VOLUME I

AN INTRODUCTION TO THE BOSONIC STRING

STRING THEORY



JOSEPH POLCHINSKI

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK
40 West 20th Street, New York, NY 10011–4211, USA
477 Williamstown Road, Port Melbourne, VIC 3207, Australia
Ruiz de Alarcón 13, 28014 Madrid, Spain
Dock House, The Waterfront, Cape Town 8001, South Africa

http://www.cambridge.org

© Cambridge University Press 1998

This book is in copyright. Subject to statutory exception
and to the provisions of relevant collective licensing agreements,
no reproduction of any part may take place without
the written permission of Cambridge University Press

First published 1998
Reprinted 1999
Reprinted with corrections 2000
Reprinted 2000 (twice), 2002 (twice)

Printed in the United Kingdom at the University Press, Cambridge

Typeface Monotype Times 11/13pt. *System* LaTeX [UPH]

A catalogue record for this book is available from the British Library

Library of Congress Cataloguing in Publication data
Polchinski, Joseph Gerard
String theory / Joseph Polchinski
p cm – (Cambridge monographs on mathematical physics)
Includes bibliographical references
Contents: v. 1 An introduction to the bosonic string
ISBN 0-521-63303-6 (hb)
1 String models I Title II. Series
QC794.6.S85P65 1998
530.14–dc21 98-4545 CIP

ISBN 0 521 63303 6 hardback

Contents

Foreword xiii

Preface xv

Notation xviii

1 A first look at strings 1
1.1 Why strings? 1
1.2 Action principles 9
1.3 The open string spectrum 16
1.4 Closed and unoriented strings 25
Exercises 30

2 Conformal field theory 32
2.1 Massless scalars in two dimensions 32
2.2 The operator product expansion 36
2.3 Ward identities and Noether's theorem 41
2.4 Conformal invariance 43
2.5 Free CFTs 49
2.6 The Virasoro algebra 52
2.7 Mode expansions 58
2.8 Vertex operators 63
2.9 More on states and operators 68
Exercises 74

3 The Polyakov path integral 77
3.1 Sums over world-sheets 77
3.2 The Polyakov path integral 82
3.3 Gauge fixing 84
3.4 The Weyl anomaly 90
3.5 Scattering amplitudes 97
3.6 Vertex operators 101

3.7 Strings in curved spacetime 108
Exercises 118

4 The string spectrum 121
4.1 Old covariant quantization 121
4.2 BRST quantization 126
4.3 BRST quantization of the string 131
4.4 The no-ghost theorem 137
Exercises 143

5 The string S-matrix 145
5.1 The circle and the torus 145
5.2 Moduli and Riemann surfaces 150
5.3 The measure for moduli 154
5.4 More about the measure 159
Exercises 164

6 Tree-level amplitudes 166
6.1 Riemann surfaces 166
6.2 Scalar expectation values 169
6.3 The *bc* CFT 176
6.4 The Veneziano amplitude 178
6.5 Chan–Paton factors and gauge interactions 184
6.6 Closed string tree amplitudes 192
6.7 General results 198
Exercises 204

7 One-loop amplitudes 206
7.1 Riemann surfaces 206
7.2 CFT on the torus 208
7.3 The torus amplitude 216
7.4 Open and unoriented one-loop graphs 222
Exercises 229

8 Toroidal compactification and T-duality 231
8.1 Toroidal compactification in field theory 231
8.2 Toroidal compactification in CFT 235
8.3 Closed strings and T-duality 241
8.4 Compactification of several dimensions 249
8.5 Orbifolds 256
8.6 Open strings 263
8.7 D-branes 268
8.8 T-duality of unoriented theories 277
Exercises 280

9	**Higher order amplitudes**	283
9.1	General tree-level amplitudes	283
9.2	Higher genus Riemann surfaces	290
9.3	Sewing and cutting world-sheets	294
9.4	Sewing and cutting CFTs	300
9.5	General amplitudes	305
9.6	String field theory	310
9.7	Large order behavior	315
9.8	High energy and high temperature	317
9.9	Low dimensions and noncritical strings	323
Exercises		327
	Appendix A: A short course on path integrals	329
A.1	Bosonic fields	329
A.2	Fermionic fields	341
Exercises		345
	References	347
	Glossary	359
	Index	389

CAMBRIDGE MONOGRAPHS ON MATHEMATICAL PHYSICS

J. Ambjørn, B. Durhuus and T. Jonsson *Quantum Geometry: A Statistical Field Theory Approach*
A. M. Anile *Relativistic Fluids and Magneto-Fluids*
J. A. de Azcárraga and J. M. Izquierdo *Lie Groups, Lie Algebras, Cohomology and Some Applications in Physics*†
J. Bernstein *Kinetic Theory in the Early Universe*
G. F. Bertsch and R. A. Broglia *Oscillations in Finite Quantum Systems*
N. D. Birrell and P. C. W. Davies *Quantum Fields in Curved Space*†
D. M. Brink *Semiclassical Methods in Nucleus–Nucleus Scattering*
J. C. Collins *Renormalization*†
M. Creutz *Quarks, Gluons and Lattices*†
F. de Felice and C. J. S. Clarke *Relativity on Curved Manifolds*†
P. G. O. Freund *Introduction to Supersymmetry*†
J. Fuchs *Affine Lie Algebras and Quantum Groups*†
J. Fuchs and C. Schweigert *Symmetries, Lie Algebras and Representations: A Graduate Course for Physicists*
R. Gambini and J. Pullin *Loops, Knots, Gauge Theories and Quantum Gravity*
M. Göckeler and T. Schücker *Differential Geometry, Gauge Theories and Gravity*†
C. Gómez, M. Ruiz Altaba and G. Sierra *Quantum Groups in Two-dimensional Physics*
M. B. Green, J. H. Schwarz and E. Witten *Superstring Theory, volume 1: Introduction*†
M. B. Green, J. H. Schwarz and E. Witten *Superstring Theory, volume 2: Loop Amplitudes, Anomalies and Phenomenology*†
S. W. Hawking and G. F. R. Ellis *The Large-Scale Structure of Space-Time*†
F. Iachello and A. Arima *The Interacting Boson Model*
F. Iachello and P. van Isacker *The Interacting Boson–Fermion Model*
C. Itzykson and J.-M. Drouffe *Statistical Field Theory, volume 1: From Brownian Motion to Renormalization and Lattice Gauge Theory*†
C. Itzykson and J.-M. Drouffe *Statistical Field Theory, volume 2: Strong Coupling, Monte Carlo Methods, Conformal Field Theory, and Random Systems*†
J. I. Kapusta *Finite-Temperature Field Theory*†
V. E. Korepin, A. G. Izergin and N. M. Boguliubov *The Quantum Inverse Scattering Method and Correlation Functions*
M. Le Bellac *Thermal Field Theory*
N. H. March *Liquid Metals: Concepts and Theory*
I. M. Montvay and G. Münster *Quantum Fields on a Lattice*
A. Ozorio de Almeida *Hamiltonian Systems: Chaos and Quantization*†
R. Penrose and W. Rindler *Spinors and Space-time, volume 1: Two-Spinor Calculus and Relativistic Fields*†
R. Penrose and W. Rindler *Spinors and Space-time, volume 2: Spinor and Twistor Methods in Space-Time Geometry*†
V. N. Popov *Functional Integrals and Collective Excitations*†
R. G. Roberts *The Structure of the Proton*†
J. M. Stewart *Advanced General Relativity*
A. Vilenkin and E. P. S. Shellard *Cosmic Strings and Other Topological Defects*
R. S. Ward and R. O. Wells Jr *Twistor Geometry and Field Theories*†
J. Polchinski *String Theory, volume 1: An Introduction to the Bosonic String*
J. Polchinski *String Theory, volume 2: Superstring Theory and Beyond*

† Issued as a paperback

STRING THEORY VOLUME II

Superstring Theory and Beyond

JOSEPH POLCHINSKI
Institute for Theoretical Physics
University of California at Santa Barbara



CAMBRIDGE UNIVERSITY PRESS

String Theory,

Superstring Theory and Beyond

The two volumes that comprise *String Theory* provide an up-to-date, comprehensive, and pedagogic introduction to string theory.

Volume I, *An Introduction to the Bosonic String*, provides a thorough introduction to the bosonic string, based on the Polyakov path integral and conformal field theory. The first four chapters introduce the central ideas of string theory, the tools of conformal field theory and of the Polyakov path integral, and the covariant quantization of the string. The next three chapters treat string interactions: the general formalism, and detailed treatments of the tree-level and one loop amplitudes. Chapter eight covers toroidal compactification and many important aspects of string physics, such as T-duality and D-branes. Chapter nine treats higher-order amplitudes, including an analysis of the finiteness and unitarity, and various nonperturbative ideas. An appendix giving a short course on path integral methods is also included.

Volume II, *Superstring Theory and Beyond*, begins with an introduction to supersymmetric string theories and goes on to a broad presentation of the important advances of recent years. The first three chapters introduce the type I, type II, and heterotic superstring theories and their interactions. The next two chapters present important recent discoveries about strongly coupled strings, beginning with a detailed treatment of D-branes and their dynamics, and covering string duality, M-theory, and black hole entropy. A following chapter collects many classic results in conformal field theory. The final four chapters are concerned with four-dimensional string theories, and have two goals: to show how some of the simplest string models connect with previous ideas for unifying the Standard Model; and to collect many important and beautiful general results on world-sheet and spacetime symmetries. An appendix summarizes the necessary background on fermions and supersymmetry.

Both volumes contain an annotated reference section, emphasizing references that will be useful to the student, as well as a detailed glossary of important terms and concepts. Many exercises are included which are intended to reinforce the main points of the text and to bring in additional ideas.

An essential text and reference for graduate students and researchers in theoretical physics, particle physics, and relativity with an interest in modern superstring theory.

Joseph Polchinski received his Ph.D. from the University of California at Berkeley in 1980. After postdoctoral fellowships at the Stanford Linear Accelerator Center and Harvard, he joined the faculty at the University of Texas at Austin in 1984, moving to his present position of Professor of Physics at the University of California at Santa Barbara, and Permanent Member of the Institute for Theoretical Physics, in 1992.

Professor Polchinski is not only a clear and pedagogical expositor, but is also a leading string theorist. His discovery of the importance of D-branes in 1995 is one of the most important recent contributions in this field, and he has also made significant contributions to many areas of quantum field theory and to supersymmetric models of particle physics.

Contents

Foreword xiii

Preface xv

Notation xviii

10 Type I and type II superstrings 1
10.1 The superconformal algebra 1
10.2 Ramond and Neveu–Schwarz sectors 5
10.3 Vertex operators and bosonization 10
10.4 The superconformal ghosts 15
10.5 Physical states 20
10.6 Superstring theories in ten dimensions 25
10.7 Modular invariance 31
10.8 Divergences of type I theory 37
Exercises 43

11 The heterotic string 45
11.1 World-sheet supersymmetries 45
11.2 The $SO(32)$ and $E_8 \times E_8$ heterotic strings 49
11.3 Other ten-dimensional heterotic strings 55
11.4 A little Lie algebra 59
11.5 Current algebras 66
11.6 The bosonic construction and toroidal compactification 73
Exercises 82

12 Superstring interactions 84
12.1 Low energy supergravity 84
12.2 Anomalies 94
12.3 Superspace and superfields 103
12.4 Tree-level amplitudes 110
12.5 General amplitudes 118

12.6 One-loop amplitudes 126
Exercises 134

13 D-branes 136
13.1 T-duality of type II strings 136
13.2 T-duality of type I strings 138
13.3 The D-brane charge and action 146
13.4 D-brane interactions: statics 152
13.5 D-brane interactions: dynamics 158
13.6 D-brane interactions: bound states 164
Exercises 175

14 Strings at strong coupling 178
14.1 Type IIB string and $SL(2,Z)$ duality 179
14.2 U-duality 187
14.3 $SO(32)$ type I–heterotic duality 190
14.4 Type IIA string and M-theory 198
14.5 The $E_8 \times E_8$ heterotic string 205
14.6 What is string theory? 208
14.7 Is M for matrix? 211
14.8 Black hole quantum mechanics 219
Exercises 226

15 Advanced CFT 228
15.1 Representations of the Virasoro algebra 228
15.2 The conformal bootstrap 233
15.3 Minimal models 236
15.4 Current algebras 243
15.5 Coset models 250
15.6 Representations of the $N = 1$ superconformal algebra 254
15.7 Rational CFT 255
15.8 Renormalization group flows 259
15.9 Statistical mechanics 266
Exercises 271

16 Orbifolds 274
16.1 Orbifolds of the heterotic string 275
16.2 Spacetime supersymmetry 281
16.3 Examples 283
16.4 Low energy field theory 292
Exercises 300

17 Calabi–Yau compactification 302
17.1 Conditions for $N = 1$ supersymmetry 302
17.2 Calabi–Yau manifolds 305
17.3 Massless spectrum 312
17.4 Low energy field theory 315

17.5 Higher corrections 321
17.6 Generalizations 324

18 Physics in four dimensions 327
18.1 Continuous and discrete symmetries 327
18.2 Gauge symmetries 335
18.3 Mass scales 343
18.4 More on unification 351
18.5 Conditions for spacetime supersymmetry 356
18.6 Low energy actions 359
18.7 Supersymmetry breaking in perturbation theory 362
18.8 Supersymmetry beyond perturbation theory 366
Exercises 373

19 Advanced topics 375
19.1 The $N = 2$ superconformal algebra 375
19.2 Type II strings on Calabi–Yau manifolds 379
19.3 Heterotic string theories with (2,2) SCFT 386
19.4 $N = 2$ minimal models 390
19.5 Gepner models 394
19.6 Mirror symmetry and applications 402
19.7 The conifold 409
19.8 String theories on K3 415
19.9 String duality below ten dimensions 421
19.10 Conclusion 429
Exercises 429

Appendix B: Spinors and SUSY in various dimensions 430
B.1 Spinors in various dimensions 430
B.2 Introduction to supersymmetry: $d = 4$ 439
B.3 Supersymmetry in $d = 2$ 449
B.4 Differential forms and generalized gauge fields 450
B.5 Thirty-two supersymmetries 452
B.6 Sixteen supersymmetries 457
B.7 Eight supersymmetries 461
Exercises 466

References 467

Glossary 488

Index 518

1

A first look at strings

1.1 Why strings?

One of the main themes in the history of science has been unification. Time and again diverse phenomena have been understood in terms of a small number of underlying principles and building blocks. The principle that underlies our current understanding of nature is quantum field theory, quantum mechanics with the basic observables living at spacetime points. In the late 1940s it was shown that quantum field theory is the correct framework for the unification of quantum mechanics and electromagnetism. By the early 1970s it was understood that the weak and strong nuclear forces are also described by quantum field theory. The full theory, the $SU(3) \times SU(2) \times U(1)$ *Model* or *Standard Model,* has been confirmed repeatedly in the ensuing years. Combined with general relativity, this theory is consistent with virtually all physics down to the scales probed by particle accelerators, roughly 10^{-16} cm. It also passes a variety of indirect tests that probe to shorter distances, including precision tests of quantum electrodynamics, searches for rare meson decays, limits on neutrino masses, limits on axions (light weakly interacting particles) from astrophysics, searches for proton decay, and gravitational limits on the couplings of massless scalars. In each of these indirect tests new physics might well have appeared, but in no case has clear evidence for it yet been seen; at the time of writing, the strongest sign is the solar neutrino problem, suggesting nonzero neutrino masses.

The Standard Model (plus gravity) has a fairly simple structure. There are four interactions based on local invariance principles. One of these, gravitation, is mediated by the spin-2 graviton, while the other three are mediated by the spin-1 $SU(3) \times SU(2) \times U(1)$ gauge bosons. In addition, the theory includes the spin-0 Higgs boson needed for symmetry breaking, and the quarks and leptons, fifteen multiplets of spin-$\frac{1}{2}$ fermions in three

generations of five. The dynamics is governed by a Lagrangian that depends upon roughly twenty free parameters such as the gauge and Yukawa couplings.

In spite of its impressive successes, this theory is surely not complete. First, it is too arbitrary: why does this particular pattern of gauge fields and multiplets exist, and what determines the parameters in the Lagrangian? Second, the union of gravity with quantum theory yields a nonrenormalizable quantum field theory, a strong signal that new physics should appear at very high energy. Third, even at the classical level the theory breaks down at the singularities of general relativity. Fourth, the theory is in a certain sense unnatural: some of the parameters in the Lagrangian are much smaller than one would expect them to be. It is these problems, rather than any positive experimental evidence, that presently must guide us in our attempts to find a more complete theory. One seeks a principle that unifies the fields of the Standard Model in a simpler structure, and resolves the divergence and naturalness problems.

Several promising ideas have been put forward. One is grand unification. This combines the three gauge interactions into one and the five multiplets of each generation into two or even one. It also successfully predicts one of the free parameters (the weak mixing angle) and possibly another (the bottom-tau mass ratio). A second idea is that spacetime has more than four dimensions, with the additional ones so highly curved as to be undetectable at current energies. This is certainly a logical possibility, since spacetime geometry is dynamical in general relativity. What makes it attractive is that a single higher-dimensional field can give rise to many four-dimensional fields, differing in their polarization (which can point along the small dimensions or the large) and in their dependence on the small dimensions. This opens the possibility of unifying the gauge interactions and gravity (the Kaluza–Klein mechanism). It also gives a natural mechanism for producing generations, repeated copies of the same fermion multiplets. A third unifying principle is supersymmetry, which relates fields of different spins and statistics, and which helps with the divergence and naturalness problems.

Each of these ideas — grand unification, extra dimensions, and supersymmetry — has attractive features and is consistent with the various tests of the Standard Model. It is plausible that these will be found as elements of a more complete theory of fundamental physics. It is clear, however, that something is still missing. Applying these ideas, either singly or together, has not led to theories that are substantially simpler or less arbitrary than the Standard Model.

Short-distance divergences have been an important issue many times in quantum field theory. For example, they were a key clue leading from the Fermi theory of the weak interaction to the Weinberg–Salam theory. Let

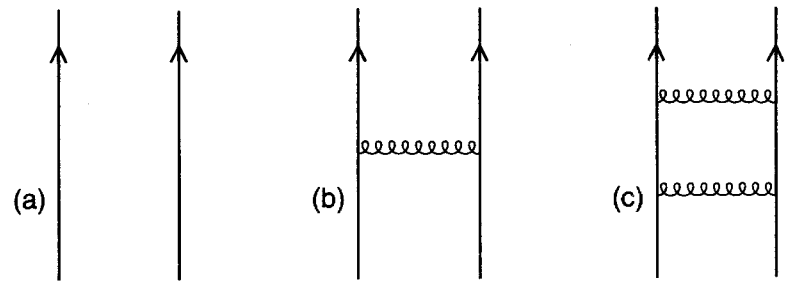


Fig. 1.1. (a) Two particles propagating freely. (b) Correction from one-graviton exchange. (c) Correction from two-graviton exchange.

us look at the short-distance problem of quantum gravity, which can be understood from a little dimensional analysis. Figure 1.1 shows a process, two particles propagating, and corrections due to one-graviton exchange and two-graviton exchange. The one-graviton exchange is proportional to Newton's constant G_N. The ratio of the one-graviton correction to the original amplitude must be governed by the dimensionless combination $G_N E^2 \hbar^{-1} c^{-5}$, where E is the characteristic energy of the process; this is the only dimensionless combination that can be formed from the parameters in the problem. Throughout this book we will use units in which $\hbar = c = 1$, defining the Planck mass

$$M_P = G_N^{-1/2} = 1.22 \times 10^{19}\,\text{GeV} \tag{1.1.1}$$

and the Planck length

$$M_P^{-1} = 1.6 \times 10^{-33}\,\text{cm}\,. \tag{1.1.2}$$

The ratio of the one-graviton to the zero-graviton amplitude is then of order $(E/M_P)^2$.

From this dimensional analysis one learns immediately that the quantum gravitational correction is an *irrelevant* interaction, meaning that it grows weaker at low energy, and in particular is negligible at particle physics energies of hundreds of GeV. By the same token, the coupling grows stronger at high energy and at $E > M_P$ perturbation theory breaks down. In the two-graviton correction of figure 1.1(c) there is a sum over intermediate states. For intermediate states of high energy E', the ratio of the two-graviton to the zero-graviton amplitude is on dimensional grounds of order

$$G_N^2 \int dE'\, E'^3 = \frac{1}{M_P^4} \int dE'\, E'^3\,, \tag{1.1.3}$$

which diverges if the theory is extrapolated to arbitrarily high energies. In position space this divergence comes from the limit where all the graviton vertices become coincident. The divergence grows worse with each additional graviton — this is the problem of nonrenormalizability.

