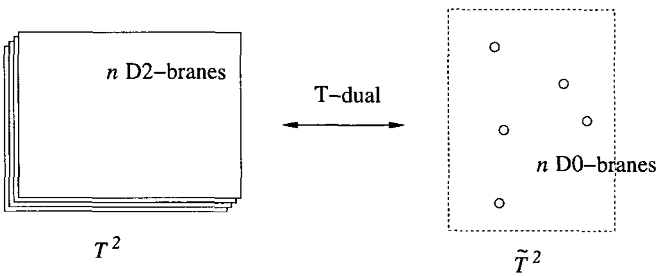


Figure 3. HYM solutions on T^2

[d]If we have included the transverse scalar fluctuations, we would also have sources for higher-dimensional branes, the so-called "Myers effect".[33]

5.3. *Holomorphic description of solutions*

In the last section we argued that solutions to the HYM equations are in one-to-one correspondence with poly-stable holomorphic vector bundles. It is useful to see how this correspondence works in this simple case. It will also lead us to a mathematical formulation of the T-duality transformation just discussed.

The key point here is that while the HYM equations $U(n)$ imply that as a real vector bundle V is trivial, as a holomorphic vector bundle this is not the case. The different Wilson lines correspond to distinct holomorphic vector bundles. The idea is as follows. First the same holonomy argument as above implies that any flat holomorphic bundle on T^2 reduces to a set of n line bundles. Then recall that having a holomorphic line bundle is equivalent to giving a complex structure on the total, in this case trivial, bundle $T^2 \times \mathbb{C}$. If $w = u + \mathrm{i}v$ is a local complex coordinate on the fiber, this means we have to give how we identify w as we move around the two closed loops in T^2. This is defined by a phase rotation on the coordinate $w \to e^{\mathrm{i}\alpha} w$. But this is nothing more than the holonomy of the $U(1)$ bundle we had before, that is a map from $\pi_1(T^2)$ to S^1. To define the complex structure we just have to give two α_i for the phase rotation about the two γ_i cycles. Since these give different complex structures they correspond to different holomorphic bundles.

Thus the (flat) holomorphic line bundle is specified by two phases, or equivalently a point q in $\tilde{T}^2$. We write

$$L_q\text{: flat line bundle on } T^2 \text{ are labelled by } q \in \tilde{T}^2$$

Again for a $SU(n)$ bundle the sum of all α_1 and α_2 phases for the n line bundles must be zero. In other words, as before we see that the sum of all the points on $\tilde{T}^2$ must equal the zero point e, corresponding to the trivial line bundle. So we have

$$V = L_{q_1} \oplus \cdots \oplus L_{q_n} \quad \text{with} \quad q_1 + \cdots + q_n = e \tag{62}$$

What though about our second condition that the holomorphic bundle must be poly-stable? First, we note that we have already argued V splits into the sum of line bundles, so the condition becomes that each line bundle L must be stable. But this is trivially satisfied: L has no sub-bundles of lower rank, so all line bundle are necessarily stable. Thus in this case, we get no additional condition from stability.

5.4. *The Fourier–Mukai transform*

Now we come to the obvious questions: is there a more formal mathematical description of the T-duality procedure? In general we might imagine starting with some rank n bundle V, say with $c_1(V) = m \neq 0$ so that we have a mixture of n D2- and m D0-branes:

$$V : \{\mathrm{rk}(V) = n, c_1(V) = m\} \quad \Leftrightarrow \quad \tilde{V} : \{\mathrm{rk}(\tilde{V}) = m, c_1(\tilde{V}) = n\} \qquad (63)$$

Is there then some way of generating the dual bundle $\tilde{V}$ with m D2-branes and n D0-branes? This is not just a map of charges but actually of points in the moduli space of configurations: for each configuration of branes given by a stable homomorphic bundle V we assign a T-dual configuration given by a stable holomorphic bundle $\tilde{V}$. Such a map does in fact exist and is known as the *Fourier–Mukai* transform.[18,19,35–37] It actually defines a duality for any holomorphic configuration, so that one can drop the stability part of the supersymmetry conditions.

Let us start by considering the duality between pure D2-branes and D0-branes discussed above. Clearly the D0-branes are not described by holomorphic bundles since their gauge fields are localized at points in $\tilde{T}^2$. Thus to define the T-duality map we need to generalize our notion of a holomorphic vector bundle to include these pure D0-brane configurations. The appropriate generalization, as was first suggested by Harvey and Moore,[34] is to *sheaves*. These are discussed in a little more detail in Appendix I. The basic point is that sheaves are defined by their group of sections rather than their fibres. These means that dimension of the fibres (or "stalks") can change as one moves over X. A pure D2-brane on T^2 is a flat holomorphic line bundle. This can equally well be defined as a sheaf in terms of the set of holomorphic sections of the line bundle. A D0-brane looks like a point in T^2 and its sections correspond to functions in T^2 which are zero everywhere except at one point q. This is known as the "skyscraper" sheaf $\mathcal{O}_q$ of a point.

We can now define the action of T-duality on sheaves. More technically, we actually want to consider *coherent sheaves* since these are the objects that have a notion of stability, generalizing the notion of a stable bundle. In other words, these are the objects which describe supersymmetric brane configurations. (See Appendix I for a definitions and discussion.)

The first step is to define a four-dimensional space $F = T^2 \times \tilde{T}^2$. (Here both T^2 and $\tilde{T}^2$ have one special point identified, the zero point under addition. There are equivalent as complex manifolds.) Then we can introduce[4,37] the so-called Poincaré line bundle $\mathcal{P}$. This has the property that restricting to $T^2 \times q \cong T^2$ where q is a point on $\tilde{T}^2$ or to $q \times \tilde{T}^2$ we get the line bundle L_q on either T^2 or $\tilde{T}^2$. In other words

$$\mathcal{P}|_{T^2 \times q} \cong \mathcal{P}|_{q \times \tilde{T}^2} \cong L_q. \tag{64}$$

As discussed in Appendix G, there is a correspondence between line bundles and divisors. In fact we find

$$\mathcal{P} = \mathcal{O}(D - e \times \tilde{T}^2 - T^2 \times e), \tag{65}$$

where D is the diagonal divisor built by taking the same points in T^2 and $\tilde{T}^2$, that is $q \times q$, for all $q \in T^2$.

We have two projections $\pi : F \to T^2$ and $\tilde{\pi} : F \to \tilde{T}^2$. We can then define the *Fourier–Mukai* transform $V = \boldsymbol{FM}(\tilde{V})$, which takes a sheaf $\tilde{V}$ on $\tilde{T}^2$ to a sheaf V on T^2, by

$$V = \boldsymbol{FM}(\tilde{V}) \equiv \pi_*(\tilde{\pi}^* \tilde{V} \otimes \mathcal{P}). \tag{66}$$

Here $\tilde{\pi}^*$ pulls sheaves from $\tilde{T}^2$ back to F while π_* pushes them forward from F to T^2 (see Appendix I).

To see how this map works, consider a single D0-brane at point q on $\tilde{T}^2$. This is described by the skyscraper sheaf $\tilde{V} = \mathcal{O}_q$. This pulls back to the sheaf $\tilde{\pi}^* \mathcal{O}_q = \mathcal{O}_{T^2 \times q}$, that is the structure sheaf of $T^2 \times q$. Now we tensor with $\mathcal{P}$. Since $\mathcal{O}_{T^2 \times q}$ is supported only on $T^2 \times q$, we have $\mathcal{F} \equiv \tilde{\pi}^* \mathcal{O}_q \otimes \mathcal{P}$ is also supported only on $T^2 \times q$. Since $\mathcal{O}_{T^2 \times q}$ on $T^2 \times q$ is the trivial bundle, $\mathcal{F}$ on $T^2 \times q$ is just the line bundle $\mathcal{P}|_{T^2 \times q} = L_q$. Finally we push forward to T^2. This process is defined in Eq. (I.1). By definition, the section of $\pi_* \mathcal{F}$ above a set $U \subset T^2$ is just the group of sections of $\mathcal{F}$ over $U \times \tilde{T}^2 \in F$. But this is just the set of sections of $L_q(U)$. Thus we have, as required, that the dual of a single D0-brane is a single D2-brane is the line bundle L_q:

$$\boldsymbol{FM}(\mathcal{O}_q) = L_q. \tag{67}$$

If we have a set of D0-branes we have $\mathcal{O}_{q_1 + \cdots + q_n}$. The transform acts as before except now, $\mathcal{F} \equiv \tilde{\pi}^* \tilde{V} \otimes \mathcal{P}$ is supported on $T^2 \times \sum_i q_i \subset F$. On each submanifold $T^2 \times q_i$, the restriction $\mathcal{F}|_{T^2 \times q_i}$ is the line bundle L_{q_i}. Pushing forward, we get sections from all n line bundles so the result is just $V = L_{q_1} \oplus \cdots \oplus L_{q_n}$ giving n D2-branes on T^2 as required.

Showing the reverse transformation from D2-branes to D0-branes $\boldsymbol{FM}(L_q) = \mathcal{O}_q$ is more involved. In particular, we have to work harder to understand the "push-forward" procedure. Consider the simplest case where we have a single D2-brane with a trivial Wilson line. We then have $\tilde{V} = L_e = \mathcal{O}_{\tilde{T}^2}$, the trivial bundle. Pulling back to F, we just get the trivial bundle on F, that is $\tilde{\pi}^*\mathcal{O}_{\tilde{T}^2} = \mathcal{O}_F$. Thus $\mathcal{F} \equiv \tilde{\pi}^*\tilde{V} \otimes \mathcal{P} = \mathcal{P}$. Finally then we need to push $\mathcal{P}$ forward to T^2. By construction, over a given point $q \in T^2$ we have $\mathcal{P}|_{q\times\tilde{T}^2} = L_q$. Thus our stalks for $\pi_*\mathcal{P}$ will be holomorphic sections of L_q. Generically, L_q has no holomorphic sections. The exception is when $q = e$ and L_q is the trivial bundle $\mathcal{O}_{\tilde{T}^2}$. Then any constant section is holomorphic. Thus is would appear that $\pi_*\mathcal{P}$ has zero stalk everywhere except at $e \in T^2$ where the stalk is a copy of $\mathbb{C}$. Thus it appears that $V = \mathcal{O}_e$ as expected. More generally, starting with L_q we would find $\mathcal{F}|_{q'\times\tilde{T}^2} = L_{q'-q}$ and then

$$\boldsymbol{FM}(L_q) = \mathcal{O}_q. \tag{68}$$

In fact this is not quite correct. If one is more careful in the definition of π_* one finds[36,37] that, as defined, $\boldsymbol{FM}(\tilde{V}) = 0$ for any vector bundle $\tilde{V}$. In practise one, instead has to modify the notion of push-forward. The idea is to define $\pi_*\mathcal{F}$ at each point $p \in T^2$ in terms of cohomology of the restriction $\mathcal{F}|_{p\times\tilde{T}^2}$ of the sheaf on F to the elliptic curve $p \times \tilde{T}^2$ above p. The usual push-forward corresponds to the zeroth order cohomology, the group of holomorphic functions. More generally one can also take first-order cohomology, the group of harmonic one-forms with values in $\mathcal{F}|_{\pi^{-1}(p)}$. In practice we will always use the transform starting from the D0-branes so this will subtlety will not actually be an issue.

Note that the Fourier–Mukai transform only required a complex structure on T^2. No mention of metric was ever made. Normally, T-duality is only defined when we have a Killing vector, which depends on a choice of metric. From this point of view, the Fourier–Mukai transform is more general than ordinary T-duality and can be defined on spaces without a Killing vector.

There is a further subtlety. We would expect that acting with $\boldsymbol{FM}$ twice, as for T-duality, to bring us back to where we started. It turns out that acting only on sheaves it is not possible to define such an operator. The solution[36,37] is to act instead on the *derived category* of coherent sheaves D^b. The objects on which $\boldsymbol{FM}$ acts are now no longer individual sheaves but rather a *complex* of sheaves

$$\dots \xrightarrow{T_3} \mathcal{F}_2 \xrightarrow{T_2} \mathcal{F}_1 \xrightarrow{T_1} \mathcal{F}_0 \xrightarrow{T_0} \dots, \tag{69}$$

that is a set of sheaves $\{\mathcal{F}_i\}$ together with maps T_i satisfying $T_{i+1}T_i = 0$. Two complexes are taken to be equivalent if the homology groups $H_i = \ker T_i / \operatorname{im} T_{i+1}$ are the same. (For more detail there are good introductions by Thomas[38] and Sharpe.[39]) On these objects a functor $\boldsymbol{FM}$ can be defined which squares to the identity (up to some signs).

It is interesting to note that there is a very natural physical interpretation of the objects in the derived category: they are simply combinations of branes and anti-branes with the T_i describing the low-energy fields for strings stretching between them. In particular, between a brane and an anti-brane these are tachyons. This is the picture due to Sen[40] that we can realize all branes as the result of annihilation of the highest dimension branes. As Witten showed[41] the brane charges are then elements in K-theory. The notion of a derived category captures a little more, in that it encodes the particular brane configuration not just the charge and was suggested by a number of authors.[42] What is interesting is that, in this context, the mathematics of T-duality is leading us to realize the branes in this annihilation picture. Derived categories also appears in formulations of mirror symmetry,[43] and understanding the exact relation to D-branes, in particular, supersymmetric configurations is an active area of research (see recent lectures by Douglas[44] and Sharpe[39] and the references therein).

In summary, we have seen that $SU(n)$ solutions of the HYM equations on T^2 are most naturally understood in terms of a T-dual picture. Here the original solutions correspond to n D2-branes wrapping T^2. Under T-duality these map to n D0-branes on a dual torus $\tilde{T}^2$. The space of solutions is then easy to define simply as the positions of the n D0-branes on $\tilde{T}^2$. Formally these give a sheaf in $\tilde{T}^2$. Mathematically the action of T-duality from D0- to D2-branes is described by the Fourier–Mukai transform.

6. Elliptically fibred Calabi–Yau manifolds

In the last section we saw how stable $SU(n)$ bundles could be built on T^2 by taking the Fourier–Mukai transform. Physically, this corresponding to constructing a set of D2-branes as the T-dual of a set of D0-branes. Our goal is to extend this construction to build bundles on a Calabi–Yau threefolds X. To do so we need X to have a fibred structure, where the fibres are two-tori, so that we can then make a T-duality transformation fibre by fibre. In fact, we will consider tori with one point marked. (This is the zero point e discussed in the last section.) Such marked tori are called *elliptic curves*. The purpose of this section is to discuss the geometry of such "elliptically fibred" Calabi–Yau manifolds.

We should note that these geometries are also central to F-theory constructions.[45,46] The basic F-theory duality is that the heterotic string compactified on $\mathbb{R}^{7,1} \times T^2$ is dual to a twelve-dimensional F-theory geometry $\mathbb{R}^{7,1} \times K$, where K is a four-dimensional elliptically fibred K3 manifold. (Recall K3 is the unique compact Calabi–Yau twofold.) The fibration is over a two-sphere $\mathbb{C}P^1$, that is $K \to \mathbb{C}P^1$. The twelve-dimensional space is not really a string geometry: there are for instance no Kähler modulus or Kaluza–Klein degrees of freedom on the elliptic fibre. Instead, one should think of these geometries as type IIB string backgrounds of the form $\mathbb{R}^{7,1} \times \mathbb{C}P^1$, with the elliptic fibres describing non-trivial IIB scalar fields. Only the complex structure of the fibre matters, and gives the value of the IIB complex field $\tau = C_0 + \mathrm{i}e^{-\phi}$, where ϕ is the dilaton and C_0 the Ramond–Ramond scalar. The type IIB $SL(2,\mathbb{Z})$ S-duality is then just a symmetry of the complex structure modulus of the torus.[45] In this picture, the heterotic string gauge degrees of freedom arise from degenerations of the fibrations where the elliptic fibre becomes singular. At these points a cycle shrinks to zero size and a new degrees of freedom can appear. (In the type IIB picture these are D7-branes.) Finally one can extend the duality[45,46] by considering the heterotic string compactified on a a Calabi–Yau n-fold X which is itself elliptically fibred over some base B. By considering the duality fibrewise this is dual to F-theory on a Calabi–Yau $n+1$-fold which is a K3 fibration over B, where the K3 fibres are themselves each elliptically fibred. It was this duality which motivated the original papers on Fourier–Mukai constructions of supersymmetric bundles on X.[4,5,6]

An elliptically fibred Calabi–Yau three-fold X consists of a base B, which is a complex two-surface, and a map

$$\pi : X \to B \tag{70}$$

with the property that for a generic point $b \in B$, the fiber

$$E_b = \pi^{-1}(b) \tag{71}$$

is an elliptic curve, with a marked "zero point" e. The set of zero points then form a global section σ. This is a map

$$\sigma : B \to X \tag{72}$$

that assigns to every point $b \in B$ the zero element $\sigma(b) = e \in E_b$. (Note we will use the same notation σ both for the map and for the image of B in X under the map.) This structure is depicted in Fig. 4. The Calabi–Yau three-fold must be a complex Kähler manifold. This implies that the base B is itself a complex manifold, while we have already assumed that the fiber is a Riemann surface and so has a complex structure. Furthermore, the

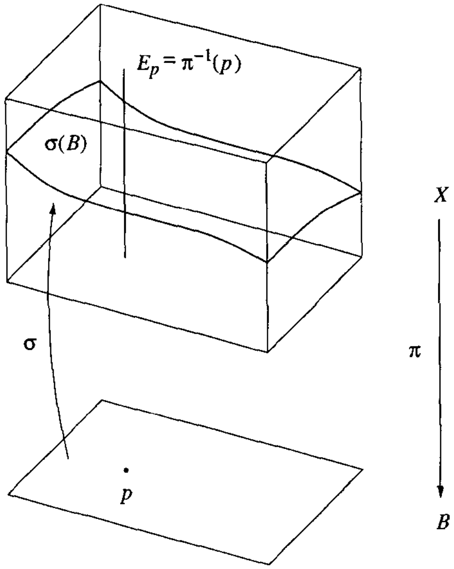


Figure 4. Elliptically fibred X

fibration must be holomorphic, that is, it must have holomorphic transition functions and both maps π and σ must be holomorphic. The condition that the Calabi–Yau three-fold has vanishing $c_1(TX) = 0$ will further constrain the types of fibration allowed as we will see.

6.1. *Elliptic curves and the Weierstrass equation*

Let us start by briefly summarizing the properties of an elliptic curve E. It is a genus one Riemann surface and can be embedded in two-dimensional complex projective space $\mathbb{C}P^2$. A simple way to do this is by using the homogeneous Weierstrass equation

$$zy^2 = 4x^3 - g_2xz^2 - g_3z^3 \tag{73}$$

where x, y and z are complex homogeneous coordinates on $\mathbb{C}P^2$ so that we identify $(\lambda x, \lambda y, \lambda z) \sim (x, y, z)$ for any non-zero complex number λ. The parameters g_2 and g_3 encode the different complex structures one can put on the torus. Note that independent of g_2 and g_3, the curve always passes through the point $(0, 1, 0)$. This is the zero point e on the curve.

Provided $z \neq 0$, we can rescale to affine coordinates where $z = 1$. We then see, viewed as a map from x to y, that there are two branch cuts in the

x-plane, linking $x = \infty$ and the three roots p, q and r of the cubic equation $4x^3 - g_2 x - g_3 = 0$ in two pairs. This is shown in Fig. 5 which also shows the generic form of the Weierstrass equation in the (x, y)-plane. The torus is then a two-fold cover of the x-plane. The two classes of non-trivial loops on the torus are those surrounding a branch cut, and those linking the two branch cuts.

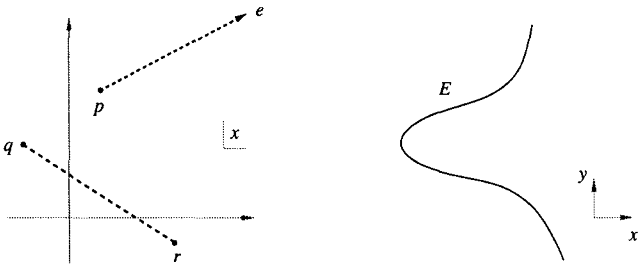


Figure 5. Wierstrass model of T^2

When any two roots coincide, the elliptic curve becomes singular. This corresponds to one of the cycles in the torus shrinking to zero. This is shown in Fig. 6. The torus has become a sphere, intersecting itself at a point. Although singular as a manifold, this space still makes sense as a complex variety since it is defined by a homogeneous polynomial in $\mathbb{C}P^2$.

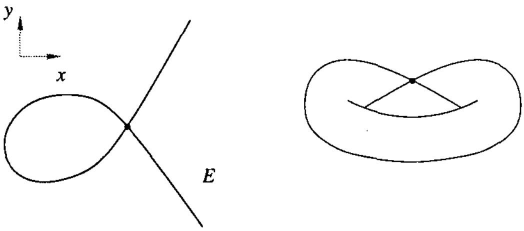


Figure 6. Single-pinch singularity

If two pairs of roots come together the singularity is even worse. The torus becomes a pair of spheres touching at two points. This is shown in Fig. 7. In the (x, y)-plane the spheres corresponds to a line and an ellipse. Such singular behaviour, where any of the roots coincide, is characterized by the discriminant of the cubic $4x^3 - g_2 x - g_3 = 0$, namely,

$$\Delta = g_2^3 - 27g_3^2 \tag{74}$$

vanishing.

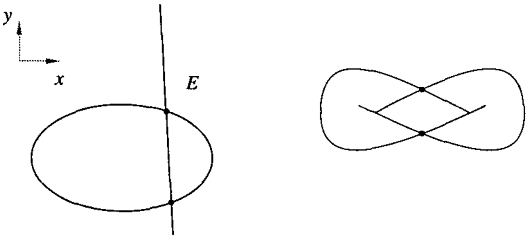


Figure 7. Double-pinch singularity

A more familiar way of describing T^2 is as the complex plane $\mathbb{C}$ modulo a lattice Λ corresponding to periodic identification in two directions. In this picture there is a natural notion of addition of points inherited from the underlying complex plane, $w' = w + a$ for $w \in \mathbb{C}$. It would be nice to relate this to the picture of T^2 as a cubic curve in $\mathbb{C}P^2$ we have just discussed. This is a famous correspondence in algebraic geometry, first discussed by Abel, and central to the evaluation of elliptic integrals. There is a map from $\mathbb{C}/\Lambda$ to $\mathbb{C}P^2$ given by

$$w \mapsto (\mathcal{P}(w), \mathcal{P}'(w), 1), \tag{75}$$

where $\mathcal{P}(w)$ is the *Weierstrass function.* It is an even function, depending, on two parameters, with a double pole at $w = 0$, with the expansion $\mathcal{P}(w) = w^{-2} + aw^2 + bw^4 + \dots$. The complex structure of T^2 depends on the parameters a and b, with the simple relation $g_2 = 20a$ and $g_3 = 28b$. Note that the identity point under addition, $w = 0$, maps to the identified point $e = (0, 1, 0)$. The condition that three points satisfy $w_1 + w_2 + w_3 = 0$ then maps to the condition that the images p_1, p_2 and p_3 are collinear.

6.2. *Elliptic fibrations and* $\mathrm{c}_1(TX) = 0$

The elliptic fibration is defined by giving the elliptic curve E over each point in the base B. If we assume the fibration has a global section, and in this paper we do, then on each coordinate patch this requires giving x and y and the parameters g_2 and g_3 in the Weierstrass equation as complex functions on the base. Globally, these all become sections of appropriate line bundles on B since there may be some none trivial patching of E as we move around the base.

We might expect that there is really only one relevant line bundle, that encoding the twisting of the direction in the T^2 fibre normal to the section. Near the section we have some coordinate w describing the position on the elliptic curve. Let $\mathcal{L}^{-1} \equiv \mathcal{N}_{\sigma/X}$ be the normal bundle describing the twisting of the coordinate. The section intersects each fibre at the point $p = (0, 1, 0)$. Near p we can solve locally $x \sim 1/w^2$ and $y \sim 1/w^3$ with $z = 1$. (This can also been seen directly from the Weierstrass function in (75).) Thus we have that x and y are sections of $\mathcal{L}^2$ and $\mathcal{L}^3$ respectively, while z is just a section of the trivial bundle. By $\mathcal{L}^i$ we mean the tensor product of the line bundle $\mathcal{L}$ with itself i times.

Similarly, from the form of the Weierstrass equation (73) we have that g_2 and g_3 should be sections of $\mathcal{L}^4$ and $\mathcal{L}^6$. Thus, we see that the elliptic fibration is characterized by a line bundle $\mathcal{L}$ over the base B together with a choice of sections g_2 and g_3 of $\mathcal{L}^4$ and $\mathcal{L}^6$.

Note that the set of points in the base over which the fibration becomes singular is given by the vanishing of the discriminant $\Delta = g_2^3 - 27g_3^2$. It follows from the above discussion that Δ is a section of the line bundle $\mathcal{L}^{12}$. The zeros of Δ then naturally define a divisor (see Appendix G), which in this case is a complex curve, in the base. This is known as the *discriminant curve.*

There are a number of different types of singularity over the discriminant curve depending on how the fibre degenerates. The type of singularity is controlled order of the zero of Δ on the discriminant curve, together with the order of the zeros of g_2 and g_3 if they also vanish there.[47] In the case where the base is a one-fold, there is a classical result due to Kodaira[48,47] classifying all the different possible types of degeneration. In this case the discriminant curve is just a set of $12\, c_1(\mathcal{L})$ of points. A given singularity looks like a singular pinched elliptic curve. Locally each pinched point is actually just an orbifold singularity of $\mathbb{C}^2$. These were discussed in Steve Gubser's lectures,[49] and are characterized by the A_n, D_n and E_n extended Dynkin diagrams. In our case, we get a family of the same type of Kodaira singularities over each disconnected component of the discriminant curve. The situation becomes more complicated at points where the curve intersects itself.

It might seem strange that we are talking about singularities given that our Calabi–Yau manifold is supposed to be smooth. The point is that this construction only really gives a *model* of X. We must resolve all the singularities by "blowing them up" to get the actual smooth X. The point is that this is a well-defined procedure, so that given the Weierstrass model we really know everything we need to know about X. (The procedure of how locally to resolve the A-D-E singularities was again discussed in Steve Gubser's lectures.[49])

Finally, we come to the important condition that on a Calabi–Yau three-fold X the first Chern class of the tangent bundle T_X must vanish or equivalently $\mathrm{Hol}(X) = SU(3)$. Now, recall that this means that the overall $U(1)$ part of the $U(n)$ holomorphic tangent bundle T_cX must be trivial. For the tangent bundle this is related to the so called *canonical bundle* K_X. Let T_c^*X be the holomorphic part of the complexified cotangent bundle. We then define on a $2n$-dimensional manifold

$$K_X = \underbrace{T_c^*X \wedge \cdots \wedge T_c^*X}_{n \text{ times}}. \tag{76}$$

Thus sections of K_X are holomorphic $(n,0)$-forms, locally given by

$$\phi = \phi(z_1, \ldots, z_n) \mathrm{d}z^1 \wedge \cdots \wedge \mathrm{d}z^n. \tag{77}$$

Equivalently, locally we simply give a single function $\phi(z_1, \ldots, z_n)$ which globally becomes a section of line bundle. Thus effectively K_X as a line bundle which precisely measures the twisting of the determinant of the T_c^*X or equivalently the overall $U(1)$ factor of T_cX. In particular, it is easy to see that $\mathrm{c}(TX) = -\mathrm{c}_1(K_X)$. It is then clear that the condition that X is Calabi–Yau implies that K_X is trivial,

$$K_X = \mathcal{O}. \tag{78}$$

We would now like to understand what this implies for our fibration. Recall that, restricting to the section, we can pull back the cotangent bundle on X to B, giving

$$T^*X|_\sigma = T^*B \oplus N^*_{\sigma/X}. \tag{79}$$

In other words, the X cotangent bundle is the combination of directions within B and directions in the conormal bundle $N^*_{\sigma/X}$. From the definition (76) of K_X, we immediate get an *adjunction* formula, true for any divisor σ,

$$K_B = K_X|_\sigma \otimes N_{\sigma/X} \tag{80}$$

where by $K_X|_\sigma$ we mean the restriction of the canonical bundle K_X to σ. Since K_X is trivial, its restriction is also trivial. Thus we find the relation

$$\mathcal{L}^{-1} \equiv N_{\sigma/X} = K_B. \tag{81}$$

The Calabi–Yau condition implies that the twisting of the normal bundle to σ must be related to the twisting of the determinant of the tangent bundle of B so that the twisting of the determinant of the tangent bundle of X is trivial. The fact that the elliptic curves are flat tori means that K_X is necessarily trivial when restricted to any fiber and ensuring that $K_X|_\sigma = 0$ is sufficient to imply that $K_X = 0$.

Nonetheless this is not quite sufficient to ensure we have an elliptically fibered Calabi–Yau manifold. In particular one requires that K_B^{-4} and K_B^{-6} must admit sections for g_2 and g_3 respectively. Also there are constraints that the singularities of the Weierstrass model (and the base) cannot be too extreme if we are going to be able to blow up to give a smooth X. This for instance imposes further restrictions on how the curves where the g_2 and g_3 sections vanish are allowed to intersect. There is theorem[46,50]

> **Theorem:** For X to be a smooth Calabi–Yau manifold the only allowed bases B are
>
> (1) "Enriques" surface E,
> (2) "Hirzebruch" surface H_n, or a blow-up thereof,
> (3) "del Pezzo" surface dP_n.

The structure of the del Pezzo surfaces is summarized in Appendix J and the other surfaces are described for instance in Donagi *et al.*[17] Here we simply note that all these bases have $b^{2,0} = 0$. It will turn out that the Enriques surface cannot be used for building three-family models.

6.3. *Algebraic classes on X*

Recall that we will need to know the Chern classes $c_2(TX)$ and $c_1(V)$, $c_2(V)$ and $c_3(V)$ if we are to check that we can satisfy the conditions given in Sec. 4.3. Since all the bundle are holomorphic, these can all be expressed in terms of algebraic classes (see Appendix E and Appendix H). Thus it is useful to have expressions for the algebraic classes in X. In general, the full set of classes will depend on the particular fibration in question. However, there is a generic set of classes which are always present, independent of the fibration, and this is what we will concentrate on. Recall that, for a Calabi–Yau threefold, the fact that we have $h^{2,0}$ means that the homology groups of algebraic classes are actually the same as the integer homology groups $H_{2p}(X,\mathbb{Z})$.

Two of the homology groups are, as always, very simple: $H_0(X,\mathbb{Z})$ is just the class of points on X. All points are homologous so $H_0(X,\mathbb{Z})=\mathbb{Z}$. Similarly $H_6(X,\mathbb{Z})$ is just the class of the manifold X itself (or multiple copies thereof) and so $H_6(X,\mathbb{Z})=\mathbb{Z}$.

Now consider $H_2(X,\mathbb{Z})$ and $H_4(X,\mathbb{Z})$, the groups of homologous classes of holomorphic curves and surfaces respectively. Simply because we have an elliptic fibration, the fiber E_p at any given point is a holomorphic curve in X. Consequently, one algebraic class in $H_2(X,\mathbb{Z})$ which is always present is the class of the fiber, which we will call F. Similarly, the existence of a section σ means there is also a holomorphic surface in X. The class of the section, which, abusing notion, we will call σ, defines an algebraic class in $H_4(X,\mathbb{Z})$. We have

$$
\begin{aligned}
F &\in H_2(X,\mathbb{Z}) \quad \text{class of elliptic fiber } E_p,\\
\sigma &\in H_4(X,\mathbb{Z}) \quad \text{class of section } \sigma.
\end{aligned}
$$

(Note that formally we are now writing σ for three different things: the section map $\sigma: B\to X$, the image of the map (the section itself), and now the homology class $[\sigma]$ of the section. Hopefully what is meant will be clear from context!)

Additional algebraic classes may also be inherited from the base B. In general, B has a set of algebraic classes, which for the B in question, since $b^{2,0}=0$, span $H_2(B,\mathbb{Z})$. These can lead to classes in X in two different ways. First, the existence of the section σ means we can lift a holomorphic curve C in B to a curve $\sigma(C)$ in X. If α is the class of C in B then we write $\sigma_*\alpha$ (the "push-forward") for the class of $\sigma(C)$ in X. Second, we can use the projection map π to pull α back to a class in $H_4(X,\mathbb{Z})$. For a given representative C, one forms the fibered surface $\pi^{-1}(C)$ over C. The homology class of this surface in $H_4(X,\mathbb{Z})$ is then denoted by $\pi^*\alpha$. Thus we have,

$$
\begin{aligned}
\sigma_*\alpha &\in H_2(X,\mathbb{Z}) \quad \text{push-forward of class } \alpha\in H_2(B,\mathbb{Z}),\\
\pi_*\alpha &\in H_4(X,\mathbb{Z}) \quad \text{pull-back of class } \alpha\in H_2(B,\mathbb{Z}).
\end{aligned}
$$

This structure is indicated in Figure 8.

In general, these maps may have kernels. For instance, two curves which are non-homologous in B, might be homologous once one embeds them in the full Calabi–Yau threefold. In fact, we will see that this is not the case. One way to show this, which will be useful in the following, is to calculate

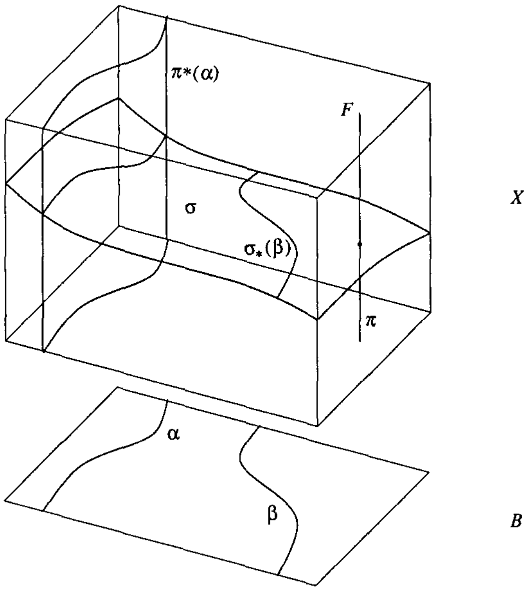


Figure 8. Algebraic classes on X

the intersection numbers between the classes in $H_2(X,\mathbb{Z})$ and $H_4(X,\mathbb{Z})$. We find

$$\begin{array}{cc|cc} & & \multicolumn{2}{c}{H_2(X,\mathbb{Z})} \\ & & \sigma_*\beta & F \\ \hline H_4(X,\mathbb{Z}) & \pi^*\alpha & \alpha\cdot\beta & 0 \\ & \sigma & -\,c_1(B)\cdot\beta & 1 \end{array} \tag{82}$$

where the entries in the first column are the intersections of $H_2(B,\mathbb{Z})$ classes in B. (Note from here on we write $c_p(B)$ for the Chern class of the base $c_p(TB)$.) The terms in (82) are all straightforward except for $\sigma\cdot\sigma_*\beta = -\,c_1(B)\cdot\beta$. Note that in general we can think of class of curves $\sigma_*\beta$ as an intersection between two classes of surfaces σ and $\pi^*\beta$. Thus $\sigma\cdot\sigma_*\beta = \sigma\cdot\sigma\cdot\pi^*\beta$. Below we will argue that $\sigma\cdot\sigma = -\sigma_*\,c_1(B)$ and hence $\sigma\cdot\sigma_*\beta = -\,c_1(B)\cdot\beta$.

Given that $H_{2p}(X,\mathbb{Z})$ and $H_{2n-2p}(X,\mathbb{Z})$ are dual as vector spaces under intersection, two classes are equivalent if they have the same intersection numbers. If we take a set of classes α_i which from a basis of $H_2(B,\mathbb{Z})$, we see that the matrix of intersection numbers of the form given in (82) is non-degenerate. Thus, for each nonzero $\alpha\in H_2(B,\mathbb{Z})$, we get nonzero classes $\sigma_*\alpha$ and $\pi^*\alpha$ in $H_2(X,\mathbb{Z})$ and $H_4(X,\mathbb{Z})$.

To complete the intersection ring of algebraic classes we need the intersection pairing on surfaces $H_4(X,\mathbb{Z}) \times H_4(X,\mathbb{Z}) \to H_2(X,\mathbb{Z})$. We find

$$\begin{array}{cc|cc} & & \multicolumn{2}{c}{H_4(X,\mathbb{Z})} \\ & & \pi^*\beta & \sigma \\ \hline H_4(X,\mathbb{Z}) & \pi^*\alpha & (\alpha\cdot\beta)F & \sigma_*\alpha \\ & \sigma & \sigma_*\beta & -\sigma_*\,\mathrm{c}_1(B) \end{array} \tag{83}$$

Again all the entries are straight-forward except for $\sigma\cdot\sigma$. This can be calculated in the following way. Near the section σ we can model X by the total space of normal bundle $N_{\sigma/X}$ over B. The section σ is then just the zero section of the line bundle $N_{\sigma/X}$. Recall that we also have the line bundle $\mathcal{O}(\sigma)$ which has holomorphic sections vanishing on σ (see Appendix G). Formally we find that the restriction of $\mathcal{O}(\sigma)$ to σ is just the normal bundle $N_{\sigma/X}$, that is, we have another *adjunction formula*,[71] true for any divisor σ,

$$\mathcal{O}(\sigma)|_\sigma = N_{\sigma/X}. \tag{84}$$

Any other holomorphic section σ' of $N_{\sigma/X}$ represents another nearby surface in the same homology class as σ. The two sections intersect when $\sigma' = 0$. From the discussion of Chern classes the homology class of these zeros in B is, by definition $\mathrm{c}_1(N_{\sigma/X})$. Lifting to X, we then have from Eq. (84)

$$\sigma\cdot\sigma = \sigma_*\,\mathrm{c}_1(N_{\sigma/X}). \tag{85}$$

In our case the Calabi–Yau condition relates $N_{\sigma/X} = K_B$ (see Eq.(81)), and so $\sigma\cdot\sigma = -\sigma_*\,\mathrm{c}_1(B)$.