There are two possible resolutions. The first is that the divergence is due to expanding in powers of the interaction and disappears when the theory is treated exactly. In the language of the renormalization group, this would be a nontrivial UV fixed point. The second is that the extrapolation of the theory to arbitrarily high energies is incorrect, and beyond some energy the theory is modified in a way that smears out the interaction in spacetime and softens the divergence. It is not known whether quantum gravity has a nontrivial UV fixed point, but there are a number of reasons for concentrating on the second possibility. One is history — the same kind of divergence problem in the Fermi theory of the weak interaction was a sign of new physics, the contact interaction between the fermions resolving at shorter distance into the exchange of a gauge boson. Another is that we need a more complete theory in any case to account for the patterns in the Standard Model, and it is reasonable to hope that the same new physics will solve the divergence problem of quantum gravity.

In quantum field theory it is not easy to smear out interactions in a way that preserves the consistency of the theory. We know that Lorentz invariance holds to very good approximation, and this means that if we spread the interaction in space we spread it in time as well, with consequent loss of causality or unitarity. Moreover we know that Lorentz invariance is actually embedded in a local symmetry, general coordinate invariance, and this makes it even harder to spread the interaction out without producing inconsistencies.

In fact, there is presently only one way known to spread out the gravitational interaction and cut off the divergence without spoiling the consistency of the theory. This is string theory, illustrated in figure 1.2. In this theory the graviton and all other elementary particles are one-dimensional objects, strings, rather than points as in quantum field theory. Why this should work and not anything else is not at all obvious *a priori*, but as we develop the theory we will see how it comes about.[1] Perhaps we merely suffer from a lack of imagination, and there are many other consistent theories of gravity with a short-distance cutoff. However, experience has shown that divergence problems in quantum field theory

[1] There is an intuitive answer to at least one common question: why not membranes, two- or higher-dimensional objects? The answer is that as we spread out particles in more dimensions we reduce the spacetime divergences, but encounter new divergences coming from the increased number of *internal* degrees of freedom. One dimension appears to be the unique case where both the spacetime and internal divergences are under control. However, as we will discuss in chapter 14, the membrane idea has resurfaced in somewhat transmuted form as *matrix theory*.

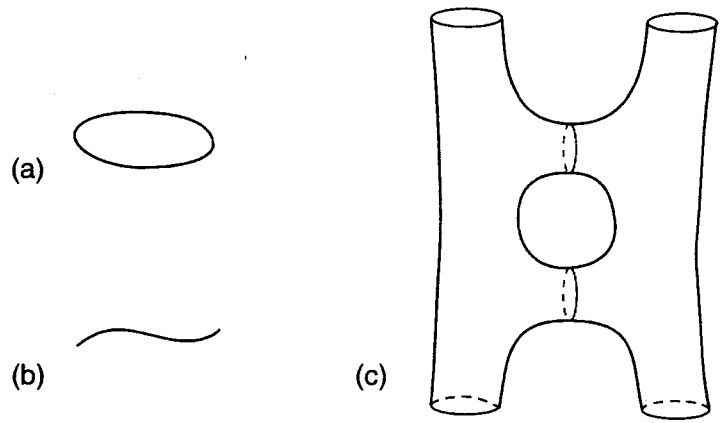


Fig. 1.2. (a) Closed string. (b) Open string. (c) The loop amplitude of fig. 1.1(c) in string theory. Each particle world-line becomes a cylinder, and the interactions no longer occur at points. (The cross-sections on the intermediate lines are included only for perspective.)

are not easily resolved, so if we have even one solution we should take it very seriously. Indeed, we are fortunate that consistency turns out to be such a restrictive principle, since the unification of gravity with the other interactions takes place at such high energy, M_P, that experimental tests will be difficult and indirect.

So what do we find if we pursue this idea? In a word, the result is remarkable. String theory dovetails beautifully with the previous ideas for explaining the patterns in the Standard Model, and does so with a structure more elegant and unified than in quantum field theory. In particular, if one tries to construct a consistent relativistic quantum theory of one-dimensional objects one finds:

1. *Gravity.* Every consistent string theory must contain a massless spin-2 state, whose interactions reduce at low energy to general relativity.

2. *A consistent theory of quantum gravity, at least in perturbation theory.* As we have noted, this is in contrast to all known quantum *field* theories of gravity.

3. *Grand unification.* String theories lead to gauge groups large enough to include the Standard Model. Some of the simplest string theories lead to the same gauge groups and fermion representations that arise in the unification of the Standard Model.

4. *Extra dimensions.* String theory requires a definite number of spacetime dimensions, ten.[2] The field equations have solutions with four large flat and six small curved dimensions, with four-dimensional physics that resembles the Standard Model.

5. *Supersymmetry.* Consistent string theories require spacetime supersymmetry, as either a manifest or a spontaneously broken symmetry.

6. *Chiral gauge couplings.* The gauge interactions in nature are parity asymmetric (chiral). This has been a stumbling block for a number of previous unifying ideas: they required parity symmetric gauge couplings. String theory allows chiral gauge couplings.

7. *No free parameters.* String theory has no adjustable constants.

8. *Uniqueness.* Not only are there no continuous parameters, but there is no discrete freedom analogous to the choice of gauge group and representations in field theory: there is a unique string theory.

In addition one finds a number of other features, such as an axion, and hidden gauge groups, that have played a role in ideas for unification.

This is a remarkable list, springing from the simple supposition of one-dimensional objects. The first two points alone would be of great interest. The next four points come strikingly close to the picture one arrives at in trying to unify the Standard Model. And as indicated by the last two points, string theory accomplishes this with a structure that is tighter and less arbitrary than in quantum field theory, supplying the element missing in the previous ideas. The first point is a further example of this tightness: string theory *must* have a graviton, whereas in field theory this and other fields are combined in a mix-and-match fashion.

String theory further has connections to many areas of mathematics, and has led to the discovery of new and unexpected relations among them. It has rich connections to the recent discoveries in supersymmetric quantum field theory. String theory has also begun to address some of the deeper questions of quantum gravity, in particular the quantum mechanics of black holes.

Of course, much remains to be done. String theory may resemble the real world in its broad outlines, but a decisive test still seems to be far away. The main problem is that while there is a unique *theory,* it has an enormous number of classical solutions, even if we restrict attention

[2] To be precise, string theory modifies the notions of spacetime topology and geometry, so what we mean by a dimension here is generalized. Also, we will see that ten dimensions is the appropriate number for weakly coupled string theory, but that the picture can change at strong coupling.

to solutions with four large flat dimensions. Upon quantization, each of these is a possible ground state (vacuum) for the theory, and the four-dimensional physics is different in each. It is known that quantum effects greatly reduce the number of stable solutions, but a full understanding of the dynamics is not yet in hand.

It is worth recalling that even in the Standard Model, the dynamics of the vacuum plays an important role in the physics we see. In the electroweak interaction, the fact that the vacuum is less symmetric than the Hamiltonian (spontaneous symmetry breaking) plays a central role. In the strong interaction, large fluctuating gauge fields in the vacuum are responsible for quark confinement. These phenomena in quantum field theory arise from having a quantum system with many degrees of freedom. In string theory there are seemingly many more degrees of freedom, and so we should expect even richer dynamics.

Beyond this, there is the question, 'what is string theory?' Until recently our understanding of string theory was limited to perturbation theory, small numbers of strings interacting weakly. It was not known how even to define the theory at strong coupling. There has been a suspicion that the degrees of freedom that we use at weak coupling, one-dimensional objects, are not ultimately the simplest or most complete way to understand the theory.

In the past few years there has been a great deal of progress on these issues, growing largely out of the systematic application of the constraints imposed by supersymmetry. We certainly do not have a complete understanding of the dynamics of strongly coupled strings, but it has become possible to map out in detail the space of vacua (when there is enough unbroken supersymmetry) and this has led to many surprises. One is the absolute uniqueness of the theory: whereas there are several weakly coupled string theories, all turn out to be limits in the space of vacua of a single theory. Another is a limit in which spacetime becomes eleven-dimensional, an interesting number from the point of view of supergravity but impossible in weakly coupled string theory. It has also been understood that the theory contains new extended objects, D-branes, and this has led to the new understanding of black hole quantum mechanics. All this also gives new and unexpected clues as to the ultimate nature of the theory.

In summary, we are fortunate that so many approaches seem to converge on a single compelling idea. Whether one starts with the divergence problem of quantum gravity, with attempts to account for the patterns in the Standard Model, or with a search for new symmetries or mathematical structures that may be useful in constructing a unified theory, one is led to string theory.

Outline

The goal of these two volumes is to provide a complete introduction to string theory, starting at the beginning and proceeding through the compactification to four dimensions and to the latest developments in strongly coupled strings.

Volume one is an introduction to bosonic string theory. This is not a realistic theory — it does not have fermions, and as far as is known has no stable ground state. The philosophy here is the same as in starting a course on quantum field theory with a thorough study of scalar field theory. That is also not the theory one is ultimately interested in, but it provides a simple example for developing the unique dynamical and technical features of quantum field theory before introducing the complications of spin and gauge invariance. Similarly, a thorough study of bosonic string theory will give us a framework to which we can in short order add the additional complications of fermions and supersymmetry.

The rest of chapter 1 is introductory. We present first the action principle for the dynamics of string. We then carry out a quick and heuristic quantization using light-cone gauge, to show the reader some of the important aspects of the string spectrum. Chapters 2–7 are the basic introduction to bosonic string theory. Chapter 2 introduces the needed technical tools in the world-sheet quantum field theory, such as conformal invariance, the operator product expansion, and vertex operators. Chapters 3 and 4 carry out the covariant quantization of the string, starting from the Polyakov path integral. Chapters 5–7 treat interactions, presenting the general formalism and applying it to tree-level and one-loop amplitudes. Chapter 8 treats the simplest compactification of string theory, making some of the dimensions periodic. In addition to the phenomena that arise in compactified field theory, such as Kaluza–Klein gauge symmetry, there is also a great deal of 'stringy' physics, including enhanced gauge symmetries, T-duality, and D-branes. Chapter 9 treats higher order amplitudes. The first half outlines the argument that string theory in perturbation theory is finite and unitary as advertised; the second half presents brief treatments of a number of advanced topics, such as string field theory. Appendix A is an introduction to path integration, so that our use of quantum field theory is self-contained.

Volume two treats supersymmetric string theories, focusing first on ten-dimensional and other highly symmetric vacua, and then on realistic four-dimensional vacua.

In chapters 10–12 we extend the earlier introduction to the supersymmetric string theories, developing the type I, II, and heterotic superstrings and their interactions. We then introduce the latest results in these subjects. Chapter 13 develops the properties and dynamics of D-branes, still using the tools of string perturbation theory as developed earlier in the book. Chapter 14 then uses arguments based on supersymmetry to understand strongly coupled strings. We find that the strongly coupled limit of any string theory is described by a dual weakly coupled string theory, or by a new eleven-dimensional theory known provisionally as M-theory. We discuss the status of the search for a complete formulation of string theory and present one promising idea, m(atrix) theory. We briefly discuss the quantum mechanics of black holes, carrying out the simplest entropy calculation. Chapter 15 collects a number of advanced applications of the various world-sheet symmetry algebras.

Chapters 16 and 17 present four-dimensional string theories based on orbifold and Calabi–Yau compactifications. The goal is not an exhaustive treatment but rather to make contact between the simplest examples and the unification of the Standard Model. Chapter 18 collects results that hold in wide classes of string theories, using general arguments based on world-sheet and spacetime gauge symmetries. Chapter 19 consists of advanced topics, including (2,2) world-sheet supersymmetry, mirror symmetry, the conifold transition, and the strong-coupling behavior of some compactified theories.

Annotated reference lists appear at the end of each volume. I have tried to assemble a selection of articles, particularly reviews, that may be useful to the student. A glossary also appears at the end of each volume.

1.2 Action principles

We want to study the classical and quantum mechanics of a one-dimensional object, a string. The string moves in D flat spacetime dimensions, with metric $\eta_{\mu\nu} = \mathrm{diag}(-,+,+,\cdots,+)$.

It is useful to review first the classical mechanics of a zero-dimensional object, a relativistic point particle. We can describe the motion of a particle by giving its position in terms of $D-1$ functions of time, $\mathbf{X}(X^0)$. This hides the covariance of the theory though, so it is better to introduce a parameter τ along the particle's world-line and describe the motion in spacetime by D functions $X^\mu(\tau)$. The parameterization is arbitrary: a different parameterization of the same path is physically equivalent, and all physical quantities must be independent of this choice. That is, for any monotonic function $\tau'(\tau)$, the two paths X'^μ and X^μ are the same, where

$$X'^\mu(\tau'(\tau)) = X^\mu(\tau) \,. \tag{1.2.1}$$

We are trading a less symmetric description for a more symmetric but redundant one, which is often a useful step. Figure 1.3(a) shows a parameterized world-line.

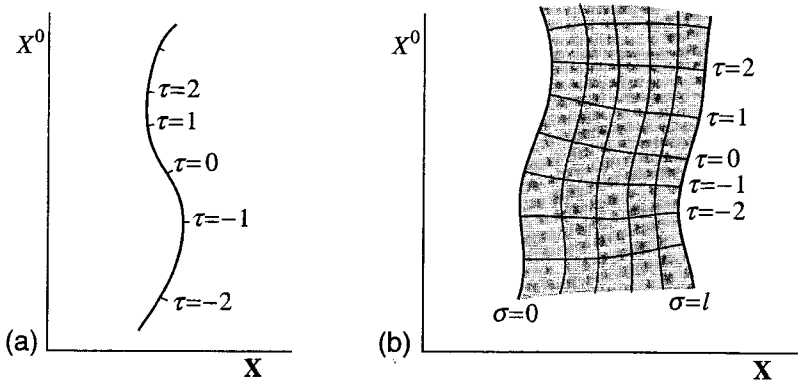


Fig. 1.3. (a) Parameterized world-line of a point particle. (b) Parameterized world-sheet of an open string.

The simplest Poincaré-invariant action that does not depend on the parameterization would be proportional to the proper time along the world-line,

$$S_{\rm pp} = -m \int d\tau\, (-\dot{X}^\mu \dot{X}_\mu)^{1/2} , \tag{1.2.2}$$

where a dot denotes a τ-derivative. The variation of the action, after an integration by parts, is

$$\delta S_{\rm pp} = -m \int d\tau\, \dot{u}_\mu \delta X^\mu , \tag{1.2.3}$$

where

$$u^\mu = \dot{X}^\mu (-\dot{X}^\nu \dot{X}_\nu)^{-1/2} \tag{1.2.4}$$

is the normalized D-velocity. The equation of motion $\dot{u}^\mu = 0$ thus describes free motion. The normalization constant m is the particle's mass, as can be checked by looking at the nonrelativistic limit (exercise 1.1).

The action can be put in another useful form by introducing an additional field on the world-line, an independent world-line metric $\gamma_{\tau\tau}(\tau)$. It will be convenient to work with the tetrad $\eta(\tau) = (-\gamma_{\tau\tau}(\tau))^{1/2}$, which is defined to be positive. We use the general relativity term *tetrad* in any number of dimensions, even though its root means 'four.' Then

$$S'_{\rm pp} = \frac{1}{2} \int d\tau \left(\eta^{-1} \dot{X}^\mu \dot{X}_\mu - \eta m^2 \right) . \tag{1.2.5}$$

This action has the same symmetries as the earlier action $S_{\rm pp}$, namely Poincaré invariance and world-line reparameterization invariance. Under the latter, $\eta(\tau)$ transforms as

$$\eta'(\tau') d\tau' = \eta(\tau) d\tau . \tag{1.2.6}$$

The equation of motion from varying the tetrad is

$$\eta^2 = -\dot{X}^\mu \dot{X}_\mu / m^2 . \tag{1.2.7}$$

Using this to eliminate $\eta(\tau)$, the action $S'_{\rm pp}$ becomes the earlier $S_{\rm pp}$. Incidentally, $S'_{\rm pp}$ makes perfect sense for massless particles, while $S_{\rm pp}$ does not work in that case.

Although the actions $S_{\rm pp}$ and $S'_{\rm pp}$ are equivalent classically, it is hard to make sense of $S_{\rm pp}$ in a path integral because of its complicated form, with derivatives inside the square root. On the other hand $S'_{\rm pp}$ is quadratic in derivatives and its path integral is fairly easily evaluated, as we will see in later chapters. Presumably, any attempt to define the quantum theory for $S_{\rm pp}$ will lead to a result equivalent to the $S'_{\rm pp}$ path integral, and we will take the latter as our starting point in defining the quantum theory.

A one-dimensional object will sweep out a two-dimensional world-sheet, which can be described in terms of two parameters, $X^\mu(\tau, \sigma)$ as in figure 1.3(b). As in the case of the particle, we insist that physical quantities such as the action depend only on the embedding in spacetime and not on the parameterization. Not only is it attractive that the theory should have this property, but we will see that it is necessary in a consistent relativistic quantum theory. The simplest invariant action, the Nambu–Goto action, is proportional to the area of the world-sheet. In order to express this action in terms of $X^\mu(\tau, \sigma)$, define first the induced metric h_{ab} where indices $a, b, \ldots$ run over values (τ, σ):

$$h_{ab} = \partial_a X^\mu \partial_b X_\mu . \tag{1.2.8}$$

Then the Nambu–Goto action is

$$S_{\rm NG} = \int_M d\tau d\sigma\, \mathscr{L}_{\rm NG} , \tag{1.2.9a}$$

$$\mathscr{L}_{\rm NG} = -\frac{1}{2\pi\alpha'} (-\det h_{ab})^{1/2} , \tag{1.2.9b}$$

where M denotes the world-sheet. The constant α', which has units of spacetime-length-squared, is the Regge slope, whose significance will be seen later. The tension T of the string is related to the Regge slope by

$$T = \frac{1}{2\pi\alpha'} \tag{1.2.10}$$

(exercise 1.1).

Let us now consider the symmetries of the action, transformations of $X^\mu(\tau, \sigma)$ such that $S_{\rm NG}[X'] = S_{\rm NG}[X]$. These are:

1. The isometry group of flat spacetime, the D-dimensional Poincaré group,
$$X'^{\mu}(\tau,\sigma) = \Lambda^{\mu}_{\nu} X^{\nu}(\tau,\sigma) + a^{\mu} \tag{1.2.11}$$
with Λ^{μ}_{ν} a Lorentz transformation and a^{μ} a translation.

2. Two-dimensional coordinate invariance, often called diffeomorphism *(diff)* invariance. For new coordinates $(\tau'(\tau,\sigma), \sigma'(\tau,\sigma))$, the transformation is
$$X'^{\mu}(\tau',\sigma') = X^{\mu}(\tau,\sigma) \,. \tag{1.2.12}$$

The Nambu–Goto action is analogous to the point particle action $S_{\rm pp}$, with derivatives in the square root. Again we can simplify it by introducing an independent world-sheet metric $\gamma_{ab}(\tau,\sigma)$. Henceforth 'metric' will always mean γ_{ab} (unless we specify 'induced'), and indices will always be raised and lowered with this metric. We will take γ_{ab} to have Lorentzian signature $(-,+)$. The action is
$$S_{\rm P}[X,\gamma] = -\frac{1}{4\pi\alpha'}\int_M d\tau\, d\sigma\, (-\gamma)^{1/2}\, \gamma^{ab}\partial_a X^{\mu}\partial_b X_{\mu} \,, \tag{1.2.13}$$
where $\gamma = \det \gamma_{ab}$. This is the Brink–Di Vecchia–Howe–Deser–Zumino action, or Polyakov action for short. It was found by Brink, Di Vecchia, and Howe and by Deser and Zumino, in the course of deriving a generalization with local world-sheet supersymmetry. Its virtues, especially for path integral quantization, were emphasized by Polyakov.

To see the equivalence to $S_{\rm NG}$, use the equation of motion obtained by varying the metric,
$$\delta_{\gamma} S_{\rm P}[X,\gamma] = -\frac{1}{4\pi\alpha'}\int_M d\tau\, d\sigma\, (-\gamma)^{1/2}\, \delta\gamma^{ab}\left(h_{ab} - \tfrac{1}{2}\gamma_{ab}\gamma^{cd}h_{cd}\right) \,, \tag{1.2.14}$$
where h_{ab} is again the induced metric (1.2.8). We have used the general relation for the variation of a determinant,
$$\delta\gamma = \gamma\gamma^{ab}\delta\gamma_{ab} = -\gamma\gamma_{ab}\delta\gamma^{ab} \,. \tag{1.2.15}$$
Then $\delta_{\gamma} S_{\rm P} = 0$ implies
$$h_{ab} = \tfrac{1}{2}\gamma_{ab}\gamma^{cd}h_{cd} \,. \tag{1.2.16}$$
Dividing this equation by the square root of minus its determinant gives
$$h_{ab}(-h)^{-1/2} = \gamma_{ab}(-\gamma)^{-1/2} \,, \tag{1.2.17}$$
so that γ_{ab} is proportional to the induced metric. This in turn can be used to eliminate γ_{ab} from the action,
$$S_{\rm P}[X,\gamma] \to -\frac{1}{2\pi\alpha'}\int d\tau\, d\sigma\, (-h)^{1/2} = S_{\rm NG}[X] \,. \tag{1.2.18}$$

The action $S_{\rm P}$ has the following symmetries:

1. D-dimensional Poincaré invariance:
$$\begin{aligned} X'^{\mu}(\tau,\sigma) &= \Lambda^{\mu}_{\nu} X^{\nu}(\tau,\sigma) + a^{\mu} \,, \\ \gamma'_{ab}(\tau,\sigma) &= \gamma_{ab}(\tau,\sigma) \,. \end{aligned} \tag{1.2.19}$$

2. Diff invariance:
$$\begin{aligned} X'^{\mu}(\tau',\sigma') &= X^{\mu}(\tau,\sigma) \,, \\ \frac{\partial\sigma'^{c}}{\partial\sigma^{a}}\frac{\partial\sigma'^{d}}{\partial\sigma^{b}}\gamma'_{cd}(\tau',\sigma') &= \gamma_{ab}(\tau,\sigma) \,, \end{aligned} \tag{1.2.20}$$
for new coordinates $\sigma'^{a}(\tau,\sigma)$.