As we mentioned, the algebraic classes we have identified so far are generic, always present independently of the exact form of the fibration. There are two obvious sources of additional classes. Consider $H_4(X,\mathbb{Z})$. First, we could have additional sections non-homologous to the zero-section σ. Second, the pull-backs of irreducible classes on B could split so that $\pi^*\alpha = \Sigma_1 + \Sigma_2$. This splitting comes from the fact that there can be curves on the base over which the elliptic curve degenerates, for example, into a pair of spheres. New classes appear from wrapping the four-cycle over either one sphere or the other. Now consider $H_2(X,\mathbb{Z})$. We see that the possibility of degeneration of the fiber means that the fiber class F can similarly split, with representatives wrapped, for instance, on one sphere or the other. Finally, the presence of new sections means there is a new way to map curves from B into X and, in general, classes in $H_2(B,\mathbb{Z})$ will map under the new section to new classes in $H_2(X,\mathbb{Z})$.

In all our discussions in this paper, we will ignore these additional classes. This will mean for instance that our bundles and fivebrane wrappings will not be in general complete. However, this restriction will allow us to analyze generic properties. It should be noted however that in constructed more sophisticated models, in particular, those with standard model gauge groups, it is often imperative to consider non-generic situations where extra classes appear.[18]

6.4. *Effective classes and* $c_2(TX)$

Recall that one of our conditions for model building (see Sec. 4.3) was that the homology class of the fivebranes $[W]$ was effective. Quite generically we can write

$$[W] = \sigma_* \Omega + fF. \tag{86}$$

where $f \in \mathbb{Z}$ and $\Omega \in H_2(B, \mathbb{Z})$. Now clearly, if Ω is a effective class in B, say with a representative curve $C_2 = \sum_i a_i C_2^{(i)}$ with $a_i > 0$, then $\sigma_* \Omega$ is an effective curve in X, since it has an effective representative $C_2' = \sum_i a_i \sigma(C_2^{(i)})$. Similarly, for fF to represent an effective curve we need $f > 0$. Thus we have the sufficient conditions

$$[W] \geq 0 \quad \Leftarrow \quad \Omega \geq 0 \text{ and } f \geq 0, \tag{87}$$

where we use the notation $[W] \geq 0$ to mean the class is effective. In general, there might be additional classes with $f < 0$ and $\Omega \neq 0$ which are still effective. In most situations this is not the case.[17] For our purposes Eq. (87) will be sufficient.

Our last problem is that, in order to discuss the anomaly cancellation condition, we will need $c_2(TX)$. Friedman, Morgan and Witten[4] show that it can be written in terms of the Chern classes of the holomorphic tangent bundle of B as

$$c_2(TX) = 12\sigma_* c_1(B) + (c_2(B) + 11\, c_1(B) \cdot c_1(B))\, F. \tag{88}$$

We will not reproduce the calculation in detail here but give a rough explanation of the terms. Recall that $c_2(TX)$ is an element of $H_2(X, \mathbb{Z})$, thus it must have the form

$$c_2(TX) = \sigma_* \alpha + kF, \tag{89}$$

for some $k \in \mathbb{Z}$ and $\alpha \in H_2(B, \mathbb{Z})$. Now recall that $TX|_\sigma = TB \oplus N_{\sigma/X}$. From the Whitney sum formula in Appendix H, and given $c_1(N_{\sigma/X}) =$

$-\,c_1(B)$, we have

$$c_2(TX|_\sigma) = c_2(B) - c_1(B)\cdot c_1(B). \tag{90}$$

Now, as an element in cohomology, we can calculate $c_2(TX|_\sigma)$ just by wedging $c_2(TX)$ with a delta-function two-form $\delta(\sigma)$ localized on σ. In homology, this simply corresponds to intersecting with the class of σ. Thus we also have, from (89) and the intersection matrix (82)

$$c_2(TX|_\sigma) = c_2(TX)\cdot\sigma = k - c_1(B)\cdot\alpha. \tag{91}$$

We can also do the same calculation, restricting to an elliptically fibred surface $\Sigma = \pi^{-1}(C)$ where C is a curve in B. We then get the analogous expression to Eq. (90) but with σ replaced with Σ. (For any divisor D in a Calabi–Yau manifold the $K_X = \mathcal{O}$ condition means we always have $c_1(N_{D/X}) = -\,c_1(D)$.) By intersecting with the discriminant curve Δ which is a section of $\mathcal{L}^{12}$, we see that Σ has $12\,c_1(B)\cdot[C]$ singular fibres (here $[C]\in H_2(B,\mathbb{Z})$ is the homology class of C in B). As discussed these give local orbifold singularities. In the simplest case, the singular fibres are all distinct and each is a A_1 singularity.[47] This is a gravitational instanton[49,63] with unit charge, so the net contribution to $c_2(\Sigma)$ is $12\,c_1\cdot[C]$. One can easily show[46] that $c_1(\Sigma)$ is proportional to the fibre F in Σ and so $c_1(\Sigma)\cdot c_1(\Sigma) = 0$. Thus we have

$$c_2(TX|_\Sigma) = 12\,c_1(B)\cdot[C] = \alpha\cdot[C], \tag{92}$$

where the last line is just $c_2(TX|_\Sigma)$ calculated from the general form (89) given $[\Sigma] = \pi^*[C]$. Finally from Eqs. (90), (91) and (92), given that α is arbitrary, we find

$$\alpha = 12\,c_1(B), \qquad k = c_2(B) + 11\,c_1(B)\cdot c_1(B) \tag{93}$$

and we reproduce the expression (88).

7. Construction of V and an $SU(5)$ GUT example

In the Sec. 5 we saw that the easiest way to describe solutions of the $SU(n)$ HYM equations on T^2 was in terms of the "T-dual" configuration as a set of D0-branes on $\tilde{T}^2$. In the previous section we then constructed elliptically fibred Calabi–Yau manifolds X with the goal that we could then build bundles on X by doing a fibrewise T-duality.

In this section, we complete the program by building the bundles explicitly and give one example with $SU(5)$ GUT group. Finally we briefly discuss how to construct more detailed models with standard model gauge groups as well as more general properties of Hořava–Witten braneworld backgrounds.

7.1. *Stability and FM on X*

Let us start by defining what we mean by the T-dual manifold. Recall that for T^2, the dual space was the Jacobian $J(T^2)$ built out of the space of Wilson lines on T^2. Given the fibred structure of X, we can construct the "associated Jacobian bundle" $\tilde{X} \equiv J(\pi)$ by taking the Jacobian $J(E_p)$ of each elliptic fiber over a point p in the base. Thus we have

$$X \leftrightarrow \tilde{X} \equiv J(\pi) \cong X \tag{94}$$

by the action of T-duality on each fiber.

Just as on the torus we can define T-duality by acting with the Fourier–Mukai transform fibrewise. Of course, the transformation is more subtle at the points where the fibration is singular, but it can nonetheless be defined. First we need to define the analogs of the space $F = T^2 \times \tilde{T}^2$ and the Poincaré bundle $\mathcal{P}$. The analog of Y is the fiber product $F = X \times_B \tilde{X}$. This is a space of complex dimension four. It is fibred over B, the fiber over $b \in B$ being the ordinary product $E_b \times \tilde{E}_b$ of the two fibers. There are two projections: $\pi : F \to X$ and $\tilde{\pi} : F \to \tilde{X}$. Now, the Poincare bundle $\mathcal{P}$ is determined by the following two properties

(1) Restricting to one or other elliptic fibre with $q \in \tilde{E}_B$ or $q \in E_b$, we have

$$\mathcal{P}|_{E_b \times q} \cong \mathcal{P}|_{q \times \tilde{E}_b} \cong L_q,$$

where L_q is the flat line bundle defined in Sec. 5.4,

(2) The restriction of $\mathcal{P}$ to $\sigma \times_B \tilde{X}$ or $X \times_B \sigma$ is the trivial bundle.

Explicitly we find $\mathcal{P} = \mathcal{O}_F(D - \sigma \times_B \tilde{X} - X \times_B \sigma) \otimes \pi^* \tilde{\pi}^* K_B$. The extra factor depending on K_B as compared to the T^2 case is required to ensure the second condition above. As for T^2, D is the diagonal divisor built from $q \times_B q$ for all $q \in E_b$.

We then define the Fourier–Mukai transform as before as a functor

$$V = \boldsymbol{FM}(\tilde{V}) \equiv \pi_*(\tilde{\pi}^* \tilde{V} \otimes \mathcal{P}). \tag{95}$$

As in the T^2 case, technically there are issues in defining the right push-forward π_* and the transform really only acts on derived categories of sheaves, but these subtleties will not concern us here.

In general $\boldsymbol{FM}$ is just a map between holomorphic sheaves which we can think of as T-duality acting on a set of D-branes. What we really want is the requirement that V is poly-stable. On T^2 this implied that $\tilde{V}$ was just the sheaf corresponding to n points on $\tilde{T}^2$. The real question is what is the constraint on $\tilde{V}$ now if V is stable?

Fibrewise the answer is clear. We know that if $\tilde{V}|_{\tilde{E}_b}$ is again just a set of n points, then $V|_{E_b}$ must be stable. Extending this smoothly over the base we see that a fibrewise stable V is equivalent to giving how the n points on $\tilde{E}_b$ vary as one moves around the base B. Globally they must define a complex surface C which is an n-fold cover of B as shown in Fig. 9.

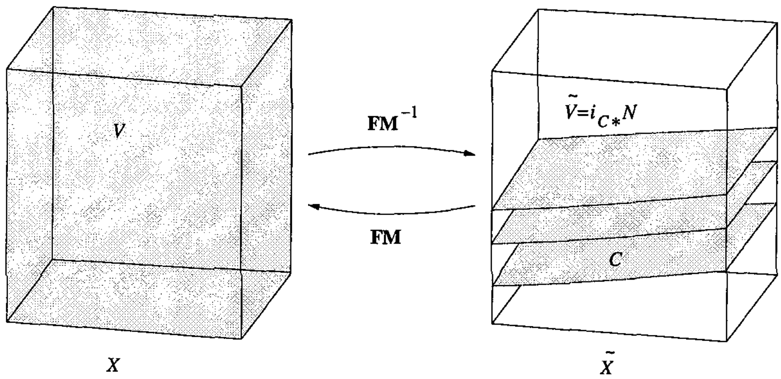


Figure 9. The Fourier–Mukai T-duality transform

Physically the surface C represents a D4-brane. This is reasonable since we recall that in the T^2 case, our original bundle corresponded to a set of D2-branes and was T-dual to a set of D0-branes. Now on our six-dimensional space X, the original bundle is a set of n D6-branes wrapping X, and taking T-duality in the fibre directions gives a set of n D4-branes. In fact, generically, one actually gets, rather than a set of branes, a single D4-brane wrapping C which is an n-fold cover of the base. For a single D4-brane there will also a line-bundle $\mathcal{N}$ on C describing the $U(1)$ gauge fields (and in general describing embedded D2- and D0-brane charge). There is an inclusion map

$$i_C : C \to \tilde{X} \tag{96}$$

describing the embedding of C in $\tilde{X}$. We can try and use this map to push the bundle on C into an object on X. As for the D0-branes in the T^2

example, the resulting object $i_{C*}\mathcal{N}$ is a actually a sheaf supported on C with the restriction $i_{C*}\mathcal{N}|_C = \mathcal{N}$. Formally we have

$$\tilde{V} = i_{C*}\mathcal{N}. \tag{97}$$

Thus is appears that to describe a fibrewise stable bundle V, the dual data is a D4-brane. That is

(1) a divisor C of $\tilde{X} \cong X$ which is an n-fold cover of the base B, known as the *spectral cover*,
(2) a line bundle $\mathcal{N}$ on C.

This pair $(C, \mathcal{N})$ is known as the *spectral data.*

As stands, though, we are not quite done. While V is fibrewise stable it is not clear that it is stable as a bundle on X. Essentially we are only solving the restriction of the HYM equations. Here though there is a very useful result which is at the heart of the spectral cover construction.[4,5,6] Provided we choose the Kähler metric on X suitably, fibrewise stability is sufficient. The point is to ensure that the dominant contribution to the slope (33) comes from the slope on the fibre. This implies that the size of the fibre directions must be chosen to be small compared with the size of the base. Stability depends in a discrete way on the Kähler form ω, thus this condition is not a limiting condition where the ratio the size of the fibre to the size of the base must go to zero, but rather only be smaller that some definite number, which in examples is not so small.

Note that the distinction between poly-stable and stable bundles corresponds to whether the hermitian Yang-Mills field strength is reducible or not. This refers to whether, globally, it can be diagonalized into parts coming from different subgroups of the full gauge group. More precisely, it refers to whether or not the holonomy commutes with more that just the center of the group. This can mean that if the bundle is only poly-stable then the GUT gauge group may be larger than expected. In the spectral cover picture it exactly corresponds to the question of whether the spectral cover C is a single surface or decomposes into more than on surface. Thus strictly one must check whether, given a particular set of spectral data, C is a single surface or a sum of surfaces. This will actually lead to an additional condition on the class of C, first derived in the context of toric manifolds X by Berglund and Mayer[51] and also discussed by Rajesh.[52]

Finally then, we have a definite procedure for constructing stable holomorphic bundles V on a large class of Calabi–Yau manifolds X.

7.2. *The spectral data and Chern classes*

To complete our construction we need to calculate the Chern classes of V in terms of the spectral data $(C, \mathcal{N})$ and in order to see if we can satisfy the conditions given in Sec. 4.3. From the Fourier–Mukai transform we have, for the Chern character,

$$\mathrm{ch}\, V = \mathrm{ch}(\pi_*(\tilde{\pi}^* i_{C*}\mathcal{N} \otimes \mathcal{P})), \tag{98}$$

There are several push-forward and pull-back maps in this expression. However, there is a standard procedure for calculating the corresponding Chern classes. For pull-backs one integrates the class (as an element in cohomology) over the fibers of the map, while for push-forwards one uses the Grothendieck–Riemann–Rock theorem.[53] Here we will simply give the results. More details can be found, in particular, in the original paper of Friedman, Morgan and Witten.[4]

Since C is an n-fold cover of the base, in order to have an $U(n)$ gauge group, in terms of the classes given in Sec. 6.3, the class $[C]$ of C must be given by

$$[C] = n\sigma + \pi^*\eta, \tag{99}$$

where $\pi^*\eta$ is the pull-back into X of some class η in the base B. Since C is an actual surface in $\tilde{X}$, we have the addition condition that η is effective on B, that is $\eta \geq 0$.

Actually there is a stronger condition on η. Recall that to ensure that the GUT group H really is the commutant of $SU(n)$ and not a larger group we must ensure that the spectral cover is a single surface. If instead it splits into more that one surface, the bundle is poly-stable, that is of the form of a product of $U(n_i)$ factors with $\sum n_i = n$, and the commutant can be larger than H. Now if C is a single analytic surface, its intersection with any other single analytic surface must be some actual positive curve (note this need not be the case for self-intersections). Consider intersecting C with σ. From Eq. (83), we have

$$[C] \cdot \sigma = \sigma_*\eta - n\sigma_*\, \mathrm{c}_1(B). \tag{100}$$

For all our base manifolds $\mathrm{c}_1(B)$ is an effective class (see for instance Appendix J). Thus, clearly a necessary condition[51,52] for C not to split is

$$\eta - n\,\mathrm{c}_1(B) \geq 0 \tag{101}$$

In fact, by considering all such intersections, one can shown that this is sufficient.

One can then calculate $c_1(V)$. One finds

$$c_1(V) = i_{C*}\left(c_1(\mathcal{N}) + \tfrac{1}{2}c_1(C) - \tfrac{1}{2}p^* c_1(B)\right) \tag{102}$$

where we have introduced the projection map

$$p : C \to B. \tag{103}$$

Since we want an $SU(n)$ bundle we require $c_1(V) = 0$. This requires that

$$c_1(\mathcal{N}) = \tfrac{1}{2}c_1(C) + \tfrac{1}{2}p^* c_1(B) + \gamma, \tag{104}$$

where γ is a class on C with the property that,

$$p_*\gamma = 0 \tag{105}$$

as an element of $H_2(B,\mathbb{Z})$.

To understand the generic solution for γ and also the form of $c_1(C)$ we need to known the generic $H_2(C,\mathbb{Z})$ classes. Clearly we can build curves in C by resticting divisors in X to C. In terms of classes this gives us two generic types of curve

$$\sigma|_C \equiv S, \qquad \pi^*\alpha|_C = p^*\alpha \tag{106}$$

where $\alpha \in H_2(B,\mathbb{Z})$. Of course other classes may also exist, but these are the generic cases. The generic solution for γ is then given by

$$\gamma = \lambda\left(nS - p^*\eta + np^* c_1(B)\right) \tag{107}$$

where λ is a rational number, since as a class in X, using the intersections (83), we have

$$\begin{aligned} i_{C*}\gamma &= \lambda(n\sigma - \pi^*\eta + n\pi^* c_1(B)) \cdot [C] \\ &= -\lambda\eta \cdot (\eta - n\,c_1(B))F \end{aligned} \tag{108}$$

so $p_*\gamma = \pi_* i_{C*}\gamma = 0$. Appropriate values for λ will emerge shortly.

Using the adjunction formulae Eqs. (80) and (84) for C, together with Eq. (99), $K_X = 0$ and $c_1(C) = -c_1(K_C)$, we find,

$$c_1(C) = nS + p^*\eta. \tag{109}$$

Combining the equations (104), (107) and (109) yields

$$c_1(\mathcal{N}) = n\left(\tfrac{1}{2} + \lambda\right)S + \left(\tfrac{1}{2} - \lambda\right)p^*\eta + \left(\tfrac{1}{2} + n\lambda\right)p^* c_1(B) \tag{110}$$

Essentially, this means that the bundle $\mathcal{N}$ is completely determined in terms of the elliptic fibration and η. It is important to note, however, that there is not always a solution for $\mathcal{N}$. The class $c_1(\mathcal{N})$ must be integral, a condition that constrains on the allowed λ and η. Since S and $p^*\alpha$ are distinct classes on C, we cannot choose η to cancel the S term and, hence, the coefficient of S must, by itself, be an integer. This implies that a consistent bundle $\mathcal{N}$ will exist if either

$$\begin{aligned} &n \text{ is odd}, \quad \lambda = m + \frac{1}{2}, \\ &n \text{ is even}, \quad \lambda = m, \quad \eta = c_1(B) + 2\alpha, \end{aligned} \tag{111}$$

where m is an integer and $\alpha \in H_2(B, \mathbb{Z})$. Thus when n is even, we cannot choose η arbitrarily. These conditions are only sufficient for the existence of a consistent line bundle $\mathcal{N}$ but are enough for the example we consider in the next section. Note that other ways of ensuring $\mathcal{N}$ is integral do exist. We could, for example take $n = 4$, $\lambda = \frac{1}{4}$ and $\eta = 2\,c_1(B) = 4\alpha$, or $n = 5$, $\lambda = \frac{1}{10}$ and $\eta = 5\alpha$.