3. Two-dimensional Weyl invariance:
$$\begin{aligned} X'^{\mu}(\tau,\sigma) &= X^{\mu}(\tau,\sigma) \,, \\ \gamma'_{ab}(\tau,\sigma) &= \exp(2\omega(\tau,\sigma))\,\gamma_{ab}(\tau,\sigma) \,, \end{aligned} \tag{1.2.21}$$
for arbitrary $\omega(\tau,\sigma)$.

The Weyl invariance, a local rescaling of the world-sheet metric, has no analog in the Nambu–Goto form. We can understand its appearance by looking back at the equation of motion (1.2.17) used to relate the Polyakov and Nambu–Goto actions. This does not determine γ_{ab} completely but only up to a local rescaling, so Weyl-equivalent metrics correspond to the *same* embedding in spacetime. This is an extra redundancy of the Polyakov formulation, in addition to the obvious diff invariance.

The variation of the action with respect to the metric defines the energy-momentum tensor,
$$\begin{aligned} T^{ab}(\tau,\sigma) &= -4\pi(-\gamma)^{-1/2}\frac{\delta}{\delta\gamma_{ab}}S_{\rm P} \\ &= -\frac{1}{\alpha'}(\partial^{a}X^{\mu}\partial^{b}X_{\mu} - \tfrac{1}{2}\gamma^{ab}\partial_{c}X^{\mu}\partial^{c}X_{\mu}) \,. \end{aligned} \tag{1.2.22}$$
This has an extra factor of -2π relative to the usual definition in field theory; this is a standard and convenient convention in string theory. It is conserved, $\nabla_a T^{ab} = 0$, as a consequence of diff invariance. The invariance of $S_{\rm P}$ under arbitrary Weyl transformations further implies that
$$\gamma_{ab}\frac{\delta}{\delta\gamma_{ab}}S_{\rm P} = 0 \;\Rightarrow\; T^{a}_{a} = 0 \,. \tag{1.2.23}$$

The actions $S_{\rm NG}$ and $S_{\rm P}$ define two-dimensional field theories on the string world-sheet. In string theory, we will see that amplitudes for *spacetime* processes are given by matrix elements in the two-dimensional quantum field theory on the *world-sheet*. While our interest is four-dimensional

spacetime, most of the machinery we use in string perturbation theory is two-dimensional. From the point of view of the world-sheet, the coordinate transformation law (1.2.12) defines $X^\mu(\tau,\sigma)$ as a scalar field, with μ going along for the ride as an internal index. From the two-dimensional point of view the Polyakov action describes massless Klein–Gordon scalars X^μ covariantly coupled to the metric γ_{ab}. Also from this point of view, the Poincaré invariance is an internal symmetry, meaning that it acts on fields at fixed τ,σ.

Varying γ_{ab} in the action gives the equation of motion

$$T_{ab} = 0 \,. \tag{1.2.24}$$

Varying X^μ gives the equation of motion

$$\partial_a(-\gamma)^{1/2}\gamma^{ab}\partial_b X^\mu = (-\gamma)^{1/2}\nabla^2 X^\mu = 0 \,. \tag{1.2.25}$$

For world-sheets with boundary there is also a surface term in the variation of the action. To be specific, take the coordinate region to be

$$-\infty \le \tau \le \infty \,, \quad 0 \le \sigma \le \ell \,. \tag{1.2.26}$$

We will think of τ as a time variable and σ as spatial, so this is a single string propagating without sources. Then

$$\delta S_{\rm P} = \frac{1}{2\pi\alpha'}\int_{-\infty}^{\infty} d\tau \int_0^\ell d\sigma\,(-\gamma)^{1/2}\delta X^\mu \nabla^2 X^\mu - \frac{1}{2\pi\alpha'}\int_{-\infty}^{\infty} d\tau\,(-\gamma)^{1/2}\delta X^\mu \partial^\sigma X^\mu\Big|_{\sigma=0}^{\sigma=\ell} \,. \tag{1.2.27}$$

In the interior, the equation of motion is as above. The boundary term vanishes if

$$\partial^\sigma X^\mu(\tau,0) = \partial^\sigma X^\mu(\tau,\ell) = 0 \,. \tag{1.2.28}$$

These are *Neumann* boundary conditions on X^μ. Stated more covariantly,

$$n^a\partial_a X_\mu = 0 \quad \text{on } \partial M \,, \tag{1.2.29}$$

where n^a is normal to the boundary ∂M. The ends of the open string move freely in spacetime.

The surface term in the equation of motion will also vanish if we impose

$$X^\mu(\tau,\ell) = X^\mu(\tau,0) \,, \quad \partial^\sigma X^\mu(\tau,\ell) = \partial^\sigma X^\mu(\tau,0) \,, \tag{1.2.30a}$$

$$\gamma_{ab}(\tau,\ell) = \gamma_{ab}(\tau,0) \,. \tag{1.2.30b}$$

That is, the fields are *periodic*. There is no boundary; the endpoints are joined to form a closed loop.

The open string boundary condition (1.2.28) and closed string boundary condition (1.2.30) are the only possibilities consistent with D-dimensional

Poincaré invariance and the equations of motion. If we relax the condition of Poincaré invariance then there are other possibilities which will be important later. Some of these are explored in the exercises.

The Nambu–Goto and Polyakov actions may be the simplest with the given symmetries, but simplicity of the action is not the right criterion in quantum field theory. Symmetry, not simplicity, is the key idea: for both physical and technical reasons we must usually consider the most general local action consistent with all of the symmetries of the theory. Let us try to generalize the Polyakov action, requiring that the symmetries be maintained and that the action be polynomial in derivatives (else it is hard to make sense of the quantum theory).[3] Global Weyl invariance, $\omega(\tau,\sigma)$ = constant, requires that the action have one more factor of γ^{ab} than γ_{ab} to cancel the variation of $(-\gamma)^{1/2}$. The extra upper indices can only contract with derivatives, so each term has precisely two derivatives. The coordinate invariance and Poincaré invariance then allow one additional term beyond the original $S_{\rm P}$:

$$\chi = \frac{1}{4\pi}\int_M d\tau\,d\sigma\,(-\gamma)^{1/2}R \,, \tag{1.2.31}$$

where R is the Ricci scalar constructed from γ_{ab}. Under a local Weyl rescaling,

$$(-\gamma')^{1/2}R' = (-\gamma)^{1/2}(R - 2\nabla^2\omega) \,. \tag{1.2.32}$$

The variation is a total derivative, because $(-\gamma)^{1/2}\nabla_a v^a = \partial_a((-\gamma)^{1/2}v^a)$ for any v^a. The integral (1.2.31) is therefore invariant for a world-sheet without boundary. With boundaries an additional surface term is needed (exercise 1.3).

Since χ is allowed by the symmetries we will include it in the action:

$$S'_{\rm P} = S_{\rm P} - \lambda\chi \,. = -\int_M d\tau\,d\sigma\,(-\gamma)^{1/2}\left(\frac{1}{4\pi\alpha'}\gamma^{ab}\partial_a X^\mu\partial_b X_\mu + \frac{\lambda}{4\pi}R\right) \,. \tag{1.2.33}$$

This is the most general (diff×Weyl)-invariant and Poincaré-invariant action with these fields and symmetries. For the moment the discussion of the symmetries is classical — we are ignoring possible quantum anomalies, which will be considered in chapter 3.

The action $S'_{\rm P}$ looks like the Hilbert action for the metric, $\int(-\gamma)^{1/2}R$, minimally coupled to D massless scalar fields X^μ. However, in two dimen-

[3] If we relax this requirement then there are various additional possibilities, such as invariants constructed from the curvature of the induced metric h_{ab}. These would be expected to appear for such one-dimensional objects as electric and magnetic flux tubes, and in that case the path integral makes sense because the thickness of the object provides a cutoff. But for infinitely thin strings, which is what we have in fundamental string theory, this is not possible.

sions, the Hilbert action depends only on the topology of the world-sheet and does not give dynamics to the metric. To see this, we use the result from general relativity that its variation is proportional to $R_{ab} - \frac{1}{2}\gamma_{ab}R$. In two dimensions, the symmetries of the curvature tensor imply that $R_{ab} = \frac{1}{2}\gamma_{ab}R$ and so this vanishes: the Hilbert action is invariant under any continuous change in the metric. It does have some global significance, as we will see in the chapter 3.

There are various further ways one might try to generalize the theory. One, suggested by our eventual interest in $D = 4$, is to allow more than four X^μ fields but require only four-dimensional Poincaré invariance. A second, motivated by the idea that symmetry is paramount, is to keep the same local symmetries plus four-dimensional Poincaré invariance, but to allow a more general field content — that is, two-dimensional fields in different representations of the world-sheet and spacetime symmetry groups. These ideas are both important and we will return to them at various times. A further idea is to enlarge the local symmetry of the theory; we will pursue this idea in volume two.

1.3 The open string spectrum

We discuss the spectrum of the open string in this section and that of the closed in the next, using light-cone gauge to eliminate the diff×Weyl redundancy. This gauge hides the covariance of the theory, but it is the quickest route to the spectrum and reveals important features like the critical dimension and the existence of massless gauge particles. The emphasis here is on the results, not the method, so the reader should not dwell on fine points. A systematic treatment using the covariant conformal gauge will begin in the next chapter.

Define light-cone coordinates in spacetime:

$$x^\pm = 2^{-1/2}(x^0 \pm x^1) , \qquad x^i , \ i = 2,\ldots,D-1 . \tag{1.3.1}$$

We are using lower case for the spacetime coordinates x^μ and upper case for the associated world-sheet fields $X^\mu(\sigma,\tau)$. In these coordinates the metric is

$$a^\mu b_\mu = -a^+b^- - a^-b^+ + a^ib^i , \tag{1.3.2a}$$

$$a_- = -a^+ , \quad a_+ = -a^- , \quad a_i = a^i . \tag{1.3.2b}$$

We will set the world-sheet parameter τ at each point of the world-sheet to be equal to the spacetime coordinate x^+. One might have thought to try $\tau = x^0$ but this does not lead to the same simplifications. So x^+ will play the role of time and p^- that of energy. The longitudinal variables x^- and

p^+ are then like spatial coordinates and momenta, as are the transverse x^i and p^i.

We start by illustrating the procedure in the case of the point particle, using the action $S'_{\rm pp}$. Fix the parameterization of the world-line by

$$X^+(\tau) = \tau . \tag{1.3.3}$$

The action then becomes

$$S'_{\rm pp} = \frac{1}{2}\int d\tau \left(-2\eta^{-1}\dot{X}^- + \eta^{-1}\dot{X}^i\dot{X}^i - \eta m^2\right) . \tag{1.3.4}$$

The canonical momenta $p_\mu = \partial L/\partial\dot{X}^\mu$ are

$$p_- = -\eta^{-1} , \quad p_i = \eta^{-1}\dot{X}^i . \tag{1.3.5}$$

Also recall the metric $p^i = p_i$ and $p^+ = -p_-$. The Hamiltonian is

$$\begin{aligned} H &= p_-\dot{X}^- + p_i\dot{X}^i - L \\ &= \frac{p^ip^i + m^2}{2p^+} . \end{aligned} \tag{1.3.6}$$

Note that there is no term $p_+\dot{X}^+$ because X^+ is not a dynamical variable in the gauge-fixed theory. Also, $\dot{\eta}$ does not appear in the action and so the momentum p_η vanishes. Since $\eta = -1/p_-$ it is clear that we should not treat η as an independent canonical coordinate; in a more complete treatment we would justify this by appealing to the theory of systems with constraints.

To quantize, impose canonical commutators on the dynamical fields,

$$[p_i, X^j] = -i\delta_i^{\ j} , \quad [p_-, X^-] = -i . \tag{1.3.7}$$

The momentum eigenstates $|k_-, k^i\rangle$ form a complete set. The remaining momentum component p_+ is determined in terms of the others as follows. The gauge choice relates τ and X^+ translations, so $H = -p_+ = p^-$. The relative sign between H and p_+ arises because the former is active and the latter passive. Then eq. (1.3.6) becomes the relativistic mass-shell condition, and we have obtained the spectrum of a relativistic scalar.

Now turn to the open string, taking again the coordinate region $-\infty \le \tau \le \infty$ and $0 \le \sigma \le \ell$. Again we must make a gauge choice to fix the redundancies in the Polyakov action, and by a good choice we can also make the equations of motion simple. Set

$$\tau = x^+ , \tag{1.3.8a}$$

$$\partial_\sigma\gamma_{\sigma\sigma} = 0 , \tag{1.3.8b}$$

$$\det\gamma_{ab} = -1 . \tag{1.3.8c}$$

That is, we choose light-cone gauge for the world-sheet time coordinate and impose two conditions on the metric. This is three conditions for

three local symmetries (the choice of two world-sheet coordinates plus the Weyl scaling).

Let us see how to reach the gauge (1.3.8). First choose τ according to (1.3.8a). Now notice that $f = \gamma_{\sigma\sigma}(-\det\gamma_{ab})^{-1/2}$ transforms as

$$f'd\sigma' = fd\sigma \tag{1.3.9}$$

under reparameterizations of σ where τ is left fixed. Thus we define an invariant length $fd\sigma = dl$. Define the σ coordinate of a point to be proportional to its invariant distance $\int dl$ from the $\sigma = 0$ endpoint; the constant of proportionality is determined by requiring that the coordinate of the right-hand endpoint remain at $\sigma = \ell$. In this coordinate system $f = dl/d\sigma$ is independent of σ. It still depends on τ; we would like the parameter region $0 \le \sigma \le \ell$ to remain unchanged, so there is not enough freedom to remove the τ-dependence. Finally, make a Weyl transformation to satisfy condition (1.3.8c). Since f is Weyl-invariant, $\partial_\sigma f$ still vanishes. With condition (1.3.8c) this implies that $\partial_\sigma\gamma_{\sigma\sigma} = 0$, and so condition (1.3.8b) is satisfied as well.[4]

We can solve the gauge condition (1.3.8c) for $\gamma_{\tau\tau}(\tau,\sigma)$, and $\gamma_{\sigma\sigma}$ is independent of σ, so the independent degrees of freedom in the metric are now $\gamma_{\sigma\sigma}(\tau)$ and $\gamma_{\sigma\tau}(\tau,\sigma)$. In terms of these, the inverse metric is

$$\begin{bmatrix} \gamma^{\tau\tau} & \gamma^{\tau\sigma} \\ \gamma^{\sigma\tau} & \gamma^{\sigma\sigma} \end{bmatrix} = \begin{bmatrix} -\gamma_{\sigma\sigma}(\tau) & \gamma_{\tau\sigma}(\tau,\sigma) \\ \gamma_{\tau\sigma}(\tau,\sigma) & \gamma_{\sigma\sigma}^{-1}(\tau)(1-\gamma_{\tau\sigma}^2(\tau,\sigma)) \end{bmatrix} . \tag{1.3.10}$$

The Polyakov Lagrangian then becomes

$$\begin{aligned} L = -\frac{1}{4\pi\alpha'}\int_0^\ell d\sigma \Big[&\gamma_{\sigma\sigma}(2\partial_\tau X^- - \partial_\tau X^i\partial_\tau X^i) \\ &-2\gamma_{\sigma\tau}(\partial_\sigma Y^- - \partial_\tau X^i\partial_\sigma X^i) + \gamma_{\sigma\sigma}^{-1}(1-\gamma_{\tau\sigma}^2)\partial_\sigma X^i\partial_\sigma X^i\Big] . \end{aligned} \tag{1.3.11}$$

Here we have separated $X^-(\tau,\sigma)$ into two pieces $x^-(\tau)$ and $Y^-(\tau,\sigma)$, the first being the mean value of X^- at given τ and the second having a mean value of zero. That is,

$$x^-(\tau) = \frac{1}{\ell}\int_0^\ell d\sigma X^-(\tau,\sigma) , \tag{1.3.12a}$$

$$Y^-(\tau,\sigma) = X^-(\tau,\sigma) - x^-(\tau) . \tag{1.3.12b}$$

[4] We have shown how to assign a definite (τ,σ) value to each point on the world-sheet, but we have not shown that this is a good gauge — that different points have different coordinates. Already for the point particle this is an issue: τ is a good coordinate only if the world-line does not double back through an x^+-hyperplane. For the string we need this and also that f is positive everywhere. It is not hard to show that for allowed classical motions, in which all points travel at less than the speed of light, these conditions are met. In the quantum theory, one can show for the point particle that the doubled-back world-lines give no net contribution. Presumably this can be generalized to the string, but this is one of those fine points not to dwell on.

The field Y^- does not appear in any terms with time derivatives and so is nondynamical. It acts as a Lagrange multiplier, constraining $\partial_\sigma\gamma_{\tau\sigma}$ to vanish. Note further that in the gauge (1.3.8) the open string boundary condition (1.2.28) becomes

$$\gamma_{\tau\sigma}\partial_\tau X^\mu - \gamma_{\tau\tau}\partial_\sigma X^\mu = 0 \quad \text{at } \sigma = 0,\ell . \tag{1.3.13}$$

For $\mu = +$ this gives

$$\gamma_{\tau\sigma} = 0 \quad \text{at } \sigma = 0,\ell \tag{1.3.14}$$

and since $\partial_\sigma\gamma_{\tau\sigma} = 0$, then $\gamma_{\tau\sigma}$ vanishes everywhere. For $\mu = i$, the boundary condition is

$$\partial_\sigma X^i = 0 \quad \text{at } \sigma = 0,\ell . \tag{1.3.15}$$

After imposing the gauge conditions and taking into account the Lagrange multiplier Y^-, we can proceed by ordinary canonical methods. The system is reduced to the variables $x^-(\tau)$, $\gamma_{\sigma\sigma}(\tau)$ and the fields $X^i(\tau,\sigma)$. The Lagrangian is

$$L = -\frac{\ell}{2\pi\alpha'}\gamma_{\sigma\sigma}\partial_\tau x^- + \frac{1}{4\pi\alpha'}\int_0^\ell d\sigma\left(\gamma_{\sigma\sigma}\partial_\tau X^i\partial_\tau X^i - \gamma_{\sigma\sigma}^{-1}\partial_\sigma X^i\partial_\sigma X^i\right) . \tag{1.3.16}$$

The momentum conjugate to x^- is

$$p_- = -p^+ = \frac{\partial L}{\partial(\partial_\tau x^-)} = -\frac{\ell}{2\pi\alpha'}\gamma_{\sigma\sigma} . \tag{1.3.17}$$

As with η in the particle example, $\gamma_{\sigma\sigma}$ is a momentum and not a coordinate. The momentum density conjugate to $X^i(\tau,\sigma)$ is

$$\Pi^i = \frac{\delta L}{\delta(\partial_\tau X^i)} = \frac{1}{2\pi\alpha'}\gamma_{\sigma\sigma}\partial_\tau X^i = \frac{p^+}{\ell}\partial_\tau X^i . \tag{1.3.18}$$

The Hamiltonian is then

$$\begin{aligned} H &= p_-\partial_\tau x^- - L + \int_0^\ell d\sigma\,\Pi_i\partial_\tau X^i \\ &= \frac{\ell}{4\pi\alpha' p^+}\int_0^\ell d\sigma\left(2\pi\alpha'\Pi^i\Pi^i + \frac{1}{2\pi\alpha'}\partial_\sigma X^i\partial_\sigma X^i\right) . \end{aligned} \tag{1.3.19}$$

This is just the Hamiltonian for $D-2$ free fields X^i, with p^+ a conserved quantity. The equations of motion are

$$\partial_\tau x^- = \frac{\partial H}{\partial p_-} = \frac{H}{p^+} , \quad \partial_\tau p^+ = -\frac{\partial H}{\partial x^-} = 0 , \tag{1.3.20a}$$

$$\partial_\tau X^i = \frac{\delta H}{\delta\Pi^i} = 2\pi\alpha' c\Pi^i , \quad \partial_\tau\Pi^i = -\frac{\delta H}{\delta X^i} = \frac{c}{2\pi\alpha'}\partial_\sigma^2 X^i , \tag{1.3.20b}$$

implying the wave equation

$$\partial_\tau^2 X^i = c^2\partial_\sigma^2 X^i \tag{1.3.21}$$

with velocity $c = \ell/2\pi\alpha' p^+$. We will not develop the interactions in the light-cone formalism, but for these it is useful to choose the coordinate length ℓ for each string to be proportional to p^+, so that $c = 1$. Since p^+ is positive and conserved, the total string length is then also conserved.