Finally let us give the remaining Chern classes for the $SU(n)$ vector bundle V. Friedman, Morgan and Witten[4] first calculated $c_2(V)$, while Curio[54] and Andreas[55] gave $c_3(V)$. The results are

$$\begin{aligned} c_1(V) &= 0, \\ c_2(V) &= \sigma_*\eta - \Big[\tfrac{1}{24}\,c_1^2(B)\left(n^3 - n\right) \\ &\qquad\qquad - \tfrac{1}{2}\left(\lambda^2 - \tfrac{1}{4}\right) n\eta \cdot (\eta - nc_1(B))\Big] F, \\ c_3(V) &= 2\lambda\eta \cdot (\eta - nc_1(B)) . \end{aligned} \tag{112}$$

We can now translate the conditions for GUT models with three families of matter given in Sec. 4.3 into conditions on our spectral data $(C, \mathcal{N})$. We need to find η, λ, such that

(1) **GUT model:** We need an $SU(n)$ bundle so $c_1(V) = 0$, implying

$$\begin{aligned} &n \text{ is odd}, \quad \lambda = m + \frac{1}{2}, \\ &n \text{ is even}, \quad \lambda = m, \quad \eta = c_1(B) + 2\alpha, \end{aligned} \tag{113}$$

We also require, so that the spectral cover C does not split and the bundle is truly $SU(n)$, the condition

$$\eta - n\,c_1(B) \geq 0. \tag{114}$$

(2) **Cohomological constraint:** We need the net fivebrane charge to vanish (51). Given Eq. (88) for $c_2(TX)$, Eq. (87) for $[W]$ and

Eq. (112) for $c_2(V)$, we have, for $[W]$ to be effective

$$\begin{aligned} \Omega &= 12\,c_1(B) - \eta \geq 0, \\ f &= c_2(B) + \left[11 + \tfrac{1}{24}\left(n^3 - n\right)\right] c_1^2(B) \\ &\qquad + \tfrac{1}{2}\left(\lambda^2 - \tfrac{1}{4}\right) n\eta \cdot (\eta - nc_1(B)) \geq 0. \end{aligned} \tag{115}$$

(3) **Three families:** Finally we need $c_3(V) = 6$ for three families of matter. Given Eq. (112), we have

$$\lambda\eta \cdot (\eta - nc_1(B)) = 3. \tag{116}$$

7.3. *An $SU(5)$ example*

Let us now look for an explicit example of a bundle where we can solve the conditions for anomaly cancellation and three families. This will be similar to the examples in Donagi *et al.*.[16,17]

We will consider the case where $H = SU(5)$ is our GUT gauge group. It then follows from (11) that we must choose the structure group of the gauge bundle also to be $SU(5)$, so

$$H = SU(5), \qquad G = SU(5), \tag{117}$$

and, hence, $n = \mathrm{rk}(V) = 5$.

Next we take the base B to be a del Pezzo surface

$$B = dP_3. \tag{118}$$

The properties of these surfaces are summarized in Appendix J. In particular, there is a basis for $H_2(dP_8, \mathbb{Z})$ composed given by l and e_i, e_2 and e_3 where

$$l \cdot l = 1, \qquad l \cdot e_i = 0, \qquad e_i \cdot e_j = -\delta_{ij}. \tag{119}$$

Furthermore we have

$$\begin{aligned} c_1(B) &= 3l - e_1 - e_2 - e_3, \\ c_2(B) &= 6. \end{aligned} \tag{120}$$

The classes l and e_i are effective, but there are also other effective classes not obtainable as a linear combination of l and e_i with non-negative integer coefficients. In particular, $c_1(B)$ is effective and in addition so are $l - e_1 - e_2$, $l - e_2 - e_3$ and $l - e_3 - e_1$ (see Appendix J). Adding these last three, we have

$$d \equiv 3l - 2(e_1 + e_2 + e_3) > 0, \qquad c_1(B) > 0. \tag{121}$$

We now must find spectral data, specified by η and λ, such that the model building conditions (113)–(116) are satisfied. Since n is odd, condition (113) tells us that $\lambda = m + \frac{1}{2}$ for integer m. Let us take

$$\begin{aligned} \eta &= 5\,c_1(B) + 6e_1 + e_2, \\ \lambda &= -\tfrac{3}{2} \end{aligned} \tag{122}$$

Since $c_1(B)$ and e_i are effective, we clearly have $\eta - 5\,c_1(B) > 0$ and so condition (114) is satisfied, and C does not split.

Now we just go ahead and calculate the other conditions. First we find that the three family condition (116) is satisfied. Then we find, since d is effective, for $[W] = \sigma_*\Omega + fF$,

$$\begin{aligned} \Omega &= 7d + e_1 + 6e_2 + 7e_3 > 0, \\ f &= 112 > 0. \end{aligned} \tag{123}$$

Thus the cohomological constraint (115) is satisfied.

Finally we have found an explicit model. It is relatively easy to find a range of other solutions both on dP_3 and for other surfaces and with other gauge groups. In hunting for solutions it soon becomes clear that the most difficult constraint is to keep $\Omega \geq 0$ and $\eta - n\,c_1(B) \geq 0$. Given the relation that $\Omega = 12\,c_1(B) - \eta$, these really translate into

$$n\,c_1(B) \leq \eta \leq 12\,c_1(B). \tag{124}$$

with $n = 3$, 4 or 5. This is relatively constraining, in particular, it does not seem to be possible to have $SU(5)$ models with dP_r for $r > 3$. (Note that in Donagi *et al.*,[16,17] examples on dP_8 and dP_9 were found. However, in fact these violate the lower, no-splitting bound on η.) Finally, note that the Enriques surface can never give models with three families.[17] The essential point is that the canonical bundle K_B is a torsion class for the Enriques surface: $2K_B = 0$. Then careful analysis of the condition (124) implies the only possibility is n is odd and $\eta = K_B$. But then $c_3(V) = 0$.

7.4. *Hořava–Witten braneworlds*

Let us end this section by briefly discussing some of the possible extensions of the models discussed here and well as properties of heterotic M-theory brane-worlds generally. The list is by no means exhaustive.

GUT models and low-energy effective theory

Thus far we have only considered building models with GUT gauge groups. Using the techniques discussed here it is quite possible to construct a variety of models.[16,17] There has not been any complete *classification of examples*, considering all possible bases and GUT groups. Nonetheless, given the bound (124), there is a finite list of possibilities which could be worked through. One definite result, as we have seen, is that no three-family models exist with an Enriques base.

The structure of the *low-energy effective theory* arising in four- and five-dimensions after compactification on X was one of the main topics of Burt Ovrut's lectures.[1] I will not expand on it here other than the very broad summary that, at least in four dimensions the effective theory is, essentially, the same in the M-theory and weak heterotic limits.[13] The only new ingredient is the presence of fivebranes, which introduces new moduli entering, for instance, the gauge kinetic functions.[10,56] Various authors have investigated supersymmetry breaking via gaugino condensation in the E_8 groups,[57] and again in the four-dimensional theory the breaking mechanism is just as in the weak limit.

$SU(3) \times SU(2) \times U(1)$ models

The next step is obviously to construct models with the standard model gauge group. Although originally considered in a heterotic M-theory context, any constructions are equally applicable to the weak heterotic string. Models have been built[13] following a very standard approach of starting with a GUT theory and then breaking the GUT group with a Wilson line.[14] For example, given an $SU(5)$ GUT group, a Wilson line of the form $g_\gamma = \text{diag}(1, 1, 1, -1, -1)$ is invariant only under $SU(3) \times SU(2) \times U(1) \subset SU(5)$, and so breaks to the standard model. As we discuss in Appendix A, this requires $\pi_1(X) \neq 0$. In particular, since g_γ generates $\mathbb{Z}_2$ we need $\mathbb{Z}_2 \subset \pi_1(X)$. Unfortunately, one can show that all elliptically fibred Calabi–Yau threefolds are simply connected (except for those with an Enriques base, but these cannot give three families of matter). Thus is appears that the Fourier–Mukai construction cannot be used.

The solution is an old one:[58] one looks for elliptically fibered threefolds Z which admit a $\mathbb{Z}_2$ involution with no fixed points. This is a map τ which maps Z to itself, which squares to the identity and has no points on Z such that $\tau(p) = p$. One can then form the quotient manifold $X = Z/\tau$ formed by identifying p with $\tau(p)$. If $\pi_1(Z) = 0$ then $\pi_1(X) = \mathbb{Z}_2$. (A good example is S^2 under the antipodal map.) If we have a bundle V_Z on X which is invariant under τ, then it descends to a bundle on X.

Thus the problem now becomes one of identifying elliptically fibred

Calabi–Yau manifolds Z which admit such involutions and also the conditions on the spectral data on Z for V_Z to be invariant τ. For τ to be fixed point free, one finds that the section σ cannot map to itself, instead Z must admit a second section ζ such that $\tau(\sigma) = \zeta$. This means the elliptic fibration must specialize. For general bases, one can find many examples of spectral data which at least leads to V_Z with invariant Chern classes.[18] However, this is not sufficient for V_Z to be invariant. To ensure the bundle itself is invariant one needs to know the explicit action of the Fourier–Mukai transform, on sheaves themselves rather than just the classes. This requires going to a very specific Calabi–Yau threefold which is a double elliptic fibration, that is as a fibre product $B \times_{\mathbb{C}P^1} B'$ where B and B' are each elliptic fibrations over the same base $\mathbb{C}P^1$. After a certain amount of work (one has to construct V_Z as a sort of twisted product, or *extension* of two simpler bundles) one finds a large class of models.[19]

Recently, this procedure has been extended[59,60] to the $\mathbb{Z}_2 \times \mathbb{Z}_2$ Wilson lines which can be used to break $SO(10)$ GUT groups to the standard model together with an additional $U(1)_{B-L}$ symmetry which can be useful for suppressing nucleon decay.[58,38] These papers also raise an additional distinction, that between invariant and equivariant bundles. It is strictly only the more constrained equivariant bundles which descend to $X = Z/\tau$. Invariant bundles descend to twisted objects, where the twisting is related to a closed two-form. This is natural in a D-brane picture since it translates into the Neveu–Schwarz two-form B_{MN} which is known to couple to the brane (see for instance Eq. (59)). The meaning of the twisting in the heterotic context remains open. Again though, many invariant models can be found.

Fivebranes, small instantons and non-perturbative potentials

There are a number of interesting phenomena which are characteristic only of heterotic M-theory models. First as we have mentioned is the presence of M5-branes. These give a number of new moduli and the moduli spaces of the branes can be calculated.[61] In general these can be quite complicated with branching points corresponding to degenerate configurations where extra light degrees of freedom can appear. One such degeneration is where a fivebrane touches one of the fixed planes. It is then possible to have a small instanton transition[26] where the brane becomes absorbed into the fixed plane, its magnetic charge being taken up by a new contribution to $c_2(V_i)$. Such transitions are particularly interesting because they can change the

$c_3(V_i)$ index. In other words, the number of light families can change dynamically. Using the Fourier–Mukai construction, these transitions can be followed explicitly.

Another aspect of interest is that one expects that non-perturbative corrections may lift the moduli space massless deformations of the compactification, thus making some vacua unstable. In M-theory the non-perturbative contributions come from wrapped membranes and fivebranes and there has been some effort to calculate their contribution both for bulk and gauge field moduli.[62] Their effects are described by a super-potential, and the contributions are related to summing over holomorphic curves in X.

In a bigger picture, the M-theory vacua we have been describing are just one corner of a large moduli space of $\mathcal{N} = 1$, $d = 4$ compactifications. Other corners are the weak heterotic string, F-theory and type II intersecting brane backgrounds. By concentrating on elliptically fibred Calabi–Yau manifolds we keep a duality to F-theory as well as the weak heterotic limit. A more general goal is to understand this moduli space of compactifications more globally, taking information from different limits. We already in the M-theory picture see a very rich structure of transitions and branches in moduli space. A very different point, brought out in the constructions we have described here is the connection between supersymmetric D-branes and sheafs, and in particular the realization of T-duality simply as a *manifest* symmetry of the space of sheafs (or more generally derived categories). Clearly there remains much to do both in understanding the phenomomenological details of these models and in elucidating the structural underpinnings of the dualities.

Acknowledgments

With regard to my original contributions to the work presented here, I would very much like to thank my collaborators Ron Donagi, André Lukas, Burt Ovrut, Tony Pantev and Kelly Stelle. I am very grateful to the organizers Steve Gubser, Joe Lykken and K.T. Mahanthappa for arranging what was an excellent Summer School, and for the opportunity to lecture. I also thank the students for their enthusiasm and insight. This work is supported by the Royal Society through a University Research Fellowship.

Appendix A. Holonomy

Steve Gubser has given a very nice introduction to supersymmetry and special holonomy as separate set of lectures at this school.[49] Here we will just recap some basic definitions[63,64] which will be helpful when discussing Calabi–Yau manifolds and supersymmetric gauge field configurations.

The notion of holonomy can be applied to any vector bundle E over a d-dimensional manifold X given a connection $D = \mathrm{d} + \omega$. The fibre E_p at a given point is a n-dimensional real or complex vector space. (Note if E_p is complex, it is more usual in the physics literature to write the connection as $\omega = \mathrm{i}A$.) Our prime example is where E is the tangent bundle TX and D is the Levi–Civita connection ∇.

Once one has a connection, one can define *parallel transport* of a vector v along a curve $\gamma(t)$ in X, by requiring that variation of v along γ vanishes, that is $D_{\dot\gamma(t)} v = 0$. If we parallel transport around a closed loop, the vector in general does not come back to itself. We then say that the manifold has non-trivial holonomy. The classic example is transporting a vector in TS^2 around an octant of the surface of a two-sphere S^2 as shown in Fig. 10, in which case the vector comes back to the starting point p rotated through a right angle.

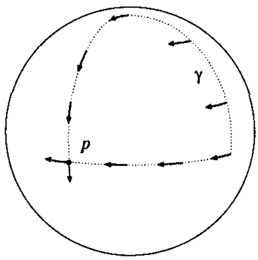


Figure 10. Holonomy on S^2

In general for a real vector bundle, a vector v comes back to itself up to a $GL(n, \mathbb{R})$-rotation g_γ. (For a complex bundle we just replace $\mathbb{R}$ with $\mathbb{C}$ here and in the following.) We can write the transformation as

$$v' = g_\gamma v = P\, \mathrm{e}^{\int_\gamma \omega}\, v \tag{A.1}$$

where P represents the path-ordered exponential. Combining the rotations coming from different loops all starting and ending at the same point p, we generate the *holonomy group*

$$\mathrm{Hol}(E, D) = \{g_\gamma : \gamma \text{ is a loop based at } p\} \subseteq GL(n, \mathbb{R}). \tag{A.2}$$

It is easy to see that the group is independent of the choice of base point p. (The representations of the holonomy group at different points are conjugates of each other, where the conjugating element is built from parallel transport along a path joining the two points.)

If D is Levi–Civita connection on a d-dimensional Riemannian manifold, then we have the additional structure of a matrix g. This gives a inner product on the vectors v which is preserved under parallel transport. This implies that $g_\gamma \in O(d)$. We write in this special case

$$\mathrm{Hol}(X) \equiv \mathrm{Hol}(TX, \nabla) \subseteq O(d). \tag{A.3}$$

More generally if actually the vector bundle $E = \rho(V)$ is the representation of some gauge group G on a vector space V and D is a G-connection then $\mathrm{Hol}(E, D) \subseteq G$ (and is realized in the representation ρ).

If X is not simply connected that there two parts to the holonomy. First there is the restricted holonomy group $\mathrm{Hol}^0(E, D)$ built out of loops which are contractible. This is a measure of the curvature of the connection. In fact, the curvature can be defined as the holonomy around an infinitesimal loop.

Even if the curvature vanishes there can still be non-trivial holonomy around non-contractible loops. In this case, the holonomy only depends on the homotopy class of the loop. Physically, these elements in $\mathrm{Hol}(E, D)$ are Wilson lines. (Recall that two loops are homotopic if one can be continuously deformed into the other. Set of homotopy classes forms a group $\pi_1(X)$ known as the fundamental group of X.) In general, even if the curvature does not vanish we have a homorphism W describing the set of *Wilson lines*

$$W : \pi_1(X) \to \mathrm{Hol}(E, D)/\,\mathrm{Hol}^0(E, D). \tag{A.4}$$

Since $\pi_1(X)$ is countable, the set of Wilson lines $\mathrm{Hol}(E, D)/\,\mathrm{Hol}^0(E, D)$ is also some countable. We also see that $\mathrm{Hol}^0(E, D)$ is the connected component of $\mathrm{Hol}(E, D)$ containing the identity.

Appendix B. *G*-structures and special holonomy

Let us concentrate on the holonomy in the tangent space $\mathrm{Hol}(TX, D)$. It is natural to ask what possible holonomy groups can appear. A completely generic connection on TX gives $\mathrm{Hol}(TX, D) = GL(d, \mathbb{R})$. If we have a metric g and $Dg = 0$, then the norm of any vector v is constant and so $\mathrm{Hol}(TX, D) \subseteq O(n) \subset GL(n, \mathbb{R})$. Furthermore, there is a unique connection, the Levi–Civita connection ∇ which has $\nabla g = 0$ and has vanishing *torsion*.[64,65] The torsion T is defined by

$$D_A D_B f - D_B D_A f = -T^C{}_{AB} D_C f, \tag{B.1}$$

for any function f on X.

Mathematically, we say that the metric g is an *invariant tensor* (under $O(n) \subset GL(n, \mathbb{R})$) and defines a *G-structure* on X, where here G is $O(n)$. A connection D such that $Dg = 0$ is said to be compatible with the G-structure, and necessarily has $\mathrm{Hol}(TX, D) \subseteq G$. The existence of ∇ means there is no obstruction to finding a compatible torsion-free connection.

The point about G-structures is that they imply that transition functions for the tangent bundle are actually elements of $G \subset GL(n, \mathbb{R})$. More formally, the frame bundle F of TX is a bundle built from set of all frames $(e_1, \ldots, e_d)$ for TX_p at each point p in X. It is a $GL(n, \mathbb{R})$ principle bundle since the fibre is just copy of $GL(n, \mathbb{R})$. We then define[64,66,67]

a *G-structure* is a principle sub-bundle P of F with fibre G.

In the example of $O(n)$ it is the set of orthonormal frames. Typically, we find

(1) a G-structure is equivalent to a set of globally defined, nowhere vanishing G-invariant tensors $\{\Xi^i\}$;
(2) given a G-structure, all tensors can be decomposed into $G \subset GL(n, \mathbb{R})$ representations;
(3) if D is compatible with the G-structure, that is $D\Xi^i = 0$, then $\mathrm{Hol}(TX, D) \subseteq G$;
(4) manifolds with a G-structure always admit some compatible connection D, but in general there is an obstruction, known as the *intrinsic torsion*, to admitting a torsion-free, compatible connection.

For instance, for $O(n)$ the metric g is a globally defined $O(n)$ invariant tensor. If we also have the volume form as well of the metric, then we have a $SO(n)$-structure. Unlike the $O(n)$ case, in general, not all manifolds admit a G-structure: it is a topological constraint that TX can actually be patched with G-valued transition functions. For instance for an $SO(n)$-structure the manifold must be orientable. Furthermore, even if X admits a G-structure it may not admit a torsion-free connection D compatible with the structure.

A classic example of this formalism is an *almost complex structure* on a $2n$-dimensional manifold X. This is a tensor $J^A{}_B$ such that $J^A{}_C J^C{}_B = \delta^A{}_B$. It is invariant under local $GL(n, \mathbb{C})$ transformations so defines a $GL(n, \mathbb{C})$-structure. In general, not all manifolds admit an almost complex structure. The existence of J allows tensors to be decomposed in $GL(n, \mathbb{C})$ representations, defined by projecting with $\frac{1}{2}(\delta^A{}_B \pm \mathrm{i}J^A{}_B)$. As a further restriction, X admits a *complex structure* only if there is a torsion free connection D compatible with J. This requires the integrability condition that the obstruction, measured by the Nijenhuis tensor,

$$N^C{}_{AB} = J^D{}_A \left(\partial_D J^C{}_B - \partial_B J^C{}_D\right) - J^D{}_B \left(\partial_D J^C{}_A - \partial_A J^C{}_D\right), \quad \text{(B.2)}$$

vanishes. Then X is a complex manifold and we can introduce complex co-ordinates. Again not all manifolds which admit almost complex structures also admit complex structures.

Of particular interest are the cases where $G \subset O(d)$. We then necessarily have a metric g and hence the natural Levi–Civita connection ∇. From what we have said so far, we see that[64]

> X has *special holonomy*, $\mathrm{Hol}(X) \subset O(d)$, if and only if it admits a G-structure with $G \subset O(n)$ which is compatible with ∇, that is with $\nabla\Xi^i = 0$.

This is equivalent to simply saying that the manifold admits a G-structure with a compatible torsion-free connection (which is then, necessarily ∇).

What possible special holonomy groups can appear? This has been answered in a theorem due to Berger and Simons.[64,68] First we have two results, the second a theorem due to de Rham, which allow us to restrict the form of X:

(1) If X is not-simply connected, we consider the covering space $\hat{X}$. Then $\mathrm{Hol}(\hat{X}) = \mathrm{Hol}^0(X)$. Thus we determine the allowed special holonomy up to the discrete part $\mathrm{Hol}(X)/\,\mathrm{Hol}^0(X)$, due to the Wilson lines, which depends on $\pi_1(X)$.
(2) If $\mathrm{Hol}(X)$ for a simply connected X is reducible (sum of two groups) then the manifold is metrically a product.

This means we can consider only simply connected X with non-reducible $\mathrm{Hol}(X)$. One simple class of special holonomy manifolds with $\mathrm{Hol}(X) = H$ are the *symmetric spaces* G/H, where $H \subset G$ are Lie groups. The theorem then states that special holonomy manifolds are symmetric spaces or one of the seven possibilities listed in Table 1. The proof depends partly on some representation theory on what G-structures are possible, and then some results about the symmetry properties of the Riemann curvature tensor. That examples of G_2 and *Spin*(7) holonomy existed was only demonstrated relatively recently.[64] Note the invariant tensors are all real degree-p forms, written as ξ_p, satisfying certain algebraic conditions, except for the metric g and the complex structure J mentioned above and Ω_n which is complex.

Table 1. Special holonomy and supersymmetry

dimension	space X	Hol(TX)	$\{\Xi^i\}$	fraction of susy
d	generic	$SO(d)$	g	none
$d=2n$	Kähler	$U(n)$	ω_2, J	none
$d=2n$	Calabi–Yau	$SU(n)$	ω_2, Ω_n	$1/2^{n-1}$
$d=4n$	quaternionic	$Sp(1)\times Sp(n)/\mathbb{Z}_2$	(ω_2^i)	none
$d=4n$	hyper-Kähler	$Sp(n)$	ω_2^i	$1/2^n$
$d=7$	Joyce	G_2	ϕ_3	$1/8$
$d=8$	Cayley	$Spin(7)$	Ψ_4	$1/16$

The forms ω_2^i are a triplet of two forms. On a quaternionic manifold the three-vector of two forms (ω_2^i) is invariant only up to an $SO(3)$ rotation on the i index. Strictly, the theorem tells us not only what the holonomy is but also in what representation of G it is realized (namely the d-dimensional one).

For physics applications, the key point is that supersymmetry is equivalent to special holonomy. A space is supersymmetric if it has a global solution to the Killing spinor equation

$$\nabla_A \eta = 0. \tag{B.3}$$

This implies we have a invariant spinor η defining a $G \subset Spin(d)$ structure which is compatible with ∇. In other words we have special holonomy. Not all the groups in Table 1 can appear as groups which leave a spinor invariant. In fact the only possibilities are $SU(n)$, $Sp(n)$, G_2 and $Spin(7)$. The corresponding fractions of preserved supersymmetry are shown in the table.

Note that all the supersymmetric manifolds are Ricci flat. From the Killing spinor equation we have

$$[\nabla_A, \nabla_B]\eta = R_{AB}{}^{CD}\gamma_{CD}\eta = 0 \tag{B.4}$$

Acting with $R^A{}_E\gamma^{EB}$ and given the symmetries of the Riemann tensor then gives $R_{AB}R^{AB} = 0$ and hence $R_{AB} = 0$. (Note on a manifold with Lorentzian signature one cannot make this final step.)

Given a Killing spinor one can construct, using the gamma matrices, the invariant forms Ξ^i in the form of spinor bilinears.[69,66] For instance, for $SU(n)$, we can define the two forms[11,66]

$$\begin{aligned} \omega_{MN} &= -\mathrm{i}\bar{\eta}\gamma_{MN}\eta && \text{Kähler form,} \\ \Omega_{M_1\ldots M_n} &= \bar{\eta}^c\gamma_{M_1\ldots M_n}\eta && \text{complex holomorphic } n\text{-form,} \end{aligned} \tag{B.5}$$

where η^c is the conjugate spinor. The forms satisfy algebraic relations[66]

$$\begin{aligned} \omega \wedge \Omega &= 0, \\ \Omega \wedge \bar{\Omega} &= 2^n \mathrm{i}^{n(n+2)} \omega^n / n!, \end{aligned} \tag{B.6}$$

where ω^n is the wedge product of ω with itself n times. The two-form ω is left invariant by $Sp(n,\mathbb{R}) \subset GL(2n,\mathbb{R})$ and Ω by $SL(n,\mathbb{C}) \subset GL(n,\mathbb{R})$. The common subgroup is $SU(n)$. Thus, in fact, the existence of ω and Ω imply the gauge group of the bundle is $SU(n)$ and hence determine a metric. The integrability conditions $\nabla\omega = \nabla\Omega = 0$, implied by $\nabla_A \eta = 0$ are equivalent to

$$\mathrm{d}\omega = \mathrm{d}\Omega = 0. \tag{B.7}$$

This is a characteristic result of supersymmetric structures. All the integrability conditions can be written as $\mathrm{d}\Xi^i = 0$.

As a final comment, we note that the G-structure formulation provides a useful way of characterizing supersymmetric backgrounds with flux.[70,66] Then the supersymmetry condition generalizes to an expression $D_A \eta = 0$ where D_A is a generalized derivative involving the flux. (See for instance Eq. (13).) The existence of η implies we still have a $G \subset Spin(d)$ structure, with G one of $SU(n)$, $Sp(n)$, G_2 or $Spin(7)$, but it is no longer integrable. The obstruction to integrability, the intrinsic torsion, is measured by $(D_A - \nabla_A)\eta = -\nabla_A \eta$. By construction it is given in terms of the flux. Thus we can characterize X, not as special holonomy manifold (which has vanishing intrinsic torsion), but as a manifold with G-structure and intrinsic torsion given by the flux. Since $\mathrm{d}\Xi^i$ measures the integrability, the intrinsic torsion conditions can be written as $\mathrm{d}\Xi^i \sim$ flux.

Appendix C. de Rham and Dolbeault cohomology

Let us very briefly recall the definitions of de Rham and Dolbeault cohomology. On a differentiable d-dimensional manifold X we have the exterior derivative operator d taking p-forms ω_p to $(p+1)$-forms ω_{p+1}, with $\mathrm{d}^2 = 0$. We define the de Rham cohomology groups $H^p(X,\mathbb{R})$ as the space of *closed forms* $\mathrm{d}\omega_p = 0$ modulo *exact* forms $\omega_p \sim \omega'_p + \mathrm{d}\beta_{p-1}$, so

$$H^p(X,\mathbb{R}) = \frac{\{\omega_p : \mathrm{d}\omega_p = 0\}}{\{\alpha_p : \alpha_p = \mathrm{d}\beta_{p-1}\}}. \tag{C.1}$$

The set of all groups $H^*(X,\mathbb{R}) = \bigoplus_i H^i(X,\mathbb{R})$ has a ring structure under the wedge product (since the wedge product of two closed or exact forms is itself closed or exact). The dimensions b_p of $H^p(X,\mathbb{R})$ are known as the *Betti numbers.*

Given an orientation on X, there is a natural inner product between elements of $H^p(X,\mathbb{R})$ and $H^{d-p}(X,\mathbb{R})$ given by $(\omega_p, \omega_{d-p}) = \int_x \omega_p \wedge \omega_{d-p}$. This then gives an identification known as *Poincaré duality* that

$$H^p(X,\mathbb{R}) \text{ and } H^{d-p}(X,\mathbb{R}) \text{ are dual vector spaces,} \tag{C.2}$$

so $b_p = b_{n-p}$.

Given a metric g, we can define *harmonic forms* γ on X, satisfying $\mathrm{d}\gamma = \mathrm{d} * \gamma = 0$. The Hodge decomposition theorem then states that there is unique harmonic representative for each element in $H^p(X,\mathbb{R})$. Poincaré duality is then realized as the natural isomorphism of the Hodge star operation on harmonic forms $\gamma \to *\gamma$.

On a complex manifold of dimension $2n$ we can introduce complex coordinates z^a and the conjugates $\bar{z}^{\bar{b}}$. This means the exterior derivative splits into holomorphic and anti-holomorphic pieces $\mathrm{d} = \partial + \bar{\partial}$ with $\partial^2 = \bar{\partial}^2 = 0$. Similarly we can decomposed complex p-forms under $GL(n,\mathbb{C})$ as (p,q)-type forms where

$$\omega^{p,q} = \omega_{a_1 \dots a_p \bar{b}_1 \dots \bar{b}_q} \mathrm{d}z^{a_1} \wedge \cdots \wedge \mathrm{d}z^{a_p} \wedge \mathrm{d}\bar{z}^{\bar{b}_1} \wedge \cdots \wedge \mathrm{d}\bar{z}^{\bar{b}_q}. \tag{C.3}$$

Since $\bar{\partial}^2 = 0$ we can define the Dolbeault cohomomology in analogy to the de Rham cohomology

$$H^{p,q}(X) = \frac{\{\omega^{p,q} : \bar{\partial}\omega^{p,q} = 0\}}{\{\alpha^{p,q} : \alpha^{p,q} = \bar{\partial}\beta^{p,q-1}\}}. \tag{C.4}$$

The dimensions $h^{p,q}$ of $H^{p,q}(X)$ are the *Hodge numbers.* As above one can define harmonic forms $\bar{\partial}\omega^{p,q} = \bar{\partial} * \omega^{p,q} = 0$ and there is a Hodge decomposition theorem implying that there is a unique harmonic form for each class in $H^{p,q}(X)$. Poincaré duality implies $h^{p,q} = h^{n-p,n-q}$.

If the manifold is Kähler (see Sec. 3.1 and Appendix B) the two Laplacians built from d and $\bar{\partial}$ agree and we get a number of further relations. First $H^{p,q}(X) = \overline{H^{q,p}(X)}$ so $h^{p,q} = h^{q,p}$. Second there is a decomposition

$$H^r(X,\mathbb{C}) = \bigoplus_{p+q=r} H^{p,q}(X), \tag{C.5}$$

so that $b_r = \sum_{p+q=r} h^{p,q}$ and $H^{p,q}(X)$ is a refinement of $H^r(X,\mathbb{C})$. (Here $H^r(X,\mathbb{C})$ is just $H^r(X,\mathbb{R})$ but with complex p-forms.) Note when $p = q$ we can define real (p,q)-forms and there is a real group $H^{2p}(X,\mathbb{R}) \cap H^{p,p}(X)$.

Appendix D. Homology of cycles

The natural duals of the de Rham cohomology groups are homology groups of cycles on an oriented manifold X. We define a p-dimensional cycle C_p as a sum

$$C_p = \sum_i a_i C_p^{(i)} \tag{D.1}$$

where $C_p^{(i)}$ are closed p-dimensional smooth oriented submanifolds of X. This implies that C_p has no boundary $\partial C_p = 0$. The coefficients a_i may be in $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ or $\mathbb{C}$ depending on what we are interested. In general we write $a_i \in \mathbb{F}$. Note the boundary of a boundary is zero, that is $\partial^2 = 0$. We then define homology groups as the group of cycles modulo the set of boundaries, that is cycles of the form $A_p = \partial B_{p+1}$. In other words we identify $C_p \sim C'_p + \partial B_{p+1}$ and define

$$H_p(X, \mathbb{F}) = \frac{\{C_p : \partial C_p = 0\}}{\{A_p : A_p = \partial B_{p+1}\}}. \tag{D.2}$$

The set of all groups $H_*(X, \mathbb{F}) = \bigoplus_p H_p(X, \mathbb{F})$ has a ring structure under intersection of cycles. (Intersection, just means you take the cycle built out of the common sub-manifolds between each combination of the $C_p^{(i)}$ and $C'^{(i)}_q$ defining the original cycles, with coefficient $\pm a_i a'_i$ where the sign is fixed by the orientation.) Given some closed cycle C_p, the corresponding homology class is written as $[C_p]$.

There is a natural inner product between elements of $H_p(X, \mathbb{R})$ and the de Rham group $H^p(X, \mathbb{R})$ given by

$$\pi(C_p, \omega_p) = \int_{C_p} \omega_p, \tag{D.3}$$

where $\pi(C_p, \omega_p)$ is called a period. This product is non-degenerate and so

$$H_p(X, \mathbb{R}) \text{ and } H^p(X, \mathbb{R}) \text{ are dual vector spaces.} \tag{D.4}$$

Since $H^{d-p}(X, \mathbb{R})$ and $H_p(X, \mathbb{R})$ are both dual to $H^p(X, \mathbb{R})$ we can identify elements ω_{d-p} and C_p of each by

$$\int_{C_p} \phi = \int_X \phi \wedge \omega_{n-p}, \tag{D.5}$$

for all $\phi \in H^p(X, \mathbb{R})$. This naturally defines an isomorphism (Poincaré duality)

$$H_p(X, \mathbb{R}) \cong H^{d-p}(X, \mathbb{R}). \tag{D.6}$$

which is used throughout the lectures.

Note this also allows us to define a integer cohomology $H^p(X, \mathbb{Z})$ as the image of the integer homology $H_p(X, \mathbb{Z})$. There is a subtlety here, that *torsion classes*, that is cycles where nC is trivial in homology for some integer n are lost in the isomorphism (D.6) since their integrals against ϕ are always zero.

Appendix E. Alegebraic and effective classes

It would be nice to be able to have a notion of homology groups to the complex case, in analogy to the relation between $H^r(X, \mathbb{R})$ and $H^{p,q}(X)$, to find some refinement of $H_r(X, \mathbb{F})$.

The natural object to consider is a *analytic cycle* C_{2p}. This is a sum as (D.1) above but now each component C_i is a closed analytic sub-manifold. That is

> $C_{2p}^{(i)}$ is defined by the vanishing of $n-p$ (and no fewer) holomorphic functions.

In particular if C_{2p} is an analytic cycle then

(1) components $C_2^{(i)}$ are called a *holomorphic curves*,
(2) components $C_4^{(i)}$ are called a *holomorphic surfaces*, and
(3) a cycle C_{2n-2} is called a *divisor*.

We define the *algebraic homology group*:[71]

> $H_{2p}(X, \mathbb{Z})_{\text{alg}} \subseteq H_{2p}(X, \mathbb{Z})$ is homology group of analytic cycles C_{2p}.

For a generic complex manifold this may be empty even if $H_p(X, \mathbb{Z})$ is not: there may be non-trivial cycles but none of them are analytic.

By considering the identification (D.5), it is easy to see that $H_{2p}(X, \mathbb{Z})_{\text{alg}} \subset H^{n-p,n-p}(X) \cap H^{2n-2p}(X, \mathbb{Z})$. (Here we are ignoring torsion classes.) What would be nice is the much stronger statement that every element of $H^{n-p,n-p}(X) \cap H^{2n-2p)}(X, \mathbb{Z})$ has a representative which is a analytic cycle. This is the famous Hodge conjecture,

$$\text{Hodge conjecture:} \quad H_{2p}(X, \mathbb{Z})_{\text{alg}} \cong H^{n-p,n-p}(X) \cap H^{2n-2p}(X, \mathbb{Z})$$

(again ignoring torsion classes). For divisors $p = n - 1$ and, by Poincaré duality, curves, $p = 1$, the conjecture has been proven to be true and is known as the *Lefschetz theorem*. For general p proving the conjecture is an unsolved problem.

On a Calabi–Yau threefold we only have divisors and curves and hence this means we can make the identification. Furthermore, we also have $h^{2,0} = h^{3,1} = 0$ so all elements in $H^{2p}(X,\mathbb{Z})$ are of type (p,p). This implies $H_{2p}(X,\mathbb{Z})_{\text{alg}} = H_{2n-2p}(X,\mathbb{Z})$. One has the following picture. Consider the divisors. In general, the image of $H^2(X,\mathbb{Z})$ is a lattice of points in $H^2(X,\mathbb{R})$. Choosing a complex structure on X corresponds to fixing an $h^{1,1}$-dimensional subspace within $H^2(X,\mathbb{R})$ describing the space $H^{1,1}(X)$. Generically, no lattice points will intersect the subspace and so there are no algebraic classes on X. The exception is when $h^{2,0} = 0$, which is the case for all Calabi–Yau manifolds. Then the subspace is the whole space $H^2(X,\mathbb{R})$ so all classes in $H^2(X,\mathbb{Z})$ are algebraic. Thus in this case, we can make the identification

$$H_{2p}(X,\mathbb{Z})_{\text{alg}} = H_{2p}(X,\mathbb{Z}) \text{ on a Calabi–Yau threefold}$$

The same applies to any complex two-fold with $h^{2,0} = 0$.

Finally, we will need a couple more definitions.