The transverse coordinates satisfy a free wave equation so it is useful to expand in normal modes. In fact $X^\pm$ also satisfy the free wave equation. For X^+ this is trivial, while for X^- it requires a short calculation. The general solution to the wave equation with boundary condition (1.3.15) is

$$X^i(\tau,\sigma) = x^i + \frac{p^i}{p^+}\tau + i(2\alpha')^{1/2}\sum_{\substack{n=-\infty \\ n\neq 0}}^{\infty} \frac{1}{n}\alpha_n^i \exp\left(-\frac{\pi i n c\tau}{\ell}\right)\cos\frac{\pi n\sigma}{\ell} . \quad (1.3.22)$$

Reality of X^i requires $\alpha_{-n}^i = (\alpha_n^i)^\dagger$. We have defined the center-of-mass variables

$$x^i(\tau) = \frac{1}{\ell}\int_0^\ell d\sigma\, X^i(\tau,\sigma) , \quad (1.3.23a)$$

$$p^i(\tau) = \int_0^\ell d\sigma\, \Pi^i(\tau,\sigma) = \frac{p^+}{\ell}\int_0^\ell d\sigma\, \partial_\tau X^i(\tau,\sigma) , \quad (1.3.23b)$$

which are the average position and the total momentum. These are the Heisenberg operators; the Schrödinger operators $x^i \equiv x^i(0)$ and $p^i \equiv p^i(0)$ appear in the expansion (1.3.22).

To quantize, impose the equal time canonical commutation relations

$$[x^-, p^+] = i\eta^{-+} = -i , \quad (1.3.24a)$$

$$[X^i(\sigma), \Pi^j(\sigma')] = i\delta^{ij}\delta(\sigma-\sigma') , \quad (1.3.24b)$$

with all other commutators between the independent variables vanishing. In terms of the Fourier components,

$$[x^i, p^j] = i\delta^{ij} , \quad (1.3.25a)$$

$$[\alpha_m^i, \alpha_n^j] = m\delta^{ij}\delta_{m,-n} . \quad (1.3.25b)$$

For each m and i the modes satisfy a harmonic oscillator algebra with nonstandard normalization

$$\alpha_m^i \sim m^{1/2} a , \quad \alpha_{-m}^i \sim m^{1/2} a^\dagger , \quad m > 0 , \quad (1.3.26)$$

where $[a, a^\dagger] = 1$. The oscillators are labeled by the direction of oscillation i and the harmonic m.

The state $|0;k\rangle$, where $k = (k^+, k^i)$, is defined to be annihilated by the lowering operators and to be an eigenstate of the center-of-mass momenta,

$$p^+|0;k\rangle = k^+|0;k\rangle , \quad p^i|0;k\rangle = k^i|0;k\rangle , \quad (1.3.27a)$$

$$\alpha_m^i|0;k\rangle = 0 , \quad m > 0 . \quad (1.3.27b)$$

A general state can be built by acting on $|0;k\rangle$ with the raising operators,

$$|N;k\rangle = \left[\prod_{i=2}^{D-1}\prod_{n=1}^{\infty}\frac{(\alpha_{-n}^i)^{N_{in}}}{(n^{N_{in}}N_{in}!)^{1/2}}\right]|0;k\rangle . \quad (1.3.28)$$

That is, the independent states can be labeled by the center-of-mass momenta k^+ and k^i, and by the occupation numbers N_{in} for each mode (i,n), where $i = 2,\ldots,D-1$ and $n = 1,\ldots,\infty$. The center-of-mass momenta are just the degrees of freedom of a point particle, while the oscillators represent an infinite number of internal degrees of freedom. Every choice of these occupation numbers corresponds, from the spacetime point of view, to a different particle or spin state.

Note that the states (1.3.28) form the Hilbert space $\mathscr{H}_1$ of a single open string, and in particular the state $|0;0\rangle$ is the ground state of a single string with zero momentum, not the zero-string vacuum state. We will call the latter $|\text{vacuum}\rangle$ to make this clear. The various operators appearing above do not create or destroy strings but act within the space of states of a single string. The n-string Hilbert space $\mathscr{H}_n$ would be formed as the product of n copies of the space (1.3.28); the wavefunction must be symmetrized because, as we will see, all these states have integer spin. The full Hilbert space of the string theory, at least in the free limit that we are considering here, is then the sum

$$\mathscr{H} = |\text{vacuum}\rangle \oplus \mathscr{H}_1 \oplus \mathscr{H}_2 \oplus \ldots . \quad (1.3.29)$$

Inserting the mode expansion (1.3.22) into the Hamiltonian (1.3.19) gives

$$H = \frac{p^i p^i}{2p^+} + \frac{1}{2p^+\alpha'}\left(\sum_{n=1}^{\infty}\alpha_{-n}^i\alpha_n^i + A\right) . \quad (1.3.30)$$

In the Hamiltonian H, the order of operators is ambiguous. We have put the lowering operators on the right and the raising operators on the left, and included an unknown constant A from the commutators. In a careful treatment of light-cone quantization, this constant is determined as follows. The choice of light-cone gauge has obscured the Lorentz invariance of the theory. It is necessary to check this by finding the operators $M^{\mu\nu}$ that generate Lorentz transformations, and verifying that they have the correct algebra with p^μ and with each other. One finds that this is the case only for the specific value $A = -1$, and only if the spacetime dimensionality is precisely $D = 26$.

We do not wish to spend this much time on light-cone quantization, so we will try to show that the result is plausible without a systematic treatment. We assure the reader that we will obtain the values of A and D honestly in the conformal gauge approach, which will also give us insight

into why Lorentz invariance would be lost in the light-cone approach for the wrong A or D.

First, we assert that the operator ordering constant in the Hamiltonian for a free field comes from summing the zero-point energies of each oscillator mode, $\frac{1}{2}\omega$ for a bosonic field like X^μ. Equivalently, it always works out that the natural operator order is averaged, $\frac{1}{2}\omega(aa^\dagger + a^\dagger a)$, which is the same as $\omega(a^\dagger a + \frac{1}{2})$. In H this would give

$$A = \frac{D-2}{2}\sum_{n=1}^{\infty} n\,, \tag{1.3.31}$$

the factor of $D-2$ coming from the sum over transverse directions. The zero-point sum diverges. It can be evaluated by regulating the theory and then being careful to preserve Lorentz invariance in the renormalization. This leads to the odd result

$$\sum_{n=1}^{\infty} n \to -\frac{1}{12}\,. \tag{1.3.32}$$

To motivate this, insert a smooth cutoff factor

$$\exp(-\epsilon\gamma_{\sigma\sigma}^{-1/2}|k_\sigma|) \tag{1.3.33}$$

into the sum, where $k_\sigma = n\pi/\ell$ and the factor of $\gamma_{\sigma\sigma}^{-1/2}$ is included to make this invariant under σ reparameterizations. The zero-point constant is then

$$\begin{aligned} A &\to \frac{D-2}{2}\sum_{n=1}^{\infty} n\,\exp\left[-\epsilon n(\pi/2p^+\alpha'\ell)^{1/2}\right] \\ &= \frac{D-2}{2}\left(\frac{2\ell p^+\alpha'}{\epsilon^2\pi} - \frac{1}{12} + O(\epsilon)\right)\,. \end{aligned} \tag{1.3.34}$$

The cutoff-dependent first term is proportional to the length ℓ of the string and can be canceled by a counterterm in the action proportional to $\int d^2\sigma(-\gamma)^{1/2}$. In fact, Weyl invariance *requires* that it be canceled, leaving only the cutoff-independent second term,

$$A = \frac{2-D}{24}\,. \tag{1.3.35}$$

This finite remainder is an example of a Casimir energy, coming from the fact that the string has a finite length.

As for the point particle, $p^- = H$, so

$$m^2 = 2p^+H - p^i p^i = \frac{1}{\alpha'}\left(N + \frac{2-D}{24}\right)\,, \tag{1.3.36}$$

where N is the *level*

$$N = \sum_{i=2}^{D-1}\sum_{n=1}^{\infty} nN_{in}\,. \tag{1.3.37}$$

The mass of each state is thus determined in terms of the level of excitation.

Now let us look at some of the light string states. The lightest is

$$|0;k\rangle\,, \quad m^2 = \frac{2-D}{24\alpha'}\,. \tag{1.3.38}$$

The mass is negative if $D > 2$: the state is a tachyon. In field theory the potential energy for a scalar field is $\frac{1}{2}m^2\phi^2$, so the negative mass-squared means that the no-string 'vacuum' is actually unstable, like the symmetric state in a spontaneously broken theory. It is a complicated question whether the bosonic string has any stable vacuum, and the answer is not known. Starting with our study of the superstring in volume two, we will see that there *are* tachyon-free string theories. For now, the simplest way to proceed is to ignore this instability and use the bosonic string theory as our model for developing string technology. Occasionally we will encounter divergences due to the tachyon, but these will not interfere with our purpose.

The lowest excited states of the string are obtained by exciting one of the $n = 1$ modes once:

$$\alpha_{-1}^i|0;k\rangle\,, \quad m^2 = \frac{26-D}{24\alpha'}\,. \tag{1.3.39}$$

Lorentz invariance now requires a specific value of D as follows. The analysis of spin is different for massive and massless particles. For a massive particle, one goes to the rest frame $p^\mu = (m,0,\ldots,0)$. The internal states then form a representation of the spatial rotation group $SO(D-1)$. For a massless particle there is no rest frame; choose the frame $p^\mu = (E,E,0,\ldots,0)$. The $SO(D-2)$ acting on the transverse directions leaves p^μ invariant, and the internal states form a representation of this smaller group. This is familiar from $D = 4$: massive particles are labeled by spin j, the $SO(3)$ representation, and so have $2j+1$ states. Massless particles are labeled by their helicity λ, which is their eigenvalue under the single generator of $SO(2)$. Lorentz invariance alone thus requires only one state, though CPT symmetry takes λ to $-\lambda$ and so requires two states for $\lambda \neq 0$.

In D dimensions, a massive vector particle thus has $D-1$ spin states while a massless vector need have only $D-2$ states. At the first level we found only the $D-2$ states $\alpha_{-1}^i|0;k\rangle$, so these must be massless and

$$A = -1, \quad D = 26\,. \tag{1.3.40}$$

This is a striking and important result: the spectrum is Lorentz-invariant *only* if the number of spacetime dimensions is $D = 26$. The classical theory

is Lorentz-invariant for any D, but there is an anomaly — the symmetry is not preserved by the quantization except when $D = 26$. In later chapters we will obtain a deeper understanding of how this comes about and what it implies.

We conclude this section with a few more comments about spin. The light-cone quantization singles out two directions and leaves manifest only the $SO(D-2)$ that acts in the transverse directions. The spin generators for the transverse directions are

$$S^{ij} = -i\sum_{n=1}^{\infty}\frac{1}{n}(\alpha^i_{-n}\alpha^j_n - \alpha^j_{-n}\alpha^i_n)\ ; \tag{1.3.41}$$

antisymmetry on ij together with Lorentz invariance allows no zero-point constant. For a massless particle, the full $SO(D-2)$ spin symmetry is made manifest by choosing the momentum to lie along the 1-direction singled out in the quantization.

For a massive particle, only an $SO(D-2)$ subgroup of the $SO(D-1)$ spin symmetry will be manifest in light-cone quantization, but this is still quite useful. For example, the $(D-1)$-dimensional vector representation of $SO(D-1)$ breaks up into an invariant and a $(D-2)$-vector under the $SO(D-2)$ acting on the transverse directions,

$$\mathbf{v} = (v^1, 0, \ldots, 0) + (0, v^2, \ldots, v^{D-1})\ . \tag{1.3.42}$$

Thus, if a massive particle is in the vector representation of $SO(D-1)$, we will see a scalar and a vector when we look at the transformation properties under $SO(D-2)$. This idea extends to any representation: one can always reconstruct the full $SO(D-1)$ spin representation from the behavior under $SO(D-2)$.

The higher excited states of the string, which are massive, do form full representations of $SO(D-1)$. One can check this by hand for a few levels, and in a more complete analysis it follows from the existence of the full set of Lorentz generators. We will show this indirectly in the coming chapters, by finding a consistent covariant quantization.

At level N, the maximum eigenvalue of a given spin component, say S^{23}, is N, obtained by acting N times with $\alpha^2_{-1} + i\alpha^3_{-1}$. Thus,

$$S^{23} \leq 1 + \alpha' m^2\ . \tag{1.3.43}$$

The slope in this inequality is known as the *Regge slope*. Meson resonances obey a linear relation of this form, with $\alpha' \sim (1\ \text{GeV})^{-2}$. For this and other reasons, string theory was originally proposed in 1970 as a theory of the strong interaction. Within a few years, however, the $SU(3)$ gauge theory QCD was discovered to be the correct theory, with string-like behavior only in certain limits.

Now that we are considering strings as the unified theory of particle physics, gravity, and quantum mechanics, α' will be of the order of the natural scale determined by the fundamental constants of gravity and quantum mechanics, M_P^{-2}. In particular, masses that are not zero in string theory are of order M_P. This is so large compared to experimentally accessible energy scales that these particles appear only in virtual states. Thus, we will be especially concerned with the massless string spectrum, since this must include all the particles of the Standard Model. Of course, most known particles are massive, but these masses are so small compared to M_P that they are zero to first approximation and become non-zero due to small symmetry-breaking effects.

Some of the interest in string theory has been driven by the possibility that the strong interaction theory QCD is equivalent to a string theory, at least in some approximation. There has been cross-fertilization between this idea and the idea of strings as a fundamental theory. The connection between QCD and string theory is outside the focus of this book, but the recent developments described in chapter 14 have also given new insight in this direction.

1.4 Closed and unoriented strings

The light-cone quantization of the closed string is quite similar to that of the open string. Again impose the gauge conditions (1.3.8). In the open string, these determined the gauge completely. In the closed string there is still some extra coordinate freedom,

$$\sigma' = \sigma + s(\tau) \bmod \ell\ , \tag{1.4.1}$$

because the point $\sigma = 0$ can be chosen anywhere along the string. Most of this remaining freedom can be fixed by the additional gauge condition

$$\gamma_{\tau\sigma}(\tau, 0) = 0\ . \tag{1.4.2}$$

That is, the line $\sigma = 0$ is orthogonal to the lines of constant τ. This determines the line $\sigma = 0$ except for an overall σ-translation. Conditions (1.3.8) and (1.4.1) thus fix all of the gauge freedom except for τ-independent translations of σ,

$$\sigma' = \sigma + s \bmod \ell\ . \tag{1.4.3}$$

We will deal with this extra gauge freedom later.

The analysis is now parallel to that for the open string. The Lagrangian, canonical momenta, Hamiltonian, and equation of motion are as in the open string, eqs. (1.3.16)–(1.3.21). The general periodic solution to the

equation of motion is

$$X^i(\tau,\sigma) = x^i + \frac{p^i}{p^+}\tau + i\left(\frac{\alpha'}{2}\right)^{1/2} \times \sum_{\substack{n=-\infty \\ n\neq 0}}^{\infty} \left\{ \frac{\alpha_n^i}{n} \exp\left[-\frac{2\pi i n(\sigma+c\tau)}{\ell}\right] + \frac{\tilde{\alpha}_n^i}{n} \exp\left[\frac{2\pi i n(\sigma-c\tau)}{\ell}\right] \right\} . \tag{1.4.4}$$

In the closed string there are two independent sets of oscillators, α_n^μ and $\tilde{\alpha}_n^\mu$, corresponding to left-moving and right-moving waves along the string. In the open string the boundary condition at the endpoints tied these together. The independent degrees of freedom are again the transverse oscillators and the transverse and longitudinal center-of-mass variables,

$$\alpha_n^i,\ \tilde{\alpha}_n^i,\ x^i,\ p^i,\ x^-,\ p^+, \tag{1.4.5}$$

with canonical commutators

$$[x^-, p^+] = -i\,, \tag{1.4.6a}$$
$$[x^i, p^j] = i\delta^{ij}\,, \tag{1.4.6b}$$
$$[\alpha_m^i, \alpha_n^j] = m\delta^{ij}\delta_{m,-n}\,, \tag{1.4.6c}$$
$$[\tilde{\alpha}_m^i, \tilde{\alpha}_n^j] = m\delta^{ij}\delta_{m,-n}\,. \tag{1.4.6d}$$

Starting from the state $|0,0;k\rangle$, which has center-of-mass momentum k^μ and is annihilated by α_m^is and $\tilde{\alpha}_m^i$s for $m>0$, the general state is

$$|N,\tilde{N};k\rangle = \left[\prod_{i=2}^{D-1}\prod_{n=1}^{\infty} \frac{(\alpha_{-n}^i)^{N_{in}}(\tilde{\alpha}_{-n}^i)^{\tilde{N}_{in}}}{(n^{N_{in}}N_{in}!n^{\tilde{N}_{in}}\tilde{N}_{in}!)^{1/2}}\right]|0,0;k\rangle\,. \tag{1.4.7}$$

The mass formula is

$$\begin{aligned} m^2 &= 2p^+H - p^ip^i \\ &= \frac{2}{\alpha'}\left[\sum_{n=1}^{\infty}(\alpha_{-n}^i\alpha_n^i + \tilde{\alpha}_{-n}^i\tilde{\alpha}_n^i) + A + \tilde{A}\right] \\ &= \frac{2}{\alpha'}(N+\tilde{N}+A+\tilde{A})\,. \end{aligned} \tag{1.4.8}$$

We have broken up the level and zero-point constant into the part from the right-moving modes and the part from the left. Summing zero-point energies again gives

$$A = \tilde{A} = \frac{2-D}{24}\,. \tag{1.4.9}$$

There is one further restriction on the state due to the remaining gauge freedom, the σ-translations (1.4.3). The physical spectrum is obtained

by restricting to gauge-invariant states. The operator that generates the σ-translations is

$$\begin{aligned} P &= -\int_0^\ell d\sigma\,\Pi^i\partial_\sigma X^i \\ &= \frac{2\pi}{\ell}\left[\sum_{n=1}^{\infty}(\alpha_{-n}^i\alpha_n^i - \tilde{\alpha}_{-n}^i\tilde{\alpha}_n^i) + A - \tilde{A}\right] \\ &= \frac{2\pi}{\ell}(N-\tilde{N})\,. \end{aligned} \tag{1.4.10}$$

States must therefore satisfy

$$N = \tilde{N}\,. \tag{1.4.11}$$

The lightest closed string state is

$$|0,0;k\rangle\,,\quad m^2 = \frac{2-D}{6\alpha'}\,, \tag{1.4.12}$$

which is again a tachyon. The first excited states are

$$\alpha_{-1}^i\tilde{\alpha}_{-1}^j|0,0;k\rangle\,,\quad m^2 = \frac{26-D}{6\alpha'}\,. \tag{1.4.13}$$

As with the open string, these states do not add up to complete representations of $SO(D-1)$ and so this level must be massless. Thus,

$$A = \tilde{A} = -1\,,\quad D = 26\,. \tag{1.4.14}$$

The states (1.4.13) transform as a 2-tensor under $SO(D-2)$. This is a reducible representation: it decomposes into a symmetric traceless tensor, an antisymmetric tensor, and a scalar. That is, any tensor e^{ij} can be decomposed

$$e^{ij} = \frac{1}{2}\left(e^{ij}+e^{ji}-\frac{2}{D-2}\delta^{ij}e^{kk}\right) + \frac{1}{2}\left(e^{ij}-e^{ji}\right) + \frac{1}{D-2}\delta^{ij}e^{kk}\,, \tag{1.4.15}$$

and the three separate terms do not mix under rotations.

At any mass level, N_{in} and $\tilde{N}_{in}$ are independent except for the $N=\tilde{N}$ constraint. Thus, the closed string spectrum at $m^2 = 4(N-1)/\alpha'$ is the product of two copies of the $m^2=(N-1)/\alpha'$ level of the open string.

Let us now discuss some general issues. The two-dimensional diff invariance removed two families of normal modes from the spectrum. If we tried to make a covariant theory without this invariance, we would have to generalize the transverse commutator (1.3.25) to

$$[\alpha_m^\mu, \alpha_n^\nu] = m\eta^{\mu\nu}\delta_{m,-n}\,. \tag{1.4.16}$$

Lorentz invariance forces the timelike oscillators to have the wrong-sign commutator. A state with an odd number of timelike excitations will then have a negative norm. This is inconsistent with quantum mechanics,

since the norm is a probability. In fact a theory based on (1.4.16) came before the string theory we have described, and the latter was discovered by requiring that the negative norm states never be produced in physical processes. The commutator (1.4.16) arises in a covariant quantization of string theory (which we shall use beginning in the next chapter), with the coordinate invariance appearing through constraints on the spectrum to eliminate the unphysical states.

The existence of a massless vector in the open string theory and of the massless symmetric and antisymmetric tensors in the closed string theory is striking. General principles require that a massless vector couple to a conserved current and therefore that the theory have a gauge invariance. This is our first example of the massless gauge particles that are present in all fundamental string theories. Actually, this particular gauge boson, call it the photon, is not so interesting, because the gauge group is only $U(1)$ (there is only one photon) and because it turns out that all particles in the theory are neutral. In chapter 6 we will discuss a simple generalization of open string theory, adding *Chan–Paton* degrees of freedom at the string endpoints, which leads to $U(n)$, $SO(n)$, and $Sp(n)$ gauge groups. Similarly, a massless symmetric tensor particle must couple to a conserved symmetric tensor source. The only such source is the energy-momentum tensor; some additional possibilities arise in special cases, but these are not relevant here. Coupling to this in a consistent way requires the theory to have *spacetime* coordinate invariance. Thus the massless symmetric tensor is the graviton, and general relativity is contained as one small piece of the closed string theory. The massless antisymmetric tensor is known as a 2-form gauge boson. As we will see in section 3.7, there is also a local spacetime symmetry associated with it.

These spacetime gauge and coordinate invariances were not evident in our starting point, the Nambu–Goto or Polyakov theory. We have certainly discovered them in a backwards way: adding up zero-point energies, requiring a Lorentz-invariant spectrum, and citing general results about interactions of massless particles. In gauge theory and in general relativity, one starts with a spacetime symmetry principle. In string theory, too, it would seem that we should first figure out what the full spacetime symmetry is, and use this to define the theory. There are various attempts in this direction, but a complete picture has not yet emerged. For now, let us note one thing: we have observed that the local *world-sheet* symmetry diff removes the unphysical timelike and longitudinal normal modes which would lead to negative norm string states. For example, the open string photon would otherwise have D states rather than $D-2$. From the spacetime point of view, it is the local *spacetime* invariances that remove these same unphysical states. So there is at least some connection between the local symmetries of the world-sheet and those of spacetime.

Anticipating our interest in four dimensions, let us look at the massless string states obtained by exciting only oscillators in the 2- and 3-directions, and classifying them by the helicity $\lambda = S^{23}$. One finds that the photon has $\lambda = \pm 1$ and the graviton $\lambda = \pm 2$, justifying the names. The closed string scalar and antisymmetric tensor each give rise to a $\lambda = 0$ state. These are called the dilaton and the axion, respectively, and we will discuss their physics at various points.

We conclude this section with a discussion of one other generalization, the unoriented string theory. The theories we have been discussing are *oriented* string theories. We have not considered the coordinate transformation

$$\sigma' = \ell - \sigma \,, \quad \tau' = \tau \,, \tag{1.4.17}$$

which changes the orientation (handedness) of the world-sheet. This symmetry is generated by the world-sheet parity operator Ω. Carrying out (1.4.17) twice gives the identity, so $\Omega^2 = 1$ and the eigenvalues of Ω are ± 1. From the mode expansions (1.3.22) and (1.4.4) we see that

$$\Omega \alpha_n^\mu \Omega^{-1} = (-1)^n \alpha_n^\mu \tag{1.4.18}$$

in the open string and

$$\Omega \alpha_n^\mu \Omega^{-1} = \tilde{\alpha}_n^\mu \,, \tag{1.4.19a}$$

$$\Omega \tilde{\alpha}_n^\mu \Omega^{-1} = \alpha_n^\mu \tag{1.4.19b}$$

in the closed string. We define the phase of Ω by fixing $\Omega = +1$ for the ground states $|0;k\rangle$ and $|0,0;k\rangle$; later we will see that this choice is required in order for Ω to be conserved by interactions. Then

$$\Omega |N;k\rangle = (-1)^N |N;k\rangle \,, \tag{1.4.20a}$$

$$\Omega |N,\tilde{N};k\rangle = |\tilde{N},N;k\rangle \,. \tag{1.4.20b}$$

There are consistent interacting string theories, the *unoriented* theories, in which only the $\Omega = +1$ states are kept. Focusing on the massless states, the open string photon has $\Omega = -1$ and is absent in the unoriented theory. Acting on the massless tensor states in the closed string, the parity operator Ω takes e^{ij} to e^{ji}, so the graviton and dilaton are present in the unoriented theory while the antisymmetric tensor is not. Notice that both the open and closed string tachyons survive in the unoriented theory; we will have to work harder to remove these.

Let us mention two other constraints that will emerge from the study of the interactions. First, it is possible to have a consistent theory with only closed strings, or with closed and open strings, but not with open strings alone: closed strings can always be produced in the scattering of open strings. Second, oriented or unoriented open strings can only couple to

closed strings of the same type. We list the possible combinations together with their massless spectra, with $G_{\mu\nu}$ representing the graviton, $B_{\mu\nu}$ the antisymmetric tensor, Φ the dilaton, and A_μ the photon:

1. Closed oriented bosonic string: $G_{\mu\nu}$, $B_{\mu\nu}$, Φ .
2. Closed unoriented bosonic string: $G_{\mu\nu}$, Φ .
3. Closed plus open oriented bosonic string: $G_{\mu\nu}$, $B_{\mu\nu}$, Φ, A_μ .
4. Closed plus open unoriented bosonic string: $G_{\mu\nu}$, Φ .

All of these have the graviton and dilaton, as well as the tachyon. In chapter 6 we will discuss more general open string theories, with Chan–Paton degrees of freedom at the endpoints. The massless oriented open strings will then be $U(n)$ gauge bosons, and the massless unoriented open strings will be $SO(n)$ or $Sp(n)$ gauge bosons.

Exercises

1.1 (a) Show that in the nonrelativistic limit the action $S_{\rm pp}$ has the usual nonrelativistic form, kinetic energy minus potential energy, with the potential energy being the rest mass.
(b) Show that for a string moving nonrelativistically, the Nambu–Goto action reduces to a kinetic term minus a potential term proportional to the length of the string. Show that the kinetic energy comes only from the *transverse* velocity of the string. Calculate the mass per unit length, as determined from the potential term and also from the kinetic term.

1.2 Show that classical equations of motion imply that the ends of the open string move at the speed of light.

1.3 For world-sheets with boundary, show that

$$\chi = \frac{1}{4\pi}\int_M d\tau\, d\sigma\, (-\gamma)^{1/2}R + \frac{1}{2\pi}\int_{\partial M} ds\, k$$

is Weyl-invariant. Here ds is the proper time along the boundary in the metric γ_{ab}, and k is the geodesic curvature of the boundary,

$$k = \pm t^a n_b \nabla_a t^b,$$

where t^a is a unit vector tangent to the boundary and n^a is an outward pointing unit vector orthogonal to t^a. The upper sign is for a timelike boundary and the lower sign for a spacelike boundary.

1.4 Show that the open string states at levels $m^2 = 1/\alpha'$ and $2/\alpha'$ form complete representations of $SO(D-1)$. You will need to work out the $SO(D-2)$ content of various symmetric and antisymmetric tensor representations of $SO(D-1)$. You can carry this to higher levels if you like, but you will need tensors of mixed symmetry and Young tableaux.

1.5 Extend the sum (1.3.32) to the 'twisted' case

$$\sum_{n=1}^{\infty}(n-\theta)$$

with θ a constant. That is, $k_\sigma = (n-\theta)\pi/\ell$. The answer is given in eq. (2.9.19). You should find that the cutoff-dependent term is independent of θ.

In the following exercises one has the usual open or closed string boundary conditions (Neumann or periodic) on X^μ for $\mu = 0,\ldots,24$ but a different boundary condition on X^{25}. Each of these has an important physical interpretation, and will be developed in detail in chapter 8. Find the mode expansion, the mass spectrum, and (for the closed string) the constraint from σ-translation invariance in terms of the occupation numbers. In some cases you need the result of exercise 1.5.

1.6 Open strings with

$$X^{25}(\tau,0) = 0\,, \quad X^{25}(\tau,\ell) = y$$

with y a constant. This is an open string with both ends on D-branes.

1.7 Open strings with

$$X^{25}(\tau,0) = 0\,, \quad \partial^\sigma X^{25}(\tau,\ell) = 0\,.$$

This is an open string with one end on a D-brane and one end free.

1.8 Closed strings with

$$X^{25}(\tau,\sigma+\ell) = X^{25}(\tau,\sigma) + 2\pi R$$

with R a constant. This is a winding string in toroidal (periodic) compactification. In this case p^{25} must be a multiple of $1/R$.

1.9 Closed strings with

$$X^{25}(\tau,\sigma+\ell) = -X^{25}(\tau,\sigma)\,.$$

This is a twisted string in orbifold compactification.

2
Conformal field theory

In this chapter we develop a number of necessary ideas and techniques from the world-sheet quantum field theory, including the operator product expansion, conformal invariance, the Virasoro algebra, and vertex operators. The focus is on conformally invariant field theory in two flat dimensions; as we will see in the next chapter, this is what we are left with after fixing the local symmetries of the string world-sheet.

2.1 Massless scalars in two dimensions

We will start with the example of D free scalar fields in two dimensions, $X^\mu(\sigma^1,\sigma^2)$. We will refer to these two dimensions as the world-sheet, anticipating the application to string theory. The action is

$$S = \frac{1}{4\pi\alpha'}\int d^2\sigma \left(\partial_1 X^\mu \partial_1 X_\mu + \partial_2 X^\mu \partial_2 X_\mu\right) . \tag{2.1.1}$$

This is the Polyakov action (1.2.13), except that the world-sheet metric γ_{ab} has been replaced with a flat Euclidean metric δ_{ab}, signature $(+,+)$. The overall sign change of the action is a result of the Euclidean convention (A.1.31). As we will see, most string calculations are carried out on a Euclidean world-sheet. At least for flat metrics, the relation between Euclidean and Minkowski amplitudes is given by a standard analytic continuation, explained in the appendix. In fact, the results of the present chapter apply equally to a Minkowski world-sheet, and in the first seven sections all equations make sense if σ^2 is replaced with $i\sigma^0$. For the index μ we still take the flat Minkowski metric.

It is straightforward to quantize the action (2.1.1) canonically, finding the spectrum, vacuum expectation values, and so on. We have done essentially this in chapter 1, after having gone to light-cone gauge. Here we will take a somewhat different route, developing first various local properties such

as equations of motion, operator products, Ward identities, and conformal invariance, before working our way around to the spectrum. It will be efficient for us to use the path integral formalism. This is reviewed in the appendix. We will be using the path integral representation primarily to derive operator equations (to be defined below); these can also be derived in a Hilbert space formalism.

It is very useful to adopt complex coordinates

$$z = \sigma^1 + i\sigma^2 , \quad \bar{z} = \sigma^1 - i\sigma^2 . \tag{2.1.2}$$

We will use a bar for the complex conjugates of z and other simple variables, and a star for the complex conjugates of longer expressions. Define also

$$\partial_z = \frac{1}{2}(\partial_1 - i\partial_2) , \quad \partial_{\bar{z}} = \frac{1}{2}(\partial_1 + i\partial_2) . \tag{2.1.3}$$

These derivatives have the properties

$$\partial_z z = 1 , \quad \partial_z \bar{z} = 0 , \quad \partial_{\bar{z}} z = 0 , \quad \partial_{\bar{z}} \bar{z} = 1 . \tag{2.1.4}$$

It is conventional to abbreviate ∂_z to ∂ and $\partial_{\bar{z}}$ to $\bar{\partial}$ when this is not ambiguous. For a general vector v^a, define in the same way

$$v^z = v^1 + iv^2 , \quad v^{\bar{z}} = v^1 - iv^2 , \quad v_z = \frac{1}{2}(v^1 - iv^2) , \quad v_{\bar{z}} = \frac{1}{2}(v^1 + iv^2) . \tag{2.1.5}$$

For the indices 1, 2 the metric is the identity and we do not distinguish between upper and lower, while the complex indices are raised and lowered with

$$g_{z\bar{z}} = g_{\bar{z}z} = \frac{1}{2} , \quad g_{zz} = g_{\bar{z}\bar{z}} = 0 , \quad g^{z\bar{z}} = g^{\bar{z}z} = 2 , \quad g^{zz} = g^{\bar{z}\bar{z}} = 0 . \tag{2.1.6}$$

Note also that

$$d^2z = 2d\sigma^1 d\sigma^2 \tag{2.1.7}$$

with the factor of 2 from the Jacobian.[1] Note also that $d^2z\,|\det g|^{1/2} = d\sigma^1 d\sigma^2$. We define

$$\int d^2z\, \delta^2(z,\bar{z}) = 1 \tag{2.1.8}$$

so that $\delta^2(z,\bar{z}) = \frac{1}{2}\delta(\sigma^1)\delta(\sigma^2)$. Another useful result is the divergence

[1] This differs from much of the literature, where d^2z is defined as $d\sigma^1 d\sigma^2$. Correlated with this, the $\frac{1}{2}$ is omitted from the definition of $\delta^2(z,\bar{z})$.

theorem in complex coordinates,

$$\int_R d^2z\,(\partial_z v^z + \partial_{\bar z} v^{\bar z}) = i\oint_{\partial R}(v^z d\bar z - v^{\bar z} dz)\,, \tag{2.1.9}$$

where the contour integral circles the region R counterclockwise.

In this notation the action is

$$S = \frac{1}{2\pi\alpha'}\int d^2z\,\partial X^\mu \bar\partial X_\mu \tag{2.1.10}$$

and the classical equation of motion is

$$\partial\bar\partial X^\mu(z,\bar z) = 0\,. \tag{2.1.11}$$

The notation $X^\mu(z,\bar z)$ may seem redundant, since the value of z determines the value of $\bar z$, but it is useful to reserve the notation $f(z)$ for fields whose equation of motion makes them *analytic* (equivalently *holomorphic*) functions of z. Writing the equation of motion as

$$\partial(\bar\partial X^\mu) = \bar\partial(\partial X^\mu) = 0\,, \tag{2.1.12}$$

it follows that ∂X^μ is holomorphic and that $\bar\partial X^\mu$ is antiholomorphic (holomorphic in $\bar z$), hence the notations $\partial X^\mu(z)$ and $\bar\partial X^\mu(\bar z)$.

Under the Minkowski continuation $\sigma^2 = i\sigma^0$, a holomorphic field becomes a function only of $\sigma^0 - \sigma^1$ and an antiholomorphic field a function only of $\sigma^0 + \sigma^1$. We thus use as synonyms

$$\text{holomorphic} = \text{left-moving}\,, \tag{2.1.13a}$$

$$\text{antiholomorphic} = \text{right-moving}\,. \tag{2.1.13b}$$

This terminology is chosen to have maximal agreement with the literature, though for it to hold literally we would need to draw σ^1 increasing from right to left. We will tend to use the Euclidean terms early on and shift to the Minkowski terms as we discuss more of the spacetime physics.

Expectation values are defined by the path integral,

$$\langle\,\mathscr{F}[X]\,\rangle = \int [dX]\,\exp(-S)\mathscr{F}[X]\,, \tag{2.1.14}$$

where $\mathscr{F}[X]$ is any functional of X, such as a product of local operators. The path integral over X^0 is a wrong-sign Gaussian, so it should be understood to be defined by the analytic continuation $X^0 \to -iX^D$. The reader should not be distracted by this; we will discuss it further in the next chapter. We do not normalize $\langle\,\mathscr{F}[X]\,\rangle$ by dividing by $\langle\,1\,\rangle$.

The path integral of a total derivative is zero. This is true for ordinary bosonic path integrals, which can be regarded as the limit of an infinite number of ordinary integrals, as well as for more formal path integrals as

with Grassmann variables. Then

$$\begin{aligned} 0 &= \int [dX]\,\frac{\delta}{\delta X_\mu(z,\bar z)}\exp(-S) \\ &= -\int [dX]\,\exp(-S)\frac{\delta S}{\delta X_\mu(z,\bar z)} \\ &= -\left\langle \frac{\delta S}{\delta X_\mu(z,\bar z)} \right\rangle \\ &= \frac{1}{\pi\alpha'}\left\langle\,\partial\bar\partial X^\mu(z,\bar z)\,\right\rangle\,. \end{aligned} \tag{2.1.15}$$

The same calculation goes through if we have arbitrary additional insertions '...' in the path integral, as long as none of these additional insertions is at z. Thus

$$\left\langle\,\partial\bar\partial X^\mu(z,\bar z)\,\ldots\,\right\rangle = 0\,. \tag{2.1.16}$$

We can regard the additional insertions as preparing arbitrary initial and final states (or we could do the same thing with boundary conditions). The path integral statement (2.1.16) is thus the same as the statement in the Hilbert space formalism that

$$\partial\bar\partial\hat X^\mu(z,\bar z) = 0 \tag{2.1.17}$$

holds for all matrix elements of the operator $\hat X^\mu(z,\bar z)$. Thus we refer to relations that hold in the sense (2.1.16) as *operator equations*. The statement (2.1.17) is Ehrenfest's theorem that the classical equations of motion translate into operator equations.

The notation '...' in the path integral (2.1.16) implicitly stands for insertions that are located away from z, but it is interesting to consider also the case in which there is an insertion that might be coincident with z:

$$\begin{aligned} 0 &= \int [dX]\,\frac{\delta}{\delta X_\mu(z,\bar z)}\left[\exp(-S)X^\nu(z',\bar z')\right] \\ &= \int [dX]\,\exp(-S)\left[\eta^{\mu\nu}\delta^2(z-z',\bar z-\bar z') + \frac{1}{\pi\alpha'}\partial_z\partial_{\bar z}X^\mu(z,\bar z)X^\nu(z',\bar z')\right] \\ &= \eta^{\mu\nu}\left\langle\,\delta^2(z-z',\bar z-\bar z')\,\right\rangle + \frac{1}{\pi\alpha'}\partial_z\partial_{\bar z}\left\langle\,X^\mu(z,\bar z)X^\nu(z',\bar z')\,\right\rangle\,. \end{aligned} \tag{2.1.18}$$

That is, the equation of motion holds except at coincident points. Again this goes through with arbitrary additional insertions '...' in the path integral, as long as none of these additional fields is at $(z,\bar z)$ or $(z',\bar z')$:

$$\frac{1}{\pi\alpha'}\partial_z\partial_{\bar z}\left\langle\,X^\mu(z,\bar z)X^\nu(z',\bar z')\ldots\,\right\rangle = -\eta^{\mu\nu}\left\langle\,\delta^2(z-z',\bar z-\bar z')\ldots\,\right\rangle\,. \tag{2.1.19}$$

Thus,

$$\frac{1}{\pi\alpha'}\partial_z\partial_{\bar{z}}X^\mu(z,\bar{z})X^\nu(z',\bar{z}') = -\eta^{\mu\nu}\delta^2(z-z',\bar{z}-\bar{z}') \tag{2.1.20}$$

holds as an operator equation. In the Hilbert space formalism, the product in the path integral becomes a time-ordered product, and the delta function comes from the derivatives acting on the time-ordering. This connection is developed further in the appendix.

In free field theory, it is useful to introduce the operation of *normal ordering*. Normal ordered operators, denoted $:\mathscr{A}:$, are defined as follows,

$$:X^\mu(z,\bar{z}): = X^\mu(z,\bar{z})\,, \tag{2.1.21a}$$

$$:X^\mu(z_1,\bar{z}_1)X^\nu(z_2,\bar{z}_2): = X^\mu(z_1,\bar{z}_1)X^\nu(z_2,\bar{z}_2) + \frac{\alpha'}{2}\eta^{\mu\nu}\ln|z_{12}|^2\,, \tag{2.1.21b}$$

where

$$z_{ij} = z_i - z_j\,. \tag{2.1.22}$$

The reader may be familiar with normal ordering defined in terms of raising and lowering operators; these two definitions will be related later. The point of this definition is the property

$$\partial_1\bar{\partial}_1\,:X^\mu(z_1,\bar{z}_1)X^\nu(z_2,\bar{z}_2): = 0\,. \tag{2.1.23}$$

This follows from the operator equation (2.1.20) and the differential equation

$$\partial\bar{\partial}\ln|z|^2 = 2\pi\delta^2(z,\bar{z})\,. \tag{2.1.24}$$

Eq. (2.1.24) is obvious for $z \neq 0$ because $\ln|z|^2 = \ln z + \ln\bar{z}$, and the normalization of the delta function is easily checked by integrating both sides using eq. (2.1.9).