(1) A analytic cycle is *reducible* if it can be written as the union of two analytic sub-manifolds. A class in $H_{2p}(X,\mathbb{R})_{\text{alg}}$ is *irreducible* if it has an irreducible representative. Note that there may be other analytic subspaces in the class which are reducible, but the class is irreducible if there is at least one irreducible representative.
(2) A class is *effective* if it can be written as the sum of irreducible classes with positive integer coefficients.

The collection of all effective classes of curves forms a cone in $H_2(X,\mathbb{Z})_{\text{alg}}$ known as the Mori cone. The Mori cone can be shown to be linearly generated by a set of effective classes. This set includes the effective basis of $H_2(X,\mathbb{Z})_{\text{alg}}$ but is, in general, larger. The Mori cone can be finitely generated, as for del Pezzo surfaces, or infinitely generated, as for dP_9 (see Appendix J).

Appendix F. Calibrations

Having defined cohomology and homology we can now define the notion of a calibrated cycle.[64,67,69,72,73] Suppose we have some p-form Ξ on a Riemannian manifold such that, if ξ is an oriented p-dimensional tangent plane at any point $p \in X$, we have

$$\begin{aligned} \Xi|_\xi &\leq \text{vol}|_\xi \quad \forall\, \xi, \\ d\Xi &= 0, \end{aligned} \tag{F.1}$$

where vol is the volume form on X. We then call Ξ a *calibration.* We note that

all the forms Ξ^i defining G-structures are calibrations.

Given a cycle C_p, we say it is calibrated if

$$\text{calibrated cycle:} \quad \Xi|_{T_x C_p} = \mathrm{vol}|_{T_x C_p} \quad \forall x \in C_p \tag{F.2}$$

If C'_p is another cycle in the same homology class we have

$$\int_{C_p} \mathrm{vol} = \int_{C_p} \Xi = \int_{C'_p} \Xi \leq \int_{C'_p} \mathrm{vol}, \tag{F.3}$$

and we get the main result that a calibrated cycle has minimum volume in its homology class.

In string theory, we define a supersymmetric cycle as a cycle in a supersymmetric manifold X on which a brane can wrap without breaking supersymmetry. As we saw in the case of M5-branes in Sec. 3.3, this reduces to a condition (37) of the form of a projector on the Killing spinor η giving zero. Using the fact[73] that for a supersymmetric manifold X, the forms Ξ^i can be written as bi-linears in η,[69] one finds that

supersymmetric cycles $\Leftrightarrow$ cycles calibrated by Ξ^i .

As specific example, consider the M5-brane wrapping a two-cycle W in a Calabi–Yau manifold X, as described in Sec. 3.3. The relevant two-form calibration is ω and one can see that the last condition in Eq. (39) is equivalent to the calibrated cycle condition (F.2). From the discussion in Sec. 3.3, we saw that W was actually a holomorphic curve. Thus it defined an element in $H_2(X,\mathbb{Z})_{\mathrm{alg}}$. Given the Lefschetz theorem, we then see that holomorphic curves are the minimum volume cycles in $H_2(X,\mathbb{Z})$.

Given the list of supersymmetric manifolds in Table 1 one finds the possible calibrations[69] given in Table 2. For the hyper-Kähler calibrations

Table 2. Supersymmetry and special holonomy

X	Hol(TX)	calibration	cycle C_p	p
Calabi–Yau	$SU(n)$	$\omega^k/k!$	holomorphic	$2k$
		$\mathrm{Re}(e^{i\theta}\Omega)$	special lagrangian	n
hyper-Kähler	$Sp(n)$	$\Theta^k/k!$	quaternionic	$4k$
		$\omega^n/n! + \mathrm{Re}(e^{i\theta}\Omega)$	complex lagrangian	$2n$
Joyce	G_2	ϕ	associative	3
		$*\phi$	coassociative	4
Cayley	$Spin(7)$	Ψ	Cayley	4

we have $\Theta = \omega_i^2/2$, which is independent of i, while ω is a unit linear combination of the ω_i and Ω is the corresponding holomorphic $2n$-form where X is viewed as a Calabi–Yau manifold. For the Calabi–Yau and hyper-Kähler cases there are also additional calibrations which come from viewing a hyper-Kähler manifold as a special case of Calabi–Yau manifold, or for $d = 8$ a Calabi–Yau manifold as a special case of a Cayley manifold.

Appendix G. Line bundles and divisors

Recall that on a *complex* vector bundle $E \xrightarrow{\pi} X$ of rank r each fibre $\pi^{-1}(x)$ over $x \in X$ is a copy of $\mathbb{C}^r$. Over each patch $U \subset X$ we have a local isomorphism

$$\phi_U : E|_U \to U \times \mathbb{C}^n. \tag{G.1}$$

For a complex bundle the transition functions g_{UV} on the overlap of two patches $U \cap V$ live in $GL(r, \mathbb{C})$, that is

$$g_{UV} := \phi_U \circ \phi_V^{-1}|_{U\cap V} : U \cap V \to GL(r, \mathbb{C}). \tag{G.2}$$

In addition, we have the usual consistency requirements on the patching that $g_{UV} \cdot g_{VU} = g_{UV} \cdot g_{VW} \cdot g_{WU} = \mathrm{id}$.

A *holomorphic bundle* is defined over a complex manifold X and is a complex vector bundle such that

$$g_{UV} \text{ are holomorphic functions on } X.$$

Such objects are important because the total space E is then a complex manifold. Since the transitions functions are holomorphic it is consistent to define holomorphic (or meromorphic[e]) sections s as holomorphic (meromorphic) maps $s : U \to E$. See also Sec. 3.2 for more details on holomorphic bundles.

A classic example of a holomorphic bundle comes from the tangent bundle TX of a $2n$-dimensional complex manifold. In complex coordinates z^a and $\bar{z}^{\bar{b}}$ we can write a complex vector v as

$$v = u^a \frac{\partial}{\partial z^a} + w^{\bar{b}} \frac{\partial}{\partial \bar{z}^{\bar{b}}} \tag{G.3}$$

If TX_c is the complex bundle spanned by $\partial/\partial z^a$ then the complexified tangent space $T_{\mathbb{C}}X$ (space of complex-valued vector fields) decomposes as

$$T_{\mathbb{C}}X = TX_c + \overline{TX}_c. \tag{G.4}$$

[e]A meromorphic function is just the ratio $h = f/g$ of two holomorphic functions, and so is allowed to have poles.

The bundle TX_c is a holomorphic vector bundle since the coordinates z^a patch holomorphically on a complex manifold X.

A *line bundle* L is a rank-one holomorphic vector bundle, that is, with a one-dimensional fibre $\pi^{-1}(x) = \mathbb{C}$. Consequently the transition functions ϕ_{UV} are just non-zero holomorphic functions on $U \cap V$. A holomorphic (or meromorphic) section of a line bundle is a holomorphic (meromorphic) function s_U on each patch U, with the patching $s_U = g_{UV} s_V$.

There is a very useful correspondence between line bundles and divisors. Recall that a divisor D is analytic $(2n-2)$-cycle of the form (D.1). By definition, each component $C^{(i)}_{2n-2}$ is defined locally on a patch U by the vanishing of a single holomorphic function f_U. More generally, we can define the whole divisor D by local *meromorphic* functions h_U, which have a zero of order a_i (if $a_i > 0$) or a pole of order a_i (if $a_i < 0$) on $C^{(i)}_{2n-2}$. This is shown in Fig. 11. Between patches the functions must be related by $h_U = g_{UV} h_V$ for some non-zero holomorphic function g_{UV} on $U \cap V$. But this is just the definition of a meromorphic section of some line bundle L, with transition functions $g_{UV} = h_U / h_V$. It is usual to denote the corresponding line bundle by $\mathcal{O}(D)$. Conversely, any meromorphic section s of a line bundle L defines a divisor simply by taking $h_U = s_U$ on each patch. We get a one-to-one correspondence

$$\text{divisor } D \;\Leftrightarrow\; \text{meromorphic section of a line bundle } L = \mathcal{O}(D) \qquad \text{(G.5)}$$

(Note, we could equally well define D by functions $h'_U = f_U h_U$ where f_U are non-zero holomorphic functions, but just corresponds to a different trivialization of the same line bundle.)

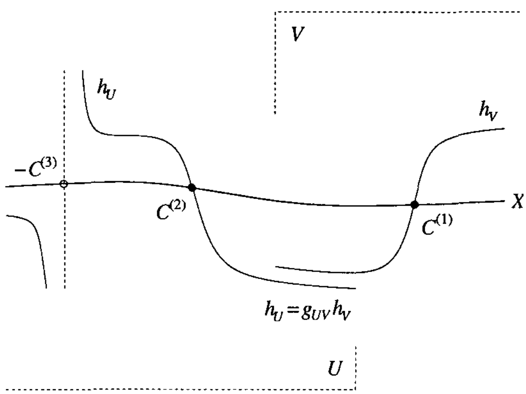


Figure 11. Divisor $D = C^{(1)} + C^{(2)} - C^{(3)}$ in X

In general, many different divisors define the same line bundle and also there are line bundles with no meromorphic sections and hence no corresponding divisor. Finally, note that an effective divisor, where all the $a_i > 0$, corresponds to a *holomorphic* section of the line bundle $\mathcal{O}(D)$.

Appendix H. Chern classes

We will mention two ways to define the Chern classes of a complex vector bundle E. The first is in terms of sections. From our discussion of line bundles L and divisors, we noted that meromorphic sections of L correspond to divisors. Furthermore it turns out that the divisors corresponding to any two sections of the same line bundle are in the same homology class. Thus it is natural to define for any line bundle with a meromorphic section

$$\mathrm{c}_1(L) = [D] \in H_{2n-2}(X, \mathbb{Z})_{\mathrm{alg}}, \tag{H.1}$$

where D is the divisor corresponding to some meromorphic section s of L, and $[D]$ represents the corresponding homology class. In fact, this definition generalizes to complex line bundles (dropping the holomorphic condition), in which case D is just defined as the cycle defined by the zeros of a general section of L, the so called "degeneracy cycle".

More generally, for rank n complex vector bundles E one can define a set of Chern classes

$$\mathrm{c}_p(E) = [D_{n-p+1}] \in H_{d-2p}(X, \mathbb{Z}), \tag{H.2}$$

where D_{r-p+1} is the degeneracy cycle given by the set of points where of $(r-p+1)$ generic sections become linearly dependent (with a sign fixed by the orientation of the sections in the $\mathbb{C}^r$-fibre). We have $\mathrm{c}_p(E) \in H_{d-2p}$, because, at a given point $x \in X$, the condition is that $(r-p+1)$ vectors in the $\mathbb{C}^r$ fibre are linear dependent, which generically corresponds to p complex conditions. This is $2p$ real conditions over X, which should define a C_{d-2p} cycle.

Given this definition one finds

(1) $\mathrm{c}_p(E) = 0$ for $p > r$ and $\mathrm{c}_0(E) = 1$, since $(r+1)$ sections sections are dependent everywhere on X;

(2) the Whitney sum formula

$$\mathrm{c}_k(E_1 \oplus E_2) = \sum_{k=p+q} \mathrm{c}_p(E_1) \cdot \mathrm{c}_q(E_2);$$

(3) for line bundles

$$\mathrm{c}_1(L_1 \otimes L_2) = \mathrm{c}_1(L_1) + \mathrm{c}_1(L_2);$$

(4) for holomorphic bundles on Kähler manifold X we have, under Poincaré duality,

$$\mathrm{c}_p(E) \in H^{p,p}(X) \cap H^{2p}(X, \mathbb{Z}),$$

and furthermore if X is an algebraic variety (so can be embedded in $\mathbb{C}P^m$), we have

$$\mathrm{c}_p(E) \in H^{2p}(X, \mathbb{Z})_{\mathrm{alg}}.$$

The second way to define Chern classes is as Poincaré dual elements in $H^{2p}(X, \mathbb{C})$. One takes the field strength F of a general $GL(r, \mathbb{C})$-connection $D = \mathrm{d} + \mathrm{i}A$. We define

$$\begin{aligned} \mathrm{c}_0(E) &= 1, \\ \mathrm{c}_1(E) &= \tfrac{1}{2\pi} \operatorname{tr} F, \\ \mathrm{c}_2(E) &= \tfrac{1}{8\pi^2} \left(\operatorname{tr} F \wedge F - \operatorname{tr} F \wedge \operatorname{tr} F \right), \\ &\cdots \end{aligned} \tag{H.3}$$

where it is understood that we really mean the cohomology class of the expressions on the right not the forms themselves. These expressions can be written more succinctly as

$$\mathrm{c}(E) \equiv \mathrm{c}_0(E) + \cdots + \mathrm{c}_r(E) = \det\left(\mathrm{id} + \tfrac{F}{2\pi}\right), \tag{H.4}$$

where the determinant is taken over the fibre indices, and the expression is understood as a (finite) expansion in F, with F^n corresponding to $F \wedge \cdots \wedge F$ n-times.

The point is that it is relatively easy to show[63,71,11] that each of these polynomials in F is, first, closed, because of the Bianchi identity, and second, choosing a different connection simply shifts the polynomial by an exact form. Then the polynomials as elements in $H^{2p}(X, \mathbb{C})$ are independent of the choice of connection. From this point of view it is much harder to see that they are integral classes. The correspondence between these two definitions of $\mathrm{c}_p(E)$ is given by the Gauss–Bonnet formulae.[71]

In general, there are other *characteristic classes* that can be defined for vector or principle bundles. In general, any polynomial expression in F which is invariant under gauge transformations is closed and defines a class in $H^{2p}(X, \mathbb{R})$ independent of the choice of connection.[63,11]

For instance on a real rank r vector bundle W one has the analogs of the Chern classes known as the *Pontrjagin classes* $p_j(W)$ with $2j \leq r$ and $4j \leq d = \dim(X)$. These are defined by first choosing a metric on on E, which is always possible, and then taking a $O(r)$-connection $D = \mathrm{d} + \omega$ which preserves the metric. If R is the corresponding field strength then

$$\begin{aligned} p(W) \equiv p_0(W) + p_1(W) + \ldots &= \det\left(\mathrm{id} - \tfrac{R}{2\pi}\right) \\ &= 1 - \tfrac{1}{8\pi^2} \operatorname{tr} R \wedge R + \ldots, \end{aligned} \tag{H.5}$$

Since $R^{\mathrm{T}} = -R$ these are non-zero only for even powers of R and so $p_j(E) \in H^{4j}(X, \mathbb{Z})$.

If we can decompose the complexification of W as $W_{\mathbb{C}} = E \oplus \bar{E}$ where E is a complex vector bundle, then there is a relation between the Pontrjagin classes of W and the Chern classes of E. In particular

$$\begin{aligned} p_1(W) &= \mathrm{c}_1^2(E) - 2\,\mathrm{c}_2(E), \\ p_2(W) &= \mathrm{c}_2^2(E) - 2\,\mathrm{c}_1(E) \cdot \mathrm{c}_3(E) + 2\,\mathrm{c}_4(E). \end{aligned} \tag{H.6}$$

Other classes which commonly appear are most easily defined by considering F (or R) diagonalized on each fibre

$$F/2\pi = \mathrm{diag}(x_1, \ldots, x_r). \tag{H.7}$$

We then have[63,74]

$$\begin{aligned} &\text{Chern character} \quad & \mathrm{ch}(E) &= \sum_i \mathrm{e}^{x_i} = \operatorname{tr} \mathrm{e}^{F/2\pi} \\ & & &= r + \mathrm{c}_1 + \tfrac{1}{2}(\mathrm{c}_1^2 - 2\,\mathrm{c}_2) + \ldots, \\ &\text{Todd class} \quad & \mathrm{td}(E) &= \prod_i \frac{x_i}{1 + \mathrm{e}^{-x_i}} \\ & & &= 1 + \tfrac{1}{2}\,\mathrm{c}_1 + \tfrac{1}{12}(\mathrm{c}_1^2 + \mathrm{c}_2) + \ldots, \\ &\text{A-roof genus} \quad & \hat{\mathrm{A}}(W) &= \prod_i \frac{x_i/2}{\sinh(x_i/2)} \\ & & &= 1 - \tfrac{1}{24} p_1 + \tfrac{1}{5760}(7p_1^2 - 4p_2) + \ldots, \\ &\text{Hirzebruch L-poly.} \quad & \mathrm{L}(W) &= \prod_i \frac{x_i}{\tanh x_i} \\ & & &= 1 + \tfrac{1}{3} p_1 + \tfrac{1}{45}(-p_1^2 + 7p_2) + \ldots. \end{aligned} \tag{H.8}$$

The Chern character has the particular property that $\mathrm{ch}(E_1 \otimes E_2) = \mathrm{ch}(E_1) \cdot \mathrm{ch}(E_2)$ and $\mathrm{ch}(E_1 \oplus E_2) = \mathrm{ch}(E_1) + \mathrm{ch}(E_2)$. All these classes appear in index theorems.[63]

Finally we note that for a complex manifold we can always define the Chern classes of the holomorphic part of the tangent bundle $c_p(TX_c)$. (Usually this is written just as $c_p(TX)$ or $c_p(X)$.) In terms of the Riemann tensor we have

$$\begin{aligned} c_1(TX) &= \tfrac{1}{2\pi}\mathcal{R}, \\ c_2(TX) &= \tfrac{1}{16\pi^2}\operatorname{tr} R\wedge R + \tfrac{1}{8\pi^2}\mathcal{R}\wedge\mathcal{R}, \end{aligned} \tag{H.9}$$

where $\mathcal{R}$ is the Ricci form $\mathcal{R}_{AB} = \frac{1}{2}R_{AB}{}^{C}{}_{D}J^{D}{}_{C}$, with J the complex structure. Recall that on a Calabi–Yau manifold $c_1(TX) = 0$ and hence $c_2(TX) = -\frac{1}{2}p_1(TX)$.

Appendix I. Sheaves

This section is a brief and somewhat sketchy introduction to sheaves. More detail can be found in Griffiths and Harris[71] or Hartshorne.[75] There is also a very good recent review of the connections between sheaves, derived categories and D-branes by Sharpe.[39]

A sheaf is a sort of generalized vector bundle defined by its sections. Formally we have

> A sheaf $\mathcal{F}$ associates to each open set $U \subset X$ a *group of sections* $\mathcal{F}(U)$, and for each pair $V \subset U$ a *restriction map* $r_{U,V} : \mathcal{F}(U) \to \mathcal{F}(V)$.

The restriction map, which tells us how to match sections on different patches has to satisfy various properties:

(1) For any $W \subset V \subset U$ we have

$$r_{U,W} = r_{V,W} \cdot r_{U,V}.$$

(2) For any U and V and sections $\sigma \in \mathcal{F}_U$ and $\tau \in \mathcal{F}_V$, if $r_{U,U\cap V}(\sigma) = r_{V,U\cap V}(\tau)$ then there exists $\rho \in \mathcal{F}(U\cup V)$ such that

$$r_{U\cup V,U}(\rho) = \sigma, \quad r_{U\cup V,V}(\rho) = \tau,$$

(3) If $\sigma \in \mathcal{F}(U\cup V)$ and

$$r_{U\cup V,U}(\sigma) = r_{U\cup V,V}(\sigma) = 0$$

then $\sigma = 0$.