2.2 The operator product expansion

The basic object of interest in string perturbation theory will be the path integral expectation value of a product of local operators,

$$\langle\, \mathscr{A}_{i_1}(z_1,\bar{z}_1)\mathscr{A}_{i_2}(z_2,\bar{z}_2)\ldots\mathscr{A}_{i_n}(z_n,\bar{z}_n)\,\rangle\,, \tag{2.2.1}$$

where $\mathscr{A}_i$ is some basis for the set of local operators. It is particularly important to understand the behavior of this expectation value in the limit that two of the operators are taken to approach one another. The tool that gives a systematic description of this limit is the *operator product expansion (OPE)*, illustrated in figure 2.1. This states that a product of two

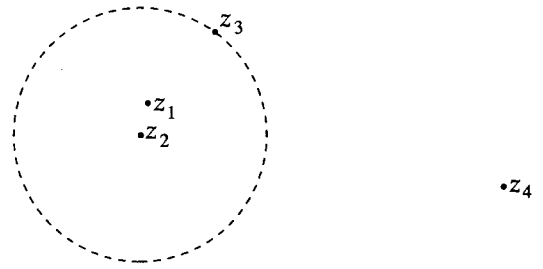


Fig. 2.1. Expectation value of a product of four local operators. The OPE gives the asymptotics as $z_1 \to z_2$ as a series where the pair of operators at z_1 and z_2 is replaced by a single operator at z_2. The radius of convergence is the distance to the nearest other operator, indicated by the dashed circle.

local operators close together can be approximated to arbitrary accuracy by a sum of local operators,

$$\mathscr{A}_i(\sigma_1)\mathscr{A}_j(\sigma_2) = \sum_k c^k_{ij}(\sigma_1-\sigma_2)\mathscr{A}_k(\sigma_2)\,. \tag{2.2.2}$$

Again this is an operator statement, meaning that it holds inside a general expectation value

$$\langle\, \mathscr{A}_i(\sigma_1)\mathscr{A}_j(\sigma_2)\ldots\rangle = \sum_k c^k_{ij}(\sigma_1-\sigma_2)\langle\, \mathscr{A}_k(\sigma_2)\ldots\rangle\,, \tag{2.2.3}$$

as long as the separation between σ_1 and σ_2 is small compared to the distance to any other operator. The *coefficient functions* $c^k_{ij}(\sigma_1-\sigma_2)$, which govern the dependence on the separation, depend on i, j, and k but not on the other operators in the expectation value; the dependence on the latter comes only through the expectation value on the right-hand side of eq. (2.2.3). The terms are conventionally arranged in order of decreasing size in the limit as $\sigma_1 \to \sigma_2$. This is analogous to an ordinary Taylor series, except that the coefficient functions need not be simple powers and can in fact be singular as $\sigma_1 \to \sigma_2$, as we will see even in the simplest example. Just as the Taylor series plays a central role in calculus, the OPE plays a central role in quantum field theory.

We will give now a derivation of the OPE for the X^μ theory, using the special properties of free field theory. In section 2.9 we will give a derivation for any conformally invariant field theory.

We have seen that the normal ordered product satisfies the naive equation of motion. Eq. (2.1.23) states that the operator product is a harmonic function of $(z_1,\bar{z}_1)$. A simple result from the theory of complex variables is that a harmonic function is locally the sum of a holomorphic and an

antiholomorphic function. In particular, this means that it is nonsingular as $z_1 \to z_2$ and can be freely Taylor expanded in z_{12} and $\bar{z}_{12}$. Thus,

$$X^\mu(z_1,\bar{z}_1)X^\nu(z_2,\bar{z}_2) = -\frac{\alpha'}{2}\eta^{\mu\nu}\ln|z_{12}|^2 + \sum_{k=1}^{\infty}\frac{1}{k!}\left[(z_{12})^k :X^\nu\partial^k X^\mu(z_2,\bar{z}_2): + (\bar{z}_{12})^k :X^\nu\bar{\partial}^k X^\mu(z_2,\bar{z}_2):\right] . \tag{2.2.4}$$

Terms with mixed $\partial\bar{\partial}$ derivatives vanish by the equation of motion. This equation and many others simplify in units in which $\alpha' = 2$, which is the most common convention in the literature. However, several other conventions are also used, so it is useful to keep α' explicit. For example, in *open* string theory equations simplify when $\alpha' = \frac{1}{2}$.

Eq. (2.2.4) has the form of an OPE. Like the equation of motion (2.1.23) from which it was derived, it is an operator statement. For an arbitrary expectation value involving the product $X^\mu(z_1,\bar{z}_1)X^\nu(z_2,\bar{z}_2)$ times fields at other points, it gives the behavior for $z_1 \to z_2$ as an infinite series, each term being a known function of z_{12} and/or $\bar{z}_{12}$ times the expectation value with a local operator replacing the pair.

OPEs are usually used as asymptotic expansions, the first few terms giving the dominant behavior at small separation. Most of our applications will be of this type, and we will often write OPEs as explicit singular terms plus unspecified nonsingular remainders. The use of '$\sim$' in place of '$=$' will mean 'equal up to nonsingular terms.' In fact, OPEs are actually convergent in conformally invariant field theories. This will be very important to us in certain applications: it makes it possible to reconstruct the entire theory from the coefficient functions. As an example, the free-field OPE (2.2.4) has a radius of convergence in any given expectation value which is equal to the distance to the nearest *other* insertion in the path integral. The operator product is harmonic except at the positions of operators, and in particular inside the dashed circle of figure 2.1, and convergence can then be shown by a standard argument from the theory of complex variables.

The various operators on the right-hand side of the OPE (2.2.4) involve products of fields at the same point. Usually in quantum field theory such a product is divergent and must be appropriately cut off and renormalized, but here the normal ordering renders it well-defined. Normal ordering is a convenient way to define composite operators in free field theory. It is of little use in most interacting field theories, because these have additional divergences from interaction vertices approaching the composite operator or one another. However, many of the field theories that we will be

interested in are free, and many others can be related to free field theories, so it will be worthwhile to develop normal ordering somewhat further.

The definition of normal ordering for arbitrary numbers of fields can be given recursively as

$$:X^{\mu_1}(z_1,\bar{z}_1)\dots X^{\mu_n}(z_n,\bar{z}_n): = X^{\mu_1}(z_1,\bar{z}_1)\dots X^{\mu_n}(z_n,\bar{z}_n) + \sum \text{subtractions} , \tag{2.2.5}$$

where the sum runs over all ways of choosing one, two, or more pairs of fields from the product and replacing each pair with $\frac{1}{2}\alpha'\eta^{\mu_i\mu_j}\ln|z_{ij}|^2$. For example,

$$:X^{\mu_1}(z_1,\bar{z}_1)X^{\mu_2}(z_2,\bar{z}_2)X^{\mu_3}(z_3,\bar{z}_3): = X^{\mu_1}(z_1,\bar{z}_1)X^{\mu_2}(z_2,\bar{z}_2)X^{\mu_3}(z_3,\bar{z}_3) + \left(\frac{\alpha'}{2}\eta^{\mu_1\mu_2}\ln|z_{12}|^2 X^{\mu_3}(z_3,\bar{z}_3) + 2 \text{ permutations}\right) . \tag{2.2.6}$$

We leave it to the reader to show that the definition (2.2.5) retains the desired property that the normal ordered product satisfies the naive equation of motion.

The definition can be compactly summarized as

$$:\mathscr{F}: = \exp\left(\frac{\alpha'}{4}\int d^2z_1 d^2z_2 \ln|z_{12}|^2 \frac{\delta}{\delta X^\mu(z_1,\bar{z}_1)}\frac{\delta}{\delta X_\mu(z_2,\bar{z}_2)}\right)\mathscr{F} , \tag{2.2.7}$$

where $\mathscr{F}$ is any functional of X. This is equivalent to eq. (2.2.5): the double derivative in the exponent contracts each pair of fields, and the exponential sums over any number of pairs with the factorial canceling the number of ways the derivatives can act. As an example of the use of this formal expression, act on both sides with the inverse exponential to obtain

$$\mathscr{F} = \exp\left(-\frac{\alpha'}{4}\int d^2z_1 d^2z_2 \ln|z_{12}|^2 \frac{\delta}{\delta X^\mu(z_1,\bar{z}_1)}\frac{\delta}{\delta X_\mu(z_2,\bar{z}_2)}\right) :\mathscr{F}: = :\mathscr{F}: + \sum \text{contractions} , \tag{2.2.8}$$

where a contraction is the opposite of a subtraction: sum over all ways of choosing one, two, or more pairs of fields from $:\mathscr{F}:$ and replacing each pair with $-\frac{1}{2}\alpha'\eta^{\mu_i\mu_j}\ln|z_{ij}|^2$.

The OPE for any pair of operators can be generated from

$$:\mathscr{F}::\mathscr{G}: = :\mathscr{F}\mathscr{G}: + \sum \text{cross-contractions} \tag{2.2.9}$$

for arbitrary functionals $\mathscr{F}$ and $\mathscr{G}$ of X. The sum now runs over all ways of contracting pairs with one field in $\mathscr{F}$ and one in $\mathscr{G}$. This can also be written

$$:\mathscr{F}::\mathscr{G}: = \exp\left(-\frac{\alpha'}{2}\int d^2z_1 d^2z_2 \ln|z_{12}|^2 \frac{\delta}{\delta X_F^\mu(z_1,\bar{z}_1)}\frac{\delta}{\delta X_{G\mu}(z_2,\bar{z}_2)}\right) :\mathscr{F}\mathscr{G}: , \tag{2.2.10}$$

where the functional derivatives act only on the fields in $\mathscr{F}$ or $\mathscr{G}$ respectively. This follows readily from eq. (2.2.7).

As an example,

$$
\begin{aligned}
&:\partial X^\mu(z)\partial X_\mu(z): \, :\partial' X^\nu(z')\partial' X_\nu(z'): \\
&\quad = :\partial X^\mu(z)\partial X_\mu(z)\partial' X^\nu(z')\partial' X_\nu(z'): \\
&\qquad -4\cdot\frac{\alpha'}{2}(\partial\partial' \ln|z-z'|^2) \, :\partial X^\mu(z)\partial' X_\mu(z'): \\
&\qquad +2\cdot\eta^\mu{}_\mu\left(-\frac{\alpha'}{2}\partial\partial' \ln|z-z'|^2\right)^2 \\
&\quad \sim \frac{D\alpha'^2}{2(z-z')^4} - \frac{2\alpha'}{(z-z')^2} \, :\partial' X^\mu(z')\partial' X_\mu(z'): \\
&\qquad -\frac{2\alpha'}{z-z'} \, :\partial'^2 X^\mu(z')\partial' X_\mu(z'): \, .
\end{aligned}
\tag{2.2.11}
$$

The second term in the equality comes from the four ways of forming a single pair and the third from the two ways of forming two pairs. In the final line we have put the OPE in standard form by Taylor expanding inside the normal ordering to express everything in terms of local operators at z' and putting the most singular terms first.

Another important example is

$$
\mathscr{F} = e^{ik_1\cdot X(z,\bar z)}, \quad \mathscr{G} = e^{ik_2\cdot X(0,0)} . \tag{2.2.12}
$$

The variations $\delta/\delta X_F^\mu$ and $\delta/\delta X_G^\mu$ give factors of ik_1^μ and ik_2^μ respectively, so the general result (2.2.10) becomes

$$
\begin{aligned}
:e^{ik_1\cdot X(z,\bar z)}: \, :e^{ik_2\cdot X(0,0)}: &= \exp\left(\frac{\alpha'}{2}k_1\cdot k_2 \ln|z|^2\right) \, :e^{ik_1\cdot X(z,\bar z)}e^{ik_2\cdot X(0,0)}: \\
&= |z|^{\alpha' k_1\cdot k_2} \, :e^{ik_1\cdot X(z,\bar z)}e^{ik_2\cdot X(0,0)}: \, .
\end{aligned}
\tag{2.2.13}
$$

To derive the OPE, Taylor expand inside the normal ordering to give

$$
:e^{ik_1\cdot X(z,\bar z)}: \, :e^{ik_2\cdot X(0,0)}: = |z|^{\alpha' k_1\cdot k_2} \, :e^{i(k_1+k_2)\cdot X(0,0)}[1+O(z,\bar z)]: \, . \tag{2.2.14}
$$

The exercises give further practice with normal ordering and the free-field OPE.

Note that the OPEs (2.2.2), (2.2.4), and so on have been written asymmetrically in σ_1 and σ_2, expanding around the latter point. They can also be cast in symmetric form by Taylor expanding the right-hand sides around $(\sigma_1+\sigma_2)/2$. The coefficient functions for the symmetric form behave simply under interchange of the two operators,

$$
c_{ij}^k(\sigma_1-\sigma_2)_{\text{sym}} = \pm c_{ji}^k(\sigma_2-\sigma_1)_{\text{sym}} , \tag{2.2.15}
$$

where the minus sign appears if $\mathscr{A}_i$ and $\mathscr{A}_j$ are both anticommuting. The asymmetric form is usually more convenient for calculation, so when the symmetry properties of the coefficient functions are needed one can work them out in the symmetric form and then convert to the asymmetric form.

2.3 Ward identities and Noether's theorem

World-sheet symmetries of course play an important role in string theory. In this section we first derive some general consequences of symmetry in field theory.

Consider a general field theory with action $S[\phi]$ in d spacetime dimensions, with $\phi_\alpha(\sigma)$ denoting general fields. Let there be a symmetry

$$
\phi'_\alpha(\sigma) = \phi_\alpha(\sigma) + \delta\phi_\alpha(\sigma) , \tag{2.3.1}
$$

where $\delta\phi_\alpha$ is proportional to an infinitesimal parameter ϵ. The product of the path integral measure and the weight $\exp(-S)$ is invariant,

$$
[d\phi'] \exp(-S[\phi']) = [d\phi] \exp(-S[\phi]) . \tag{2.3.2}
$$

A continuous symmetry in field theory implies the existence of a conserved current (*Noether's theorem*) and also *Ward identities*, which constrain the operator products of the current. To derive these results consider the change of variables,

$$
\phi'_\alpha(\sigma) = \phi_\alpha(\sigma) + \rho(\sigma)\delta\phi_\alpha(\sigma) . \tag{2.3.3}
$$

This is *not* a symmetry, the transformation law being altered by the inclusion of an arbitrary function $\rho(\sigma)$. The path integral measure times $\exp(-S)$ would be invariant if ρ were a constant, so its variation must be proportional to the gradient $\partial_a\rho$,

$$
\begin{aligned}
&[d\phi'] \exp(-S[\phi']) \\
&\quad = [d\phi] \exp(-S[\phi]) \left[1 + \frac{i\epsilon}{2\pi}\int d^d\sigma \, g^{1/2} j^a(\sigma)\partial_a\rho(\sigma) + O(\epsilon^2)\right] .
\end{aligned}
\tag{2.3.4}
$$

The unknown coefficient $j^a(\sigma)$ comes from the variation of the measure and the action, both of which are local, and so it must be a local function of the fields and their derivatives. Take the function ρ to be nonzero only in a small region, and consider a path integral with general insertions '...' *outside* this region; the insertions are therefore invariant under (2.3.3). Invariance of the path integral under change of variables gives

$$
\begin{aligned}
0 &= \int [d\phi'] \exp(-S[\phi']) \ldots - \int [d\phi] \exp(-S[\phi]) \ldots \\
&= \frac{\epsilon}{2\pi i}\int d^d\sigma \, g^{1/2}\rho(\sigma) \langle \nabla_a j^a(\sigma) \ldots \rangle ,
\end{aligned}
\tag{2.3.5}
$$

where the limited support of ρ has allowed us to integrate by parts. Thus

we have

$$\nabla_a j^a = 0 \tag{2.3.6}$$

as an operator equation. This is Noether's theorem, which is developed further in exercise 2.5.

To derive the Ward identity, let $\rho(\sigma)$ be 1 in some region R and 0 outside R. Also, include in the path integral some general local operator $\mathscr{A}(\sigma_0)$ at a point σ_0 inside R, and the usual arbitrary insertions '...' outside. Proceeding as above we obtain the operator relation

$$\delta\mathscr{A}(\sigma_0) + \frac{\epsilon}{2\pi i}\int_R d^d\sigma\, g^{1/2}\nabla_a j^a(\sigma)\mathscr{A}(\sigma_0) = 0\ . \tag{2.3.7}$$

Equivalently,

$$\nabla_a j^a(\sigma)\mathscr{A}(\sigma_0) = g^{-1/2}\delta^d(\sigma-\sigma_0)\frac{2\pi}{i\epsilon}\delta\mathscr{A}(\sigma_0) + \text{total } \sigma\text{-derivative}\ . \tag{2.3.8}$$

The divergence theorem gives

$$\int_{\partial R} dA\, n_a j^a \mathscr{A}(\sigma_0) = \frac{2\pi}{i\epsilon}\delta\mathscr{A}(\sigma_0) \tag{2.3.9}$$

with dA the area element and n^a the outward normal. This relates the integral of the current around the operator to the variation of the operator. Going to two flat dimensions this becomes

$$\oint_{\partial R}(j dz - \tilde{\jmath} d\bar{z})\mathscr{A}(z_0,\bar{z}_0) = \frac{2\pi}{\epsilon}\delta\mathscr{A}(z_0,\bar{z}_0)\ . \tag{2.3.10}$$

Again we drop indices, $j \equiv j_z$, $\tilde{\jmath} \equiv j_{\bar{z}}$; notice that on a current the omitted indices are implicitly lower. We use a tilde rather than a bar on $\tilde{\jmath}$ because this is *not* the adjoint of j. The Minkowski density j_0 is in general Hermitean, so the Euclidean j_2 with an extra factor of i is anti-Hermitean, and $(j_z)^\dagger = \frac{1}{2}(j_1 - ij_2)^\dagger = j_z$.

It is important that Noether's theorem and the Ward identity are local properties, which do not depend on whatever boundary conditions we might have far away, nor even on whether these are invariant under the symmetry. In particular, since the function $\rho(\sigma)$ is nonzero only inside R, the symmetry transformation need only be defined there.

In conformally invariant theories it is usually the case that j_z is holomorphic and $j_{\bar{z}}$ antiholomorphic (except for singularities at the other fields). In this case the currents $(j_z, 0)$ and $(0, j_{\bar{z}})$ are separately conserved. The integral (2.3.10) then picks out the residues in the OPE,

$$\mathrm{Res}_{z\to z_0}\, j(z)\mathscr{A}(z_0,\bar{z}_0) + \overline{\mathrm{Res}}_{\bar{z}\to\bar{z}_0}\,\tilde{\jmath}(\bar{z})\mathscr{A}(z_0,\bar{z}_0) = \frac{1}{i\epsilon}\delta\mathscr{A}(z_0,\bar{z}_0)\ . \tag{2.3.11}$$

Here 'Res' and '$\overline{\mathrm{Res}}$' are the coefficients of $(z-z_0)^{-1}$ and $(\bar{z}-\bar{z}_0)^{-1}$ respectively. This form of the Ward identity is particularly convenient.

The world-sheet current was defined with an extra factor of $2\pi i$ relative to the usual definition in field theory in order to make this OPE simple.

As an example, return to the free massless scalar and consider the *spacetime* translation $\delta X^\mu = \epsilon a^\mu$. Under $\delta X^\mu(\sigma) = \epsilon\rho(\sigma)a^\mu$,

$$\delta S = \frac{\epsilon a_\mu}{2\pi\alpha'}\int d^2\sigma\, \partial^a X^\mu \partial_a \rho\ . \tag{2.3.12}$$

This is of the claimed form (2.3.5) with Noether current $a_\mu j_a^\mu$, where

$$j_a^\mu = \frac{i}{\alpha'}\partial_a X^\mu\ . \tag{2.3.13}$$

The components are holomorphic and antiholomorphic as expected. For the OPE of this current with the exponential operator one finds

$$j^\mu(z)\ :e^{ik\cdot X(0,0)}: \sim \frac{k^\mu}{2z}\ :e^{ik\cdot X(0,0)}:\ , \tag{2.3.14a}$$

$$\tilde{\jmath}^\mu(\bar{z})\ :e^{ik\cdot X(0,0)}: \sim \frac{k^\mu}{2\bar{z}}\ :e^{ik\cdot X(0,0)}:\ , \tag{2.3.14b}$$

from terms with a single contraction. This OPE is in agreement with the general identity (2.3.11).