Essentially these just allow one to think of a section σ defined over the whole space X, with a unique restriction $\sigma|_U$ to any given patch. The group of sections over a given point $p \in X$ is called a *stalk*. Note the definition is very general. It can also be generalized to define sheaves of rings, sets of modules etc instead of groups. We will always be thinking of the sections as functions or particular sections of vector bundles.

Clearly, sections of a vector bundle form a sheaf. We will be most interested in holomorphic objects. Given a holomorphic bundle E over X, we write

$$\mathcal{O}_X(E) = \text{holomorphic sections of } E.$$

The simplest case is when E is the trivial line bundle. The corresponding sheaf, built from holomorphic functions on X, is written $\mathcal{O}_X$ and is call the *structure sheaf* of X. Note given a divisor D we write $\mathcal{O}_X(D)$ for both the corresponding line bundle (see Eq. (G.5)) and the corresponding sheaf of holomorphic sections of that bundle. Often the subscript X is dropped. In general, the nomenclature is that

$$\textit{locally-free} \text{ sheaf} \Leftrightarrow \text{holomorphic bundle } E.$$

Vector bundles give sheaves where the dimension of the stalk never changes. More generally one could consider, for instance, given a complex sub-manifold V, the *ideal sheaf*,

$$\mathcal{I}_V = \text{holomorphic functions vanishing on } V.$$

Away from V, the stalks are just copies of $\mathbb{C}$ and the sheaf just looks like $\mathcal{O}_X$, however on V the stalk is just a single point. From this point of view, we see that a sheaf is like a vector bundle where the dimension of the fibre can change. This is shown in Fig. 12.

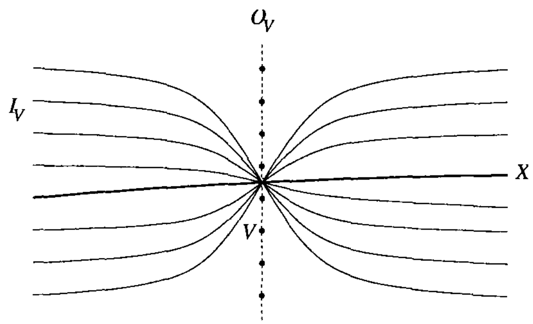


Figure 12. The sheaves $\mathcal{I}_V$ and $\mathcal{O}_V$

Even more radically, we could define a sheaf $\mathcal{O}_V$ (the structure sheaf of V) by, for each open set $U \subset X$,

$$\mathcal{O}_V(U) = \begin{cases} \text{holomorphic functions on } U \cap V & \text{if } U \cap V \neq \emptyset \\ 0 & \text{otherwise} \end{cases}$$

Over most points in X the stalks of $\mathcal{O}_V$ are zero, but over V they are $\mathbb{C}$. Thus we have a sort of line bundle localized on V. The most extreme case where V, a single point p, we get the so-called *skyscraper sheaf* $\mathcal{O}_p$, which has zero stalk everywhere except at p.

In general one defines the *support* of a sheaf $\mathcal{F}$

$$\mathrm{supp}(\mathcal{F}) = \{x \in X : \text{stalk over } x \text{ is not zero}\}.$$

so that, for instance, $\mathrm{supp}\, \mathcal{O}_V = V$. One also defines

$$\textit{torsion} \text{ sheaf } \mathcal{F} \Leftrightarrow \mathrm{supp}(\mathcal{F}) \text{ is a submanifold of } X.$$

Thus, a torsion sheaf has stalks which are generically zero.

Given a continuous map $i : Y \to X$, we can *push forward* sheaves from Y to X by, for each open set $U \in X$,

$$i_*\mathcal{F}(U) = \mathcal{F}(i^{-1}(U)), \tag{I.1}$$

defining the group of sections over U as the corresponding group of sections over $i^{-1}(U)$. Given a vector bundle E on a submanifold V with the embedding $i : V \hookrightarrow X$, we can take then push-forward the vector bundle to a sheaf i_*E on X. Clearly, if $U \cap V = \emptyset$ then $i_*E(U) = 0$, since $i^{-1}(U) = \emptyset$. We have $\mathrm{supp}(i_*E) = V$ and i_*E is effectively a vector bundle localized on V.

We can also define maps of sheaves in terms of maps of their sections. For instance we have an inclusion map $i : \mathcal{I}_V \to \mathcal{O}_X$ and also a restriction map $r : \mathcal{O}_X \to \mathcal{O}_V$. In general we have a *exact sequence* of sheaves $\mathcal{F}_i$ is a set of maps α_i

$$\cdots \to \mathcal{F}_n \xrightarrow{\alpha_n} \mathcal{F}_{n+1} \xrightarrow{\alpha_{n+1}} \mathcal{F}_n \to \ldots, \tag{I.2}$$

with $\alpha_{n+1} \circ \alpha_n = 0$. The maps i and r form a *short exact sequence*

$$0 \to \mathcal{I}_V \xrightarrow{i} \mathcal{O}_X \xrightarrow{r} \mathcal{O}_V \to 0, \tag{I.3}$$

since $r \circ i = 0$. As for any short exact sequence, this means that $\mathcal{I}_V$ is a *sub-sheaf* of $\mathcal{O}_X$ and $\mathcal{O}_V$ is the *quotient sheaf* $\mathcal{O}_X/\mathcal{I}_V$, that is holomorphic functions on X modulo holomorphic functions on X which vanish on V.

It will be useful to define two further classes of sheaf. First we have *coherent* sheaves. A coherent sheaf is one that can be realized as a finite *resolution* in terms of vector bundles. This means there is a finite exact sequence

$$0 \to E_n \to E_{n-1} \to \cdots \to E_0 \to \mathcal{F} \to 0 \tag{I.4}$$

where E_i are holomorphic vector bundles. Such a resolution allows us to define Chern classes of $\mathcal{F}$ by

$$\mathrm{c}_p(\mathcal{F}) = \sum_{i=0}^{n} (-1)^i \, \mathrm{c}_p(E_i). \tag{I.5}$$

It can be shown that this is independent of the particular resolution.

Finally we define a *torsion-free coherent sheaf* as a coherent sheaf which has no torsion sub-sheaves. If $\mathrm{vect}(X)$ is the space of holomorphic vector bundles on X, $\mathrm{coh}(X)$ the space of coherent sheaves, and $\mathrm{coh}_{\mathrm{tf}}(X)$ the space of torsion-free coherent sheaves then

$$\mathrm{vect}(X) \subset \mathrm{coh}_{\mathrm{tf}}(X) \subset \mathrm{coh}(X). \tag{I.6}$$

The key point is that for coherent sheaves, since we can define $\mathrm{c}_1(\mathcal{F})$ we can define the slope of the sheaf as in Eq. (33). Furthermore, if we have a torsion-free coherent sheaf we can sensibly define stability, since any sub-sheaf is torsion-free and we then don't have to worry about the slope of torsion sub-sheaves. In other words, there is a sensible notion of a supersymmetric torsion-free coherent sheaf.

Appendix J. Properties of del Pezzo surfaces

A del Pezzo surface is a complex manifold B of complex dimension two the canonical bundle K_B of which is negative. This means that the class of the dual anticanonical bundle K_B^{-1} has positive intersection with every curve in the surface. The del Pezzo surfaces which will concern us in this paper are the surfaces dP_r constructed from complex projective space $\mathbb{C}P^2$ by blowing up r points $p_1, \ldots, p_r$ in general position where $r = 0, 1, \ldots, 8$. We will not discuss the process of blowing up in detail, but just recall that it is a procedure by which we replace a point in $\mathbb{C}P^2$ with a two-sphere $\mathbb{C}P^1$. The sphere has self-intersection -1 and is called an "exceptional divisor". The homology class of the sphere has only one representative, the sphere itself. More details can be found in Griffiths and Harris.[71]

We will also consider the rational elliptic surface, which we denote dP_9, although it is not a del Pezzo surface in the strict sense. It can be obtained as the blow-up of $\mathbb{C}P^2$ at nine points which form the complete intersection of two cubic curves, and which are otherwise in general position. (If (x, y, z) are homogeneous coordinates on $\mathbb{C}P^2$, then the cubic curves are given by a pair of homogeneous cubic equations $f(x, y, z) = g(x, y, z) = 0$. As we saw for the Weierstrass model in Sec. 6.1, cubic equations generically describe two-tori.) For a dP_9 surface, the anticanonical bundle is no longer positive but, rather, it is "nef", which means that its intersection with every curve on the dP_9 surface is non-negative. In fact, a dP_9 surface is elliptically fibered over $\mathbb{C}P^1$ and the elliptic fibers (cubic curves of the form $af + bg = 0$, which all pass though the nine blown up points) are in the anticanonical class. This description fails when the nine points are in completely general position, which is why we require them to be the complete intersection of two cubics.

Of particular interest is the homology group of curves $H_2(dP_r, \mathbb{Z})$ on the del Pezzo surface. First we note that $H_2(\mathbb{C}P^2, \mathbb{Z}) = \mathbb{Z}$. There is just a single class l corresponding to a line $ax + by + cz = 0$. Any higher order homogeneous equation always has a limit where it factors into a product of lines. Thus all curves (divisors) in $\mathbb{C}P^2$ have homology classes which are just multiples of the "hyperplane" divisor l. (This is true for any $\mathbb{C}P^n$.) Since a new cycle is created each time a point is blown up, we see that the dimension of $H_2(dP_r, \mathbb{Z})$ is thus $b_2 = r + 1$. Explicitly, we write l for the class inherited from $\mathbb{C}P^2$. The blow-up of the i-th point p_i corresponds to an exceptional divisor e_i. Hence, for dP_r, there are r exceptional divisors e_i, $i = 1, \ldots, r$. The curves l and e_i where $i = 1, \ldots, r$ form a basis of homology classes. Since $h^{2,0} = 0$ the Lefschetz theorem relates elements of $H^{1,1}$ to algebraic classes, and we have $H_2(dP_r, \mathbb{Z}) = H_2(dP_r, \mathbb{Z})_{\text{alg}}$. A particularly important element of $H_2(dP_r, \mathbb{Z})$ is the anticanonical class $F = -c_1(K_{dP_r})$, given by

$$F = -c_1(K_{dP_r}) = 3l - \sum e_i. \tag{J.1}$$

Let us consider the intersection numbers of the basis of curves l and e_i, $i = 1, \ldots, r$ of $H_2(dP_r, \mathbb{Z})$. Any two lines in $\mathbb{C}P^2$ generically intersect once. As we mentioned exceptional divisors have self-intersect -1. Furthermore, two exceptional divisors at different points cannot intersect. Finally, a generic line in $\mathbb{C}P^2$ will miss all the blow-up points. Thus we expect

$$l \cdot l = 1, \qquad e_i \cdot e_j = -\delta_{ij}, \qquad e_i \cdot l = 0. \tag{J.2}$$

It is important to explicitly know the set of effective divisors on dP_r. By definition, l and e_i are effective, as is the anticanonical class F. Now consider a line l in $\mathbb{C}P^2$ which passes through the i-th blown up point p_i. Such a line is still effective. The class of such a curve is given by $l - e_i$ and, hence, this is an effective divisor for any $i = 1, \ldots, r$. In general, a line can pass through at most two points, say p_i and p_j where $i \neq j$. The properties of blow-ups then imply that the class of such a curve is

$$l - e_i - e_j \tag{J.3}$$

which, by construction, is an effective divisor for any $i \neq j = 1, \ldots, r$.

In general, a class $\alpha \in H_2(dP_r, \mathbb{Z})$ is called exceptional if it satisfies

$$\alpha \cdot \alpha = -1, \qquad \alpha \cdot F = 1. \tag{J.4}$$

The classes of the exceptional curves e_i certainly are of this type, but there are others; for example, the class $l - e_i - e_j$ just described satisfies these properties as well. In fact, any exceptional class on a general del Pezzo surface is the class of a unique, irreducible, non-singular curve which is the image of some curve on $\mathbb{C}P^2$, perhaps passing through some number of blow-up points (or is the blow-up itself). Each such curve in dP_r is in fact a $\mathbb{C}P^1$. Even though this is not apparent from our description, all these exceptional curves look exactly alike and, in fact, can be interchanged by the Weyl group which acts as a symmetry group of the family of del Pezzo surfaces. So, for example, our del Pezzo surface admits another description, in which the line $l - e_i - e_j$ appears as the blow-up of some point, while one or more of the exceptional divisors e_i appears as a line or higher degree curve.

For $r \leq 4$, all exceptional curves are of the types already discussed. But consider, for $r \geq 5$, a conic in $\mathbb{C}P^2$; that is, a curve defined by a quadratic equation. The conic is denoted by $2l$. A conic can pass through at most five blown up points, say p_i, p_j, p_k, p_l and p_m. If they are all different, then the curves

$$2l - e_i - e_j - e_k - e_l - e_m \tag{J.5}$$

are exceptional divisors. These are easily seen to be effective as well. Similarly, consider a cubic in $\mathbb{C}P^2$; that is, a curve defined by a cubic equation. The cubic is denoted by $3l$. When $r = 7$, 8 or 9, a cubic can be chosen to pass through one of the blown up points, say p_i, twice (that is, it will be a singular cubic curve, with singular point at p_i), while also passing (once) through six more of the blown up points, say p_j, p_k, p_l, p_m, p_n and p_o. Therefore, we see that, for $r = 7$, 8 or 9, we also get exceptional divisors of the form

$$3l - 2e_i - e_j - e_k - e_l - e_m - e_n - e_o \tag{J.6}$$

where all the points are different. Again, these are easily seen to be effective classes. Yet more examples of exceptional curves are obtained, for $r = 8$ or 9, by considering appropriate plane curves of degrees 4, 5 or 6. The complete list of exceptional curves for $r \leq 8$ can be found, for example, in Table 3, page 35 of Demazure.[76] All these classes are effective.

We can now complete the description of the set of effective classes on a del Pezzo surface. These classes are precisely the linear combinations, with non-negative integer coefficients, of the anticanonical class F and of the exceptional classes, including the e_i, the curves in (J.3), (J.5) and (J.6), and their more complicated cousins for large r. For $r \leq 8$ this gives us an explicit, finite set which generates the Mori cone. (Note also that the exceptional classes are actually the roots of the E_r lattice.)

The above statement, that is, that the effective classes are precisely the linear combinations, with non-negative integer coefficients, of the anticanonical class F and of the exceptional classes, remains true for the rational elliptic surface dP_9. The new and, perhaps, surprising feature is that on a dP_9 surface there are infinitely many exceptional classes. This is easiest to see using the elliptic fibration structure. Each of the nine exceptional divisors e_i has intersection number 1 with the elliptic fiber F, so it gives a section of the fibration. Conversely, it is easy to see that any section is an exceptional curve. But since each fiber, an elliptic curve, is a group, it follows that the set of sections is itself a group under the operation of pointwise addition of sections. We are free to designate one of our nine sections, say e_9, as the "zero" section. The other eight sections then generate an infinite group of sections, which generically will be $\mathbb{Z}^8$. The Mori cone in this case is not generated by any finite set of effective curves.

Finally, we list the formulas for the Chern classes on dP_r. We find that

$$\begin{aligned} \mathrm{c}_1(dP_r) &= -\,\mathrm{c}_1(K_{dP_r}) = 3l - \sum e_i \\ \mathrm{c}_2(dP_r) &= 3 + r \end{aligned} \tag{J.7}$$

are the first and second Chern classes of dP_r respectively. The second Chern class is simply a number, since there is only one class in $H_0(dP_r, \mathbb{Z})$.

References

1. B. A. Ovrut's TASI 2001 lectures, in this volume.

2. P. Horava and E. Witten, "Heterotic and type I string dynamics from eleven dimensions," Nucl. Phys. B **460** (1996) 506 [arXiv:hep-th/9510209];
P. Horava and E. Witten, "Eleven-Dimensional Supergravity on a Manifold with Boundary," Nucl. Phys. B **475** (1996) 94 [arXiv:hep-th/9603142].
3. See for instance Green, Schwarz and Witten[11] or Polchinski[28] or for a review of convention string models see:
K. R. Dienes, "String Theory and the Path to Unification: A Review of Recent Developments," Phys. Rept. **287** (1997) 447 [arXiv:hep-th/9602045].
4. R. Friedman, J. Morgan and E. Witten, "Vector bundles and F theory," Commun. Math. Phys. **187**, 679 (1997). [hep-th/9701162].
5. R. Donagi, Asian J. Math. **1**, 214 (1997).
6. M. Bershadsky, A. Johansen, T. Pantev and V. Sadov, "On four-dimensional compactifications of F-theory," Nucl. Phys. **B505**, 165 (1997) [hep-th/9701165].
7. There is a very large literature on this subject. For more recent papers see the following and the references therein:
R. Blumenhagen, L. Goerlich, B. Kors and D. Lust, JHEP **0010** (2000) 006;
C. Angelantonj, I. Antoniadis, E. Dudas and A. Sagnotti, Phys. Lett. B **489** (2000) 223;
G. Aldazabal, S. Franco, L. E. Ibanez, R. Rabadan and A. M. Uranga, J. Math. Phys. **42** (2001) 3103;
G. Aldazabal, S. Franco, L. E. Ibanez, R. Rabadan and A. M. Uranga, JHEP **0102** (2001) 047;
R. Blumenhagen, B. Kors and D. Lust, JHEP **0102** (2001) 030;
R. Blumenhagen, B. Kors, D. Lust and T. Ott, Nucl. Phys. B **616** (2001) 3;
C. Kokorelis, JHEP **0208** (2002) 018;
M. Cvetic, G. Shiu and A. M. Uranga, Nucl. Phys. B **615** (2001) 3.
8. E. Witten, "Strong Coupling Expansion Of Calabi-Yau Compactification," Nucl. Phys. **B471**, 135 (1996) [hep-th/9602070].
9. A. Lukas, B. A. Ovrut, K. S. Stelle and D. Waldram, "The universe as a domain wall," Phys. Rev. **D59**, 086001 (1999) [hep-th/9803235];
A. Lukas, B. A. Ovrut, K. S. Stelle and D. Waldram, "Heterotic M-theory in five dimensions," Nucl. Phys. **B552**, 246 (1999) [hep-th/9806051].
10. A. Lukas, B. A. Ovrut and D. Waldram, "Non-standard embedding and five-branes in heterotic M-theory," Phys. Rev. D **59** (1999) 106005 [arXiv:hep-th/9808101].
11. M. Green, J. Schwarz and E. Witten, Superstring Theory, vol. 2, Cambridge University Press, Cambridge UK (1987).
12. J. O. Conrad, "Brane tensions and coupling constants from within M-theory," Phys. Lett. B **421** (1998) 119 [arXiv:hep-th/9708031];
T. Harmark, "Coupling constants and brane tensions from anomaly cancellation in M-theory," Phys. Lett. B **431** (1998) 295 [arXiv:hep-th/9802190];
A. Bilal, J. P. Derendinger and R. Sauser, "M-theory on S(1)/Z(2): Facts and fakes," Nucl. Phys. B **576** (2000) 347 [arXiv:hep-th/9912150].
13. T. J. Li, J. L. Lopez and D. V. Nanopoulos, "M-theory inspired no-scale supergravity," Mod. Phys. Lett. A **12** (1997) 2647 [arXiv:hep-ph/9702237];

E. Dudas and C. Grojean, "Four-dimensional M-theory and supersymmetry breaking," Nucl. Phys. B **507** (1997) 553 [arXiv:hep-th/9704177];
H. P. Nilles, M. Olechowski and M. Yamaguchi, "Supersymmetry breaking and soft terms in M-theory," Phys. Lett. B **415** (1997) 24 [arXiv:hep-th/9707143];
I. Antoniadis and M. Quiros, "On the M-theory description of gaugino condensation," Phys. Lett. B **416** (1998) 327 [arXiv:hep-th/9707208];
Z. Lalak and S. Thomas, "Gaugino condensation, moduli potential and supersymmetry breaking in M-theory models," Nucl. Phys. B **515** (1998) 55 [arXiv:hep-th/9707223];
A. Lukas, B. A. Ovrut and D. Waldram, "On the four-dimensional effective action of strongly coupled heterotic string theory," Nucl. Phys. B **532** (1998) 43 [arXiv:hep-th/9710208].