Another example is the *world-sheet* translation $\delta\sigma^a = \epsilon v^a$, under which $\delta X^\mu = -\epsilon v^a \partial_a X^\mu$. The Noether current is

$$j_a = i v^b T_{ab}\ , \tag{2.3.15a}$$

$$T_{ab} = -\frac{1}{\alpha'}:\left(\partial_a X^\mu \partial_b X_\mu - \frac{1}{2}\delta_{ab}\partial_c X^\mu \partial^c X_\mu\right):\ . \tag{2.3.15b}$$

Here T_{ab} is the world-sheet energy-momentum tensor.[2]

2.4 Conformal invariance

The energy-momentum tensor (2.3.15b) is traceless, $T_a^a = 0$. In complex coordinates this is

$$T_{z\bar{z}} = 0\ . \tag{2.4.1}$$

The conservation $\partial^a T_{ab} = 0$ then implies that in any theory with $T_a^a = 0$,

$$\bar{\partial}T_{zz} = \partial T_{\bar{z}\bar{z}} = 0\ . \tag{2.4.2}$$

Thus

$$T(z) \equiv T_{zz}(z)\ , \quad \tilde{T}(\bar{z}) \equiv T_{\bar{z}\bar{z}}(\bar{z}) \tag{2.4.3}$$

[2] In T_{ab} we have used normal ordering to define the product of operators at a point. The only possible ambiguity introduced by the renormalization is a constant times δ_{ab}, from the subtraction. Adding such a constant gives a different energy-momentum tensor which is also conserved. We choose to focus on the tensor (2.3.15b), for reasons that will be explained more fully in the next chapter.

are respectively holomorphic and antiholomorphic. For the free massless scalar,

$$T(z) = -\frac{1}{\alpha'} :\partial X^\mu \partial X_\mu: , \quad \tilde{T}(\bar{z}) = -\frac{1}{\alpha'} :\bar{\partial} X^\mu \bar{\partial} X_\mu: , \tag{2.4.4}$$

which are indeed holomorphic and antiholomorphic as a consequence of the equation of motion.

The tracelessness of T_{ab} implies a much larger symmetry. The currents

$$j(z) = iv(z)T(z) , \quad \tilde{j}(\bar{z}) = iv(z)^* \tilde{T}(\bar{z}) \tag{2.4.5}$$

are conserved for any holomorphic $v(z)$. For the free scalar theory, one finds the OPE

$$T(z)X^\mu(0) \sim \frac{1}{z} \partial X^\mu(0) , \quad \tilde{T}(\bar{z})X^\mu(0) \sim \frac{1}{\bar{z}} \bar{\partial} X^\mu(0) . \tag{2.4.6}$$

The Ward identity then gives the transformation

$$\delta X^\mu = -\epsilon v(z) \partial X^\mu - \epsilon v(z)^* \bar{\partial} X^\mu . \tag{2.4.7}$$

This is an infinitesimal coordinate transformation $z' = z + \epsilon v(z)$. The finite transformation is

$$X'^\mu(z', \bar{z}') = X^\mu(z, \bar{z}) , \quad z' = f(z) \tag{2.4.8}$$

for any holomorphic $f(z)$. This is known as a *conformal transformation.*

The conformal symmetry we have found should not be confused with the diff invariance of general relativity. We are in the flat space theory, with no independent metric field to vary, so the transformation $z \to z'$ actually changes the distances between points. We would not ordinarily have such an invariance; it is a nontrivial statement about the dynamics. For the scalar action (2.1.10), the conformal transformation of ∂ and $\bar{\partial}$ just balances that of d^2z. A mass term $m^2 X^\mu X_\mu$ would not be invariant. Obviously there will in the end be a close relation with the diff×Weyl symmetry of the Polyakov string, but we leave that for the next chapter.

Consider the special case

$$z' = \zeta z \tag{2.4.9}$$

for complex ζ. The phase of ζ is a rotation of the system, while its magnitude is a rescaling of the size of the system. Such a scale invariance has occasionally been considered as an approximate symmetry in particle physics, and statistical systems at a critical point are described by scale-invariant field theories.

To get some insight into the general conformal transformation, consider its effect on infinitesimal distances $ds^2 = d\sigma^a d\sigma^a = dz d\bar{z}$. Conformal transformations rescale this by a *position-dependent* factor,

$$ds'^2 = dz' d\bar{z}' = \frac{\partial z'}{\partial z} \frac{\partial \bar{z}'}{\partial \bar{z}} dz d\bar{z} . \tag{2.4.10}$$

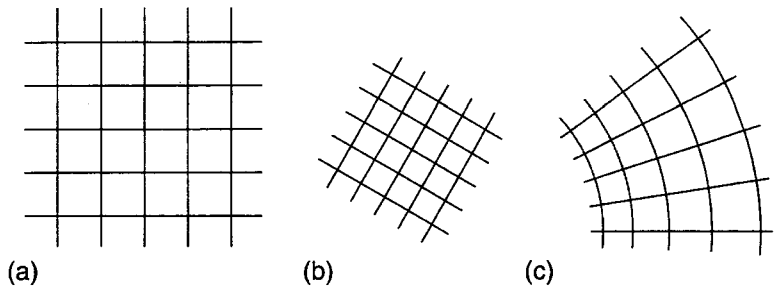


Fig. 2.2. (a) Two-dimensional region. (b) Effect of the special conformal transformation (2.4.9). (c) Effect of a more general conformal transformation.

Thus, as indicated in figure 2.2, a conformal transformation takes infinitesimal squares into infinitesimal squares, but rescales them by a position-dependent factor. An antiholomorphic function $z' = f(z)^*$ has the same property, but changes the orientation. Most systems that are invariant under the rigid scaling (2.4.9) are actually invariant under the much larger conformal symmetry. A theory with this invariance is called a *conformal field theory (CFT).* Conformal invariance in more than two dimensions is developed in exercise 2.6.

Conformal invariance and the OPE

Conformal invariance puts strong constraints on the form of the OPE, and in particular on the OPEs of the energy-momentum tensor. Consider the OPE of T with the general operator $\mathscr{A}$. Because $T(z)$ and $\tilde{T}(\bar{z})$ are (anti)holomorphic except at insertions, the corresponding coefficient functions must also have this property. The OPE of T with a general $\mathscr{A}$ is therefore a *Laurent expansion,* in integer but possibly negative powers of z. Further, all the singular terms are determined by the conformal transformation of $\mathscr{A}$. To see this, let us write a general expansion of the singular terms,

$$T(z)\mathscr{A}(0,0) \sim \sum_{n=0}^{\infty} \frac{1}{z^{n+1}} \mathscr{A}^{(n)}(0,0) \tag{2.4.11}$$

and similarly for $\tilde{T}$; the operator coefficients $\mathscr{A}^{(n)}$ remain to be determined. Under an infinitesimal conformal transformation $z' = z + \epsilon v(z)$, a single pole in $v(z)T(z)\mathscr{A}(0,0)$ arises when the z^{-n-1} term of the $T\mathscr{A}$ OPE multiplies the term of order z^n in $v(z)$. Thus, the Ward identity in the

form (2.3.11) implies that

$$\delta\mathscr{A}(z,\bar{z}) = -\epsilon\sum_{n=0}^{\infty}\frac{1}{n!}\left[\partial^n v(z)\mathscr{A}^{(n)}(z,\bar{z}) + \bar{\partial}^n v(z)^*\tilde{\mathscr{A}}^{(n)}(z,\bar{z})\right] . \qquad (2.4.12)$$

Thus the operators $\mathscr{A}^{(n)}$ are determined by the conformal transformation of $\mathscr{A}$.

It is convenient to take a basis of local operators that are eigenstates under rigid transformation (2.4.9),

$$\mathscr{A}'(z',\bar{z}') = \zeta^{-h}\bar{\zeta}^{-\tilde{h}}\mathscr{A}(z,\bar{z}) . \qquad (2.4.13)$$

The $(h,\tilde{h})$ are known as the *weights* of $\mathscr{A}$. The sum $h+\tilde{h}$ is the dimension of $\mathscr{A}$, determining its behavior under scaling, while $h-\tilde{h}$ is the spin, determining its behavior under rotations. The derivative ∂_z increases h by one, and the derivative $\partial_{\bar{z}}$ increases $\tilde{h}$ by one. The Ward identities for the transformation (2.4.13) and for the translation $\delta\mathscr{A} = -v^a\partial_a\mathscr{A}$ determine part of the OPE,

$$T(z)\mathscr{A}(0,0) = \ldots + \frac{h}{z^2}\mathscr{A}(0,0) + \frac{1}{z}\partial\mathscr{A}(0,0) + \ldots , \qquad (2.4.14)$$

and similarly for $\tilde{T}$.

An important special case is a *tensor operator* or *primary field*[3] $\mathscr{O}$, on which a general conformal transformation acts as

$$\mathscr{O}'(z',\bar{z}') = (\partial_z z')^{-h}(\partial_{\bar{z}}\bar{z}')^{-\tilde{h}}\mathscr{O}(z,\bar{z}) . \qquad (2.4.15)$$

The OPE (2.4.11) reduces to

$$T(z)\mathscr{O}(0,0) = \frac{h}{z^2}\mathscr{O}(0,0) + \frac{1}{z}\partial\mathscr{O}(0,0) + \ldots , \qquad (2.4.16)$$

the more singular terms in the general OPE (2.4.14) being absent.

Taking again the example of the free X^μ CFT, the weights of some typical operators are

$$\left.\begin{array}{llll} X^\mu & (0,0) , & \partial X^\mu & (1,0) , \\ \bar{\partial}X^\mu & (0,1) , & \partial^2 X^\mu & (2,0) , \\ :e^{ik\cdot X}: & \left(\frac{\alpha' k^2}{4},\frac{\alpha' k^2}{4}\right) . & & \end{array}\right\} \qquad (2.4.17)$$

[3] In quantum field theory, one usually distinguishes the fundamental fields (the variables of integration in the path integral) from more general operators which may be composite. Actually, this distinction is primarily useful in weakly coupled field theories in three or more dimensions. It is of little use in CFT, in particular because of equivalences between different field theories, and so the term *field* is used for any local operator.

All transform as tensors except $\partial^2 X^\mu$. More generally, an exponential times a general product of derivatives,

$$:\left(\prod_i \partial^{m_i} X^{\mu_i}\right)\left(\prod_j \bar{\partial}^{n_j} X^{\nu_j}\right)e^{ik\cdot X}: , \qquad (2.4.18)$$

has weight

$$\left(\frac{\alpha' k^2}{4} + \sum_i m_i , \frac{\alpha' k^2}{4} + \sum_j n_j\right) . \qquad (2.4.19)$$

For any pair of operators, applying rigid translations, scale transformations, and rotations to both sides of the OPE determines the z-dependence of the coefficient functions completely,

$$\mathscr{A}_i(z_1,\bar{z}_1)\mathscr{A}_j(z_2,\bar{z}_2) = \sum_k z_{12}^{h_k-h_i-h_j}\bar{z}_{12}^{\tilde{h}_k-\tilde{h}_i-\tilde{h}_j}c^k_{ij}\,\mathscr{A}_k(z_2,\bar{z}_2) , \qquad (2.4.20)$$

where the c^k_{ij} are now constants. In all cases of interest the weights appearing on the right-hand side of the OPE (2.4.20) are bounded below, and so the degree of singularity in the operator product is bounded. More general conformal transformations put further constraints on the OPE: they determine the OPEs of *all* fields in terms of those of the primary fields. We will develop this in chapter 15.

Notice that the conformal transformation properties of normal ordered products are not in general given by the naive transformations of the product. For example, the transformation law (2.4.7) for X^μ would naively imply

$$\delta e^{ik\cdot X} = -\epsilon v(z)\partial e^{ik\cdot X} - \epsilon v(z)^*\bar{\partial}e^{ik\cdot X} \quad \text{(naive)} \qquad (2.4.21)$$

for the exponential, making it a tensor of weight $(0,0)$. The modification is a quantum effect, due to the renormalization needed to define the product of operators at a point. Specifically it enters here because the subtraction $\ln|z_{12}|^2$ in : : makes explicit reference to the coordinate frame.

Conformal properties of the energy-momentum tensor

The OPE of the energy-momentum tensor with itself was obtained in eq. (2.2.11),

$$\begin{aligned} T(z)T(0) &= \frac{\eta^\mu_\mu}{2z^4} - \frac{2}{\alpha' z^2}:\partial X^\mu(z)\partial X_\mu(0): + :T(z)T(0): \\ &\sim \frac{D}{2z^4} + \frac{2}{z^2}T(0) + \frac{1}{z}\partial T(0) . \end{aligned} \qquad (2.4.22)$$

A similar result holds for $\tilde{T}$. The $T(z)\tilde{T}(\bar{z}')$ OPE must be nonsingular. It cannot have poles in $(z-z')$ because it is antiholomorphic in z' at nonzero separation; similarly it cannot have poles in $(\bar{z}-\bar{z}')$ because it is

holomorphic in z at nonzero separation.[4] The same holds for any OPE between a holomorphic and an antiholomorphic operator.

Thus T is *not* a tensor. Rather, the OPE (2.4.22) implies the transformation law

$$\delta T(z) = -\frac{D}{12}\partial_z^3 v(z) - 2\partial_z v(z) T(z) - v(z)\partial_z T(z) . \tag{2.4.23}$$

In a general CFT, $T(z)$ transforms as

$$\delta T(z) = -\frac{c}{12}\partial_z^3 v(z) - 2\partial_z v(z) T(z) - v(z)\partial_z T(z) , \tag{2.4.24}$$

with c a constant known as the *central charge*. The central charge of a free scalar is 1; for D free scalars it is D. The transformation (2.4.24) is the most general form that is linear in v, is consistent with the symmetry of the TT OPE, and has three lower z indices as required by rigid scale and rotation invariance. The scale, rotation, and translation symmetries determine the coefficients of the second and third terms. Further, by considering the commutator of two such transformations one can show that $\partial_a c = 0$, and it is a general result in quantum field theory that an operator that is independent of position must be a c-number.[5] The corresponding TT OPE is

$$T(z)T(0) \sim \frac{c}{2z^4} + \frac{2}{z^2}T(0) + \frac{1}{z}\partial T(0) . \tag{2.4.25}$$

The finite form of the transformation law (2.4.24) is

$$(\partial_z z')^2 T'(z') = T(z) - \frac{c}{12}\{z', z\} , \tag{2.4.26}$$

where $\{f, z\}$ denotes the *Schwarzian derivative*

$$\{f, z\} = \frac{2\partial_z^3 f \partial_z f - 3\partial_z^2 f \partial_z^2 f}{2\partial_z f \partial_z f} . \tag{2.4.27}$$

One can check this by verifying that it has the correct infinitesimal form and that it composes correctly under successive transformations (so that one can integrate the infinitesimal transformation). The corresponding forms hold for $\tilde{T}$, possibly with a different central charge $\tilde{c}$ in the general CFT.

The nontensor behavior of the energy-momentum tensor has a number of important physical consequences, as we will see. We should emphasize that 'nontensor' refers to conformal transformations. The energy-momentum tensor will have its usual tensor property under coordinate transformations.

[4] Unless otherwise stated OPEs hold only at *nonzero* separation, ignoring possible delta functions. For our applications the latter will not matter. Occasionally it is useful to include the delta functions, but in general these depend partly on definitions so one must be careful.

[5] The argument is this. Suppose that $\mathscr{A}(\sigma)$ is independent of position, and consider its equal time commutator with any other local operator $\mathscr{B}(\sigma')$. These commute at spacelike separation, by locality. Since $\mathscr{A}(\sigma)$ is actually independent of position it also commutes at zero separation. Therefore $\mathscr{A}(\sigma)$ commutes with all local operators, and so must be a c-number.

2.5 Free CFTs

In this section we discuss three families of free-field CFTs — the linear dilaton, bc, and $\beta\gamma$ theories. The bc theory is the one of most immediate interest, as it will appear in the next chapter when we gauge-fix the Polyakov string, but all have a variety of applications in string theory.

Linear dilaton CFT

This family of CFTs is based on the same action (2.1.10) but with energy-momentum tensor

$$T(z) = -\frac{1}{\alpha'} :\partial X^\mu \partial X_\mu: + V_\mu \partial^2 X^\mu , \tag{2.5.1a}$$

$$\tilde{T}(\bar{z}) = -\frac{1}{\alpha'} :\bar{\partial} X^\mu \bar{\partial} X_\mu: + V_\mu \bar{\partial}^2 X^\mu , \tag{2.5.1b}$$

where V_μ is some fixed D-vector. Working out the TT OPE, one finds that it is of the standard form (2.4.25), but with central charge

$$c = \tilde{c} = D + 6\alpha' V_\mu V^\mu . \tag{2.5.2}$$

The TX^μ OPE and the Ward identity (2.4.11) imply the conformal transformation

$$\delta X^\mu = -\epsilon v \partial X^\mu - \epsilon v^* \bar{\partial} X^\mu - \frac{\epsilon}{2}\alpha' V^\mu [\partial v + (\partial v)^*] . \tag{2.5.3}$$

This is a *different* conformal symmetry of the same action. The field X^μ no longer transforms as a tensor, its variation now having an inhomogeneous piece. Incidentally, the free massless scalar in two dimensions has a remarkably large amount of symmetry — much more than we will have occasion to mention.

The energy-momentum tensor plays a special role in string theory — in particular, it tells us how to couple to a curved metric — so different values of V^μ are to be regarded as different CFTs. The vector V^μ picks out a direction in spacetime. This CFT is therefore not Lorentz-invariant and not of immediate interest to us. We will see one physical interpretation of this *linear dilaton* CFT in section 3.7, and encounter it in some technical applications later.

A different variation of the free scalar CFT is to take some of the X^μ to be periodic; we will take this up in chapter 8.

bc CFT

The second family of CFTs has *anticommuting* fields b and c with action

$$S = \frac{1}{2\pi}\int d^2z\, b\bar{\partial}c\,. \tag{2.5.4}$$

This is conformally invariant for b and c transforming as tensors of weights $(h_b, 0)$ and $(h_c, 0)$ such that

$$h_b = \lambda\,, \quad h_c = 1-\lambda \tag{2.5.5}$$

for any given constant λ. Thus we have another family of CFTs (which is secretly the same as the linear dilaton family, as we will learn in chapter 10). The operator equations of motion, obtained by the same method (2.1.15), (2.1.18) as before, are

$$\bar{\partial}c(z) = \bar{\partial}b(z) = 0\,, \tag{2.5.6a}$$

$$\bar{\partial}b(z)c(0) = 2\pi\delta^2(z,\bar{z})\,. \tag{2.5.6b}$$

The bb and cc OPEs satisfy the equation of motion without source. The normal ordered bc product is

$$:b(z_1)c(z_2): = b(z_1)c(z_2) - \frac{1}{z_{12}}\,. \tag{2.5.7}$$

This satisfies the naive equations of motion as a consequence of

$$\bar{\partial}\frac{1}{z} = \partial\frac{1}{\bar{z}} = 2\pi\delta^2(z,\bar{z})\,, \tag{2.5.8}$$

which can be verified by integrating over a region containing the origin and integrating the derivative by parts. Normal ordering of a general product of fields is combinatorially the same as for the X^μ CFT, a sum over contractions or subtractions. One must be careful because b and c are anticommuting so that interchange of fields flips the sign: one should anticommute the fields being paired until they are next to each other before replacing them with the subtraction (2.5.7).

The operator products are

$$b(z_1)c(z_2) \sim \frac{1}{z_{12}}\,, \quad c(z_1)b(z_2) \sim \frac{1}{z_{12}}\,, \tag{2.5.9}$$

where in the second OPE there have been two sign flips, one from anticommutation and one from $z_1 \leftrightarrow z_2$. Other operator products are nonsingular:

$$b(z_1)b(z_2) = O(z_{12})\,, \quad c(z_1)c(z_2) = O(z_{12})\,. \tag{2.5.10}$$

These are not only holomorphic but have a zero due to antisymmetry.

Noether's theorem gives the energy-momentum tensor

$$T(z) = :(\partial b)c: - \lambda\partial(:bc:)\,, \tag{2.5.11a}$$

$$\tilde{T}(\bar{z}) = 0\,. \tag{2.5.11b}$$

One can also verify (2.5.11) by working out the OPE of T with b and c; it has the standard tensor form (2.4.16) with the given weights. The TT OPE is of standard form (2.4.25) with

$$c = -3(2\lambda-1)^2 + 1\,, \quad \tilde{c} = 0\,. \tag{2.5.12}$$

This is a purely holomorphic CFT, and is an example where $c \neq \tilde{c}$. There is of course a corresponding antiholomorphic theory

$$S = \frac{1}{2\pi}\int d^2z\, \tilde{b}\partial\tilde{c}\,, \tag{2.5.13}$$

which is the same as the above with $z \leftrightarrow \bar{z}$.

The bc theory has a *ghost number* symmetry $\delta b = -i\epsilon b$, $\delta c = i\epsilon c$. The corresponding Noether current is

$$j = -:bc:\,. \tag{2.5.14}$$

Again the components are separately holomorphic and antiholomorphic, the latter vanishing. When there are both holomorphic and antiholomorphic bc fields, the ghost numbers are separately conserved.