14. E. Witten, "Symmetry Breaking Patterns In Superstring Models," Nucl. Phys. B **258** (1985) 75;
A. Sen, "The Heterotic String In Arbitrary Background Field," Phys. Rev. D **32** (1985) 2102;
J. D. Breit, B. A. Ovrut and G. C. Segre, "E(6) Symmetry Breaking In The Superstring Theory," Phys. Lett. B **158** (1985) 33.
15. T. Banks and M. Dine, "Couplings and Scales in Strongly Coupled Heterotic String Theory," Nucl. Phys. B **479** (1996) 173 [arXiv:hep-th/9605136].
16. R. Donagi, A. Lukas, B. A. Ovrut and D. Waldram, "Non-perturbative vacua and particle physics in M-theory," JHEP **9905** (1999) 018 [arXiv:hep-th/9811168].
17. R. Donagi, A. Lukas, B. A. Ovrut and D. Waldram, "Holomorphic vector bundles and non-perturbative vacua in M-theory," JHEP **9906** (1999) 034 [arXiv:hep-th/9901009].
18. R. Donagi, B. A. Ovrut, T. Pantev and D. Waldram, "Standard models from heterotic M-theory," Adv. Theor. Math. Phys. **5** (2002) 93 [arXiv:hep-th/9912208].
19. R. Donagi, B. A. Ovrut, T. Pantev and D. Waldram, "Standard-model bundles on non-simply connected Calabi-Yau threefolds," JHEP **0108** (2001) 053 [arXiv:hep-th/0008008];
R. Donagi, B. A. Ovrut, T. Pantev and D. Waldram, "Standard-model bundles," Adv. Theor. Math. Phys. **5** (2002) 563 [arXiv:math.ag/0008010];
R. Donagi, B. A. Ovrut, T. Pantev and D. Waldram, "Spectral involutions on rational elliptic surfaces," Adv. Theor. Math. Phys. **5** (2002) 499 [arXiv:math.ag/0008011].
20. S. Donaldson, Proc. London Math. Soc. **3**, 1 (1985).
21. K. Uhlenbeck and S.-T. Yau, Comm. Pure App. Math. **39**, 275 (1986), **42**, 703 (1986).
22. K. Becker, M. Becker and A. Strominger, "Five-branes, membranes and nonperturbative string theory," Nucl. Phys. B **456** (1995) 130 [arXiv:hep-th/9507158];
M. Bershadsky, C. Vafa and V. Sadov, "D-Branes and Topological Field Theories," Nucl. Phys. B **463** (1996) 420 [arXiv:hep-th/9511222].

23. I. Bandos, K. Lechner, A. Nurmagambetov, P. Pasti, D. P. Sorokin and M. Tonin, "Covariant action for the super-five-brane of M-theory," Phys. Rev. Lett. **78** (1997) 4332 [arXiv:hep-th/9701149];
M. Aganagic, J. Park, C. Popescu and J. H. Schwarz, "World-volume action of the M-theory five-brane," Nucl. Phys. B **496** (1997) 191 [arXiv:hep-th/9701166].
24. See for instance: P. Claus, R. Kallosh and A. Van Proeyen, "M 5-brane and superconformal (0,2) tensor multiplet in 6 dimensions," Nucl. Phys. B **518** (1998) 117 [arXiv:hep-th/9711161].
25. E. Witten, "Solutions of four-dimensional field theories via M-theory," Nucl. Phys. B **500** (1997) 3 [arXiv:hep-th/9703166];
A. Klemm, W. Lerche, P. Mayr, C. Vafa and N. P. Warner, "Self-Dual Strings and $N = 2$ Supersymmetric Field Theory," Nucl. Phys. B **477** (1996) 746 [arXiv:hep-th/9604034].
26. E. Witten, "Small Instantons in String Theory," Nucl. Phys. B **460** (1996) 541 [arXiv:hep-th/9511030];
O. J. Ganor and A. Hanany, "Small E_8 Instantons and Tensionless Noncritical Strings," Nucl. Phys. B **474** (1996) 122 [arXiv:hep-th/9602120];
N. Seiberg and E. Witten, "Comments on String Dynamics in Six Dimensions," Nucl. Phys. B **471** (1996) 121 [arXiv:hep-th/9603003];
S. Kachru and E. Silverstein, "Chirality-changing phase transitions in 4d string vacua," Nucl. Phys. B **504** (1997) 272 [arXiv:hep-th/9704185];
B. A. Ovrut, T. Pantev and J. Park, "Small instanton transitions in heterotic M-theory," JHEP **0005** (2000) 045 [arXiv:hep-th/0001133].
27. For a review of grand unified theories, see for instance:
R. N. Mohapatra, "Unification and supersymmetry," TASI 97: Supersymmetry, supersymmetry and supercolliders, ed. J. Bagger, World Scientific, Singapore (1998). [arXiv:hep-ph/9911272].
28. J. Polchinski, String Theory, vol 1 and 2, Cambridge University Press, Cambridge, UK (1998).
29. W. I. Taylor, "D-brane field theory on compact spaces," Phys. Lett. B **394** (1997) 283 [arXiv:hep-th/9611042].
30. M. Cederwall, A. von Gussich, B. E. W. Nilsson and A. Westerberg, "The Dirichlet super-three-brane in ten-dimensional type IIB supergravity," Nucl. Phys. B **490** (1997) 163 [arXiv:hep-th/9610148];
M. Cederwall, A. von Gussich, B. E. W. Nilsson, P. Sundell and A. Westerberg, "The Dirichlet super-p-branes in ten-dimensional type IIA and IIB supergravity," Nucl. Phys. B **490** (1997) 179 [arXiv:hep-th/9611159];
E. Bergshoeff and P. K. Townsend, "Super D-branes," Nucl. Phys. B **490** (1997) 145 [arXiv:hep-th/9611173].
31. M. Marino, R. Minasian, G. W. Moore and A. Strominger, "Nonlinear instantons from supersymmetric p-branes," JHEP **0001** (2000) 005 [arXiv:hep-th/9911206].
32. M. Bershadsky, C. Vafa and V. Sadov, "D-Strings on D-Manifolds," Nucl. Phys. B **463** (1996) 398 [arXiv:hep-th/9510225];
M. B. Green, J. A. Harvey and G. W. Moore, "I-brane inflow and anoma-

lous couplings on D-branes," Class. Quant. Grav. **14** (1997) 47 [arXiv:hep-th/9605033];
R. Minasian and G. W. Moore, "K-theory and Ramond-Ramond charge," JHEP **9711** (1997) 002 [arXiv:hep-th/9710230];
J. F. Morales, C. A. Scrucca and M. Serone, "Anomalous couplings for D-branes and O-planes," Nucl. Phys. B **552** (1999) 291 [arXiv:hep-th/9812071].

33. R. C. Myers, "Dielectric-branes," JHEP **9912** (1999) 022 [arXiv:hep-th/9910053].
34. J. A. Harvey and G. W. Moore, "On the algebras of BPS states," Commun. Math. Phys. **197** (1998) 489 [arXiv:hep-th/9609017].
35. S. Mukai, "Duality between $D(X)$ and $D(\hat{X})$ with its application to Picard sheaves," Nagoya Math. J. **81** (1981) 153.
36. C. Bartocci, U. Bruzzo, D. Hernandez Ruiperez and J. M. Munoz Porras, "Mirror symmetry on elliptic K3 surfaces via Fourier-Mukai transform," Commun. Math. Phys. **195** (1998) 79.
37. P. S. Aspinwall and R. Y. Donagi, Adv. Theor. Math. Phys. **2** (1998) 1041 [arXiv:hep-th/9806094].
38. R.P. Thomas, "Derived categories for the working mathematician," arXiv:math.ag/0001045.
39. E. Sharpe, "Lectures on D-branes and sheaves," arXiv:hep-th/0307245.
40. A. Sen, "Tachyon condensation on the brane antibrane system," JHEP **9808** (1998) 012 [arXiv:hep-th/9805170].
41. E. Witten, "D-branes and K-theory," JHEP **9812** (1998) 019 [arXiv:hep-th/9810188];
for a review of K-theory and D-branes see G. Moore, "K-theory from a physical perspective," arXiv:hep-th/0304018.
42. M. R. Douglas, "D-branes, categories and $N = 1$ supersymmetry," J. Math. Phys. **42** (2001) 2818 [arXiv:hep-th/0011017]. E. R. Sharpe, "D-branes, derived categories, and Grothendieck groups," Nucl. Phys. B **561** (1999) 433 [arXiv:hep-th/9902116]. P. S. Aspinwall and A. E. Lawrence, "Derived categories and zero-brane stability," JHEP **0108** (2001) 004 [arXiv:hep-th/0104147].
43. M. Kontsevich, "Homological algebra of mirror symmetry," in Proceedings of the International Congress of Mathematicians, Birkhäuser (1995) [alg-geom/9411018].
44. M. R. Douglas, "Lectures on D-branes on Calabi-Yau manifolds," Prepared for ICTP Spring School on Superstrings and Related Matters, Trieste, Italy, 2-10 Apr 2001
http://www.ictp.trieste.it/~pub_off/lectures/lns007/Douglas/Douglas.pdf.
45. C. Vafa, "Evidence for F-Theory," Nucl. Phys. B **469** (1996) 403 [arXiv:hep-th/9602022].
46. D. R. Morrison and C. Vafa, Nucl. Phys. B **473** (1996) 74 [arXiv:hep-th/9602114];
D. R. Morrison and C. Vafa, Nucl. Phys. B **476** (1996) 437 [arXiv:hep-th/9603161].
47. See Vafa[45] or for instance: P. S. Aspinwall and D. R. Morrison, "Point-

like instantons on K3 orbifolds," Nucl. Phys. B **503** (1997) 533 [arXiv:hep-th/9705104].

48. K. Kodaira, Annals of Math. **77** (1963) 563; Annals of Math. **78** (1963) 1.
49. S. Gubser's TASI 2001 lectures, in this volume.
50. A. Grassi, Int. J. Math. **4** (1993) 203.
51. P. Berglund and P. Mayr, "Heterotic string/F-theory duality from mirror symmetry," Adv. Theor. Math. Phys. **2** (1999) 1307 [arXiv:hep-th/9811217]; P. Berglund and P. Mayr, "Stability of vector bundles from F-theory," JHEP **9912** (1999) 009 [arXiv:hep-th/9904114].
52. G. Rajesh, "Toric geometry and F-theory/heterotic duality in four dimensions," JHEP **9812** (1998) 018 [arXiv:hep-th/9811240].
53. W. Fulton, Intersection theory, vol. 2 of A Series of Modern Surveys in Mathematics, Springer–Verlag, New York (1997).
54. G. Curio, "Chiral matter and transitions in heterotic string models," Phys. Lett. B **435** (1998) 39 [arXiv:hep-th/9803224].
55. B. Andreas, "On vector bundles and chiral matter in N = 1 heterotic compactifications," JHEP **9901** (1999) 011 [arXiv:hep-th/9802202].
56. 5brane moduli M. Brandle and A. Lukas, "Five-branes in heterotic braneworld theories," Phys. Rev. D **65** (2002) 064024 [arXiv:hep-th/0109173].
57. See references in Burt Ovrut's lectures[1] as well as papers listed above.[13]
58. NPB 258 (1985) E. Witten, "Symmetry Breaking Patterns In Superstring Models," Nucl. Phys. B **258** (1985) 75.
59. B. A. Ovrut, T. Pantev and R. Reinbacher, "Torus-fibered Calabi-Yau threefolds with non-trivial fundamental group," JHEP **0305** (2003) 040 [arXiv:hep-th/0212221];
    B. A. Ovrut, T. Pantev and R. Reinbacher, "Invariant homology on standard model manifolds," arXiv:hep-th/0303020.
60. R. Donagi, B. A. Ovrut, T. Pantev and R. Reinbacher, "SU(4) instantons on Calabi-Yau threefolds with Z(2) x Z(2) fundamental group," arXiv:hep-th/0307273.
61. R. Donagi, B. A. Ovrut and D. Waldram, "Moduli spaces of fivebranes on elliptic Calabi-Yau threefolds," JHEP **9911** (1999) 030 [arXiv:hep-th/9904054].
62. K. Becker, M. Becker and A. Strominger, "Five-branes, membranes and nonperturbative string theory," Nucl. Phys. B **456** (1995) 130 [arXiv:hep-th/9507158];
    J. A. Harvey and G. W. Moore, "Superpotentials and membrane instantons," arXiv:hep-th/9907026;
    G. W. Moore, G. Peradze and N. Saulina, "Instabilities in heterotic M-theory induced by open membrane instantons," Nucl. Phys. B **607** (2001) 117 [arXiv:hep-th/0012104];
    E. Lima, B. A. Ovrut, J. Park and R. Reinbacher, "Non-perturbative superpotential from membrane instantons in heterotic M-theory," Nucl. Phys. B **614** (2001) 117 [arXiv:hep-th/0101049];
    E. Lima, B. A. Ovrut and J. Park, "Five-brane superpotentials in heterotic M-theory," Nucl. Phys. B **626** (2002) 113 [arXiv:hep-th/0102046].

63. T. Eguchi, P. B. Gilkey and A. J. Hanson, "Gravitation, Gauge Theories And Differential Geometry," Phys. Rept. **66** (1980) 213.
64. D. D. Joyce, Compact manifolds with special holonomy, Oxford Univeristy Press, Oxford UK (2000).
65. R. M. Wald General relativity, Unverversity of Chicago Press, Chicago IL. (1984).
66. J. P. Gauntlett, D. Martelli and D. Waldram, "Superstrings with intrinsic torsion," arXiv:hep-th/0302158.
67. J. P. Gauntlett, "Branes, calibrations and supergravity," arXiv:hep-th/0305074.
68. A.L. Besse, *Einstein manifolds*, Springer-Verlag, New York, (1987).
69. J. Dadok and R. Harvey, "Calibrations and spinors," Acta Math. **170** (1993) 83 ;
R. Harvey, "Spinors and calibrations," Perspectives in Mathematics, vol. 9, Academic Press Inc., Boston, MA (1990).
70. J. P. Gauntlett, N. Kim, D. Martelli and D. Waldram, "Fivebranes wrapped on SLAG three-cycles and related geometry," JHEP **0111** (2001) 018 [arXiv:hep-th/0110034];
J. P. Gauntlett, D. Martelli, S. Pakis and D. Waldram, "G-structures and wrapped NS5-branes," arXiv:hep-th/0205050.
71. P. Griffiths and J. Harris, Principles of algebraic geometry, John Wiley and Sons, Inc., New York, NY (1978).
72. R. Harvey and H.B. Lawson, "Calibrated geometries," Acta Math. **148** (1982) 47.
73. G. W. Gibbons and G. Papadopoulos, "Calibrations and intersecting branes," Commun. Math. Phys. **202** (1999) 593 [arXiv:hep-th/9803163];
J. P. Gauntlett, N. D. Lambert and P. C. West, "Branes and calibrated geometries," Commun. Math. Phys. **202** (1999) 571 [arXiv:hep-th/9803216];
B. S. Acharya, J. M. Figueroa-O'Farrill and B. Spence, "Branes at angles and calibrated geometry," JHEP **9804** (1998) 012 [arXiv:hep-th/9803260].
74. M. Nakahara, Geometry, topology and physics, IOP Publishing Ltd., Bristol, UK (1990).
75. R. Hartshorne, Algebraic geometry, vol 52 of Graduate Texts in Mathematics, Springer–Verlag, New York (1977).
76. M. Demazure, "Surfaces de del Pezzo," Seminaire sur les Singularities des Surfaces, Springer Lecture Notes in Math. vol 777, Springer–Verlag (1980).

This page intentionally left blank

This book covers some recent advances in string theory and extra dimensions. Intended mainly for advanced graduate students in theoretical physics, it presents a rare combination of formal and phenomenological topics, based on the annual lectures given at the School of the Theoretical Advanced Study Institute (2001) — a traditional event that brings together graduate students in high energy physics for an intensive course of advanced learning. The lecturers in the School are leaders in their fields.

The first lecture, by E. D'Hoker and D. Freedman, is a systematic introduction to the gauge–gravity correspondence, focusing in particular on correlation functions in the conformal case. The second, by L. Dolan, provides an introduction to perturbative string theory, including recent advances on backgrounds involving Ramond–Ramond fluxes. The third, by S. Gubser, explains some of the basic facts about special holonomy and its uses in string theory and M-theory. The fourth, by J. Hewett, surveys the TeV phenomenology of theories with large extra dimensions. The fifth, by G. Kane, presents the case for supersymmetry at the weak scale and some of its likely experimental consequences. The sixth, by A. Liddle, surveys recent developments in cosmology, particularly with regard to recent measurements of the CMB and constraints on inflation. The seventh, by B. Ovrut, presents the basic features of heterotic M-theory, including constructions that contain the Standard Model. The eighth, by K. Rajagopal, explains the recent advances in understanding QCD at low temperatures and high densities in terms of color superconductivity. The ninth, by M. Sher, summarizes grand unified theories and baryogenesis, including discussions of supersymmetry breaking and the Standard Model Higgs mechanism. The tenth, by M. Spiropulu, describes collider physics, from a survey of current and future machines to examples of data analyses relevant to theories beyond the Standard Model. The eleventh, by M. Strassler, is an introduction to supersymmetric gauge theory, focusing on Wilsonian renormalization and analogies between three- and four-dimensional theories. The twelfth, by W. Taylor and B. Zwiebach, introduces string field theory and discusses recent advances in understanding open string tachyon condensation. The thirteenth, by D. Waldram, discusses explicit model building in heterotic M-theory, emphasizing the role of the E_8 gauge fields.

The written presentation of these lectures is detailed yet straightforward, and they will be of use to both students and experienced researchers in high-energy theoretical physics for years to come.

ISBN 981-238-788-9

www.worldscientific.com

5495 hc