This current is not a tensor,

$$T(z)j(0) \sim \frac{1-2\lambda}{z^3} + \frac{1}{z^2}j(0) + \frac{1}{z}\partial j(0)\,. \tag{2.5.15}$$

This implies the transformation law

$$\delta j = -v\partial j - j\partial v + \frac{2\lambda-1}{2}\partial^2 v\,, \tag{2.5.16}$$

whose finite form is

$$(\partial_z z')j_{z'}(z') = j_z(z) + \frac{2\lambda-1}{2}\frac{\partial_z^2 z'}{\partial_z z'}\,. \tag{2.5.17}$$

The one case where b and c have equal weight is $h_b = h_c = \frac{1}{2}$, for which the central charge $c = 1$. Here we will often use the notation $b \to \psi$, $c \to \bar{\psi}$. For this case the bc CFT can be split in two in a conformally invariant way,

$$\psi = 2^{-1/2}(\psi_1 + i\psi_2)\,, \quad \bar{\psi} = 2^{-1/2}(\psi_1 - i\psi_2)\,, \tag{2.5.18a}$$

$$S = \frac{1}{4\pi}\int d^2z\, \left(\psi_1\bar{\partial}\psi_1 + \psi_2\bar{\partial}\psi_2\right)\,, \tag{2.5.18b}$$

$$T = -\frac{1}{2}\psi_1\partial\psi_1 - \frac{1}{2}\psi_2\partial\psi_2\,. \tag{2.5.18c}$$

Each ψ theory has central charge $\frac{1}{2}$.

The bc theory for $\lambda = 2$, weights $(h_b, h_c) = (2, -1)$, will arise in the next chapter as the Faddeev–Popov ghosts from gauge-fixing the Polyakov string. The ψ theory will appear extensively in the more general string theories of volume two.

$\beta\gamma$ CFT

The third family of CFTs is much like the bc theory but with *commuting* fields; β is an $(h_\beta, 0)$ tensor and γ an $(h_\gamma, 0)$ tensor, where

$$h_\beta = \lambda\,, \quad h_\gamma = 1 - \lambda\,. \tag{2.5.19}$$

The action is

$$S = \frac{1}{2\pi}\int d^2z\, \beta\bar{\partial}\gamma\,. \tag{2.5.20}$$

These fields are again holomorphic by the equations of motion,

$$\bar{\partial}\gamma(z) = \bar{\partial}\beta(z) = 0\,. \tag{2.5.21}$$

The equations of motion and operator products are derived in the standard way. Because the statistics are changed, some signs in operator products are different,

$$\beta(z_1)\gamma(z_2) \sim -\frac{1}{z_{12}}\,, \quad \gamma(z_1)\beta(z_2) \sim \frac{1}{z_{12}}\,. \tag{2.5.22}$$

The energy-momentum tensor is

$$T = :(\partial\beta)\gamma: - \lambda\partial(:\beta\gamma:)\,, \tag{2.5.23a}$$

$$\tilde{T} = 0\,. \tag{2.5.23b}$$

The central charge is simply reversed in sign,

$$c = 3(2\lambda - 1)^2 - 1\,, \quad \tilde{c} = 0\,. \tag{2.5.24}$$

The $\beta\gamma$ theory for $\lambda = \frac{3}{2}$, weights $(h_\beta, h_\gamma) = (\frac{3}{2}, -\frac{1}{2})$, will arise in chapter 10 as the Faddeev–Popov ghosts from gauge-fixing the superstring.

2.6 The Virasoro algebra

Thus far in this chapter we have studied local properties of the two-dimensional field theory. We now are interested in the spectrum of the theory. The spatial coordinate will be periodic, as in the closed string, or bounded, as in the open string.

For the periodic case let

$$\sigma^1 \sim \sigma^1 + 2\pi\,. \tag{2.6.1}$$

Let the Euclidean time coordinate run

$$-\infty < \sigma^2 < \infty \tag{2.6.2}$$

so that the two dimensions form an infinite cylinder. It is again useful to form a complex coordinate, and there are two natural choices. The first is

$$w = \sigma^1 + i\sigma^2\,, \tag{2.6.3}$$

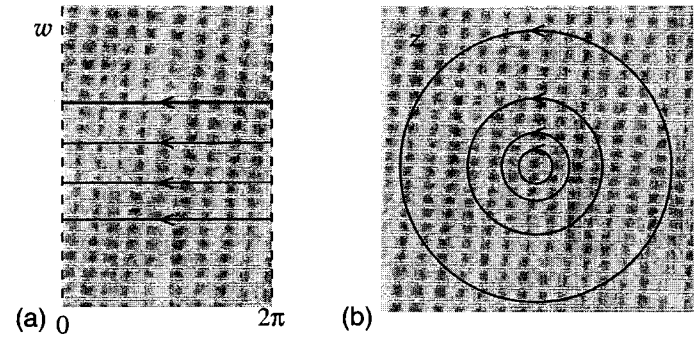


Fig. 2.3. Closed string coordinates. (a) Equal time contours in the w-plane. The dashed lines are identified. (b) The same contours in the z-plane.

so that $w \sim w + 2\pi$. The second is

$$z = \exp(-iw) = \exp(-i\sigma^1 + \sigma^2)\,. \tag{2.6.4}$$

These two coordinate systems are shown in figure 2.3. In terms of the w coordinate, time corresponds to translation of $\sigma^2 = \mathrm{Im}\, w$. In terms of z, time runs radially, the origin being the infinite past. These coordinates are related by a conformal transformation. The w coordinate is natural for the canonical interpretation of the theory, but the z coordinate is also quite useful and most expressions are written in this frame.

For a holomorphic or antiholomorphic operator we can make a Laurent expansion,

$$T_{zz}(z) = \sum_{m=-\infty}^{\infty} \frac{L_m}{z^{m+2}}\,, \quad \tilde{T}_{\bar{z}\bar{z}}(\bar{z}) = \sum_{m=-\infty}^{\infty} \frac{\tilde{L}_m}{\bar{z}^{m+2}}\,. \tag{2.6.5}$$

The Laurent coefficients, known as the *Virasoro generators*, are given by the contour integrals

$$L_m = \oint_C \frac{dz}{2\pi i z}\, z^{m+2} T_{zz}(z)\,, \tag{2.6.6}$$

where C is any contour encircling the origin counterclockwise. The Laurent expansion is just the same as an ordinary Fourier transformation in the w frame at time $\sigma^2 = 0$:

$$T_{ww}(w) = -\sum_{m=-\infty}^{\infty} \exp(im\sigma^1 - m\sigma^2) T_m\,, \tag{2.6.7a}$$

$$T_{\bar{w}\bar{w}}(\bar{w}) = -\sum_{m=-\infty}^{\infty} \exp(-im\sigma^1 - m\sigma^2) \tilde{T}_m\,, \tag{2.6.7b}$$

where

$$T_m = L_m - \delta_{m,0}\frac{c}{24}\,, \quad \tilde{T}_m = \tilde{L}_m - \delta_{m,0}\frac{\tilde{c}}{24}\,. \tag{2.6.8}$$

The additive shift of T_0 is from the nontensor transformation (2.4.26),

$$T_{ww} = (\partial_w z)^2 T_{zz} + \frac{c}{24}\,. \tag{2.6.9}$$

The Hamiltonian H of time translation in the $w = \sigma^1 + i\sigma^2$ frame is

$$H = \int_0^{2\pi} \frac{d\sigma^1}{2\pi} T_{22} = L_0 + \tilde{L}_0 - \frac{c+\tilde{c}}{24}\,. \tag{2.6.10}$$

Notice that the $+2$s from the Laurent expansions (2.6.5) have canceled in the Fourier expansions (2.6.7) due to the conformal transformation of T. Similarly the Laurent expansion for a holomorphic field of weight h would include $+h$ in the exponent.

Cutting open the path integral on circles of constant time $\operatorname{Im} w = \ln|z|$, the Virasoro generators become operators in the ordinary sense. (This idea of cutting open a path integral is developed in the appendix.) By holomorphicity the integrals (2.6.6) are independent of C and so in particular are invariant under time translation (radial rescaling). That is, they are conserved charges, the charges associated with the conformal transformations.

It is an important fact that the OPE of currents determines the algebra of the corresponding charges. Consider general charges Q_i, $i = 1, 2$, given as contour integrals of holomorphic currents,

$$Q_i\{C\} = \oint_C \frac{dz}{2\pi i}\, j_i\,. \tag{2.6.11}$$

Consider the combination

$$Q_1\{C_1\}Q_2\{C_2\} - Q_1\{C_3\}Q_2\{C_2\}\,, \tag{2.6.12}$$

where the contours are shown in figure 2.4(a). The order in which the factors are written is irrelevant, as these are just variables of integration in a path integral (unless both charges are anticommuting, in which case there is an additional sign and all the commutators become anticommutators). As discussed in the appendix, when we slice open the path integral to make an operator interpretation, what determines the operator ordering is the time ordering, which here is $t_1 > t_2 > t_3$. The path integral with the combination (2.6.12) thus corresponds to a matrix element of

$$\hat{Q}_1\hat{Q}_2 - \hat{Q}_2\hat{Q}_1 \equiv [\hat{Q}_1, \hat{Q}_2]\,. \tag{2.6.13}$$

Now, for a given point z_2 on the contour C_2, we can deform the difference of the C_1 and C_3 contours as shown in figure 2.4(b), so the commutator

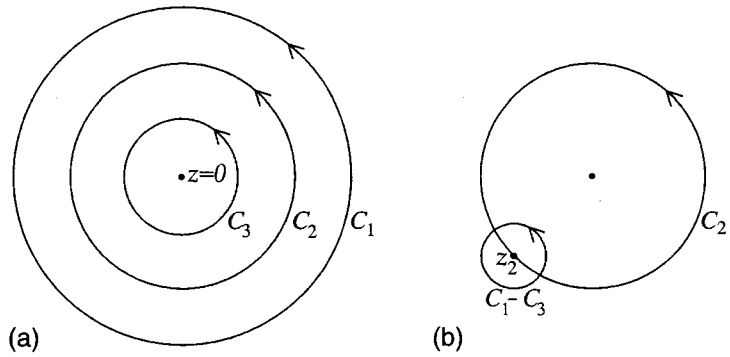


Fig. 2.4. (a) Contours centered on $z = 0$. (b) For given z_2 on contour C_2, contour $C_1 - C_3$ is contracted.

is given by the residue of the OPE,

$$[Q_1, Q_2]\{C_2\} = \oint_{C_2} \frac{dz_2}{2\pi i} \operatorname{Res}_{z_1 \to z_2} j_1(z_1) j_2(z_2)\,. \tag{2.6.14}$$

This contour argument allows us to pass back and forth between OPEs and commutation relations. Let us emphasize this: for conserved currents, knowing the singular terms in the OPE is equivalent to knowing the commutator algebra of the corresponding charges. The calculation of figure 2.4 also applies with the conserved charge $Q_2\{C_2\}$ replaced by any operator,

$$[Q, \mathscr{A}(z_2, \bar{z}_2)] = \operatorname{Res}_{z_1 \to z_2} j(z_1)\mathscr{A}(z_2, \bar{z}_2) = \frac{1}{i\epsilon}\delta\mathscr{A}(z_2, \bar{z}_2)\,. \tag{2.6.15}$$

This is just the familiar statement that a charge Q generates the corresponding transformation δ. Similarly for the contour integral of an antiholomorphic current

$$\tilde{Q}\{C\} = -\oint_C \frac{d\bar{z}}{2\pi i}\, \tilde{j}_i\,, \tag{2.6.16}$$

the Ward identity and contour argument imply

$$[\tilde{Q}, \mathscr{A}(z_2, \bar{z}_2)] = \operatorname{Res}_{\bar{z}_1 \to \bar{z}_2} \tilde{j}(\bar{z}_1)\mathscr{A}(z_2, \bar{z}_2) = \frac{1}{i\epsilon}\delta\mathscr{A}(z_2, \bar{z}_2)\,. \tag{2.6.17}$$

Apply this to the Virasoro generators (2.6.6), where $j_m(z) = z^{m+1}T(z)$:

$$\operatorname{Res}_{z_1 \to z_2} z_1^{m+1} T(z_1)\, z_2^{n+1} T(z_2)$$
$$= \operatorname{Res}_{z_1 \to z_2} z_1^{m+1} z_2^{n+1} \left(\frac{c}{2z_{12}^4} + \frac{2}{z_{12}^2} T(z_2) + \frac{1}{z_{12}} \partial T(z_2) \right)$$

$$= \frac{c}{12}(\partial^3 z_2^{m+1}) z_2^{n+1} + 2(\partial z_2^{m+1}) z_2^{n+1} T(z_2) + z_2^{m+n+2} \partial T(z_2)$$
$$= \frac{c}{12}(m^3 - m) z_2^{m+n-1} + (m-n) z_2^{m+n+1} T(z_2) + \text{total derivative} . \tag{2.6.18}$$

The z_2 contour integral of the right-hand side then gives the *Virasoro algebra*,

$$[L_m, L_n] = (m-n) L_{m+n} + \frac{c}{12}(m^3 - m)\delta_{m,-n} . \tag{2.6.19}$$

The $\tilde{L}_m$ satisfy the same algebra with central charge $\tilde{c}$.

Any CFT thus has an infinite set of conserved charges, the Virasoro generators, which act in the Hilbert space and which satisfy the algebra (2.6.19). Let us for now notice just a few simple properties. Generally we work with eigenstates of L_0 and $\tilde{L}_0$. The generator L_0 satisfies

$$[L_0, L_n] = -n L_n . \tag{2.6.20}$$

If $|\psi\rangle$ is an eigenstate of L_0 with eigenvalue h, then

$$L_0 L_n |\psi\rangle = L_n (L_0 - n)|\psi\rangle = (h-n) L_n |\psi\rangle , \tag{2.6.21}$$

so that $L_n|\psi\rangle$ is an eigenstate with eigenvalue $h-n$. The generators with $n<0$ raise the L_0 eigenvalue and those with $n>0$ lower it.

The three generators L_0 and $L_{\pm 1}$ form a closed algebra without central charge,

$$[L_0, L_1] = -L_1 , \quad [L_0, L_{-1}] = L_{-1} , \quad [L_1, L_{-1}] = 2L_0 . \tag{2.6.22}$$

This is the algebra $SL(2,\mathbf{R})$, which differs from $SU(2)$ by signs. For the Laurent coefficients of a holomorphic tensor field $\mathcal{O}$ of weight $(h,0)$,

$$\mathcal{O}(z) = \sum_{m=-\infty}^{\infty} \frac{\mathcal{O}_m}{z^{m+h}} , \tag{2.6.23}$$

one finds from the OPE (2.4.16) the commutator

$$[L_m, \mathcal{O}_n] = [(h-1)m - n]\mathcal{O}_{m+n} . \tag{2.6.24}$$

Again modes with $n>0$ reduce L_0, while modes with $n<0$ increase it.

In the open string, let

$$0 \le \mathrm{Re}\, w \le \pi \quad \Leftrightarrow \quad \mathrm{Im}\, z \ge 0 , \tag{2.6.25}$$

where $z = -\exp(-iw)$. These coordinate regions are shown in figure 2.5. At a boundary, the energy-momentum tensor satisfies

$$T_{ab} n^a t^b = 0 , \tag{2.6.26}$$

where n^a and t^b are again normal and tangent vectors. To see this, consider a coordinate system in which the boundary is straight. The presence of

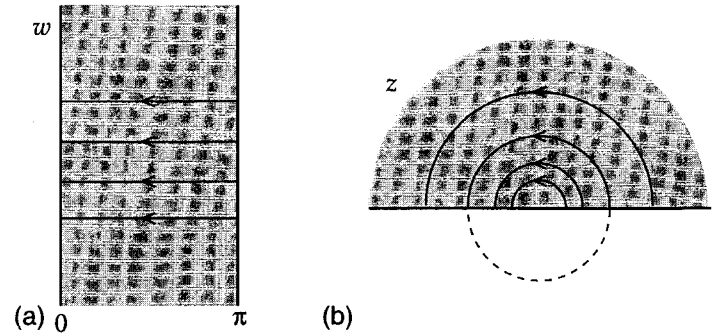


Fig. 2.5. Open string coordinates. (a) Equal time contours in the w-plane. (b) The same contours in the z-plane. The dashed line shows the extension of one contour, as used in the doubling trick.

the boundary breaks translation invariance in the normal direction but not the tangential, so that the current $T_{ab}t^b$ is still conserved. Then the boundary condition (2.6.26) is just the statement that the flow of this current out of the boundary is zero. In the present case, this becomes

$$T_{ww} = T_{\bar{w}\bar{w}} , \ \mathrm{Re}\, w = 0, \pi \quad \Leftrightarrow \quad T_{zz} = T_{\bar{z}\bar{z}} , \ \mathrm{Im}\, z = 0 . \tag{2.6.27}$$

It is convenient to use the *doubling trick*. Define T_{zz} in the lower half-z-plane as the value of $T_{\bar{z}\bar{z}}$ at its image in the upper half-z-plane, $z' = \bar{z}$:

$$T_{zz}(z) \equiv T_{\bar{z}\bar{z}}(\bar{z}') , \quad \mathrm{Im} z < 0 . \tag{2.6.28}$$

The equation of motion and boundary condition are then summarized by the statement that T_{zz} is holomorphic in the *whole* complex plane. There is only one set of Virasoro generators, because the boundary condition couples T and $\tilde{T}$,

$$L_m = \frac{1}{2\pi i}\int_C \left(dz\, z^{m+1} T_{zz} - d\bar{z}\, \bar{z}^{m+1} T_{\bar{z}\bar{z}} \right)$$
$$= \frac{1}{2\pi i}\oint dz\, z^{m+1} T_{zz}(z) . \tag{2.6.29}$$

In the first line, the contour C is a semi-circle centered on the origin; in the second line, we have used the doubling trick to write L_m in terms of a closed contour. Again, these satisfy the Virasoro algebra

$$[L_m, L_n] = (m-n) L_{m+n} + \frac{c}{12}(m^3 - m)\delta_{m,-n} . \tag{2.6.30}$$

2.7 Mode expansions

Free scalars

In free field theory, the fields break up into harmonic oscillators, and the spectrum and energy-momentum tensor can be given in terms of the modes. We start with the closed string. In the X^μ theory, ∂X and $\bar\partial X$ are (anti)holomorphic and so have Laurent expansions like that for T,

$$\partial X^\mu(z) = -i\left(\frac{\alpha'}{2}\right)^{1/2}\sum_{m=-\infty}^{\infty}\frac{\alpha_m^\mu}{z^{m+1}}\,, \quad \bar\partial X^\mu(\bar z) = -i\left(\frac{\alpha'}{2}\right)^{1/2}\sum_{m=-\infty}^{\infty}\frac{\tilde\alpha_m^\mu}{\bar z^{m+1}}\,. \tag{2.7.1}$$

Equivalently,

$$\alpha_m^\mu = \left(\frac{2}{\alpha'}\right)^{1/2}\oint\frac{dz}{2\pi}\,z^m\partial X^\mu(z)\,, \tag{2.7.2a}$$

$$\tilde\alpha_m^\mu = -\left(\frac{2}{\alpha'}\right)^{1/2}\oint\frac{d\bar z}{2\pi}\,\bar z^m\bar\partial X^\mu(\bar z)\,. \tag{2.7.2b}$$

Single-valuedness of X^μ implies that $\alpha_0^\mu = \tilde\alpha_0^\mu$. Moreover, the Noether current for spacetime translations is $i\partial_a X^\mu/\alpha'$, so the spacetime momentum is

$$p^\mu = \frac{1}{2\pi i}\oint_C(dz\,j^\mu - d\bar z\,\tilde\jmath^\mu) = \left(\frac{2}{\alpha'}\right)^{1/2}\alpha_0^\mu = \left(\frac{2}{\alpha'}\right)^{1/2}\tilde\alpha_0^\mu\,. \tag{2.7.3}$$

Integrating the expansions (2.7.1) gives

$$X^\mu(z,\bar z) = x^\mu - i\frac{\alpha'}{2}p^\mu\ln|z|^2 + i\left(\frac{\alpha'}{2}\right)^{1/2}\sum_{\substack{m=-\infty\\ m\neq 0}}^{\infty}\frac{1}{m}\left(\frac{\alpha_m^\mu}{z^m}+\frac{\tilde\alpha_m^\mu}{\bar z^m}\right). \tag{2.7.4}$$

Either from standard canonical commutation, or from the contour argument and the XX OPE, one derives

$$[\alpha_m^\mu,\alpha_n^\nu] = [\tilde\alpha_m^\mu,\tilde\alpha_n^\nu] = m\delta_{m,-n}\eta^{\mu\nu}\,, \tag{2.7.5a}$$

$$[x^\mu,p^\nu] = i\eta^{\mu\nu}, \tag{2.7.5b}$$

with other commutators vanishing. The spectrum is given by starting with a state $|0;k\rangle$ that has momentum k^μ and is annihilated by all of the lowering modes, α_n^μ for $n>0$, and acting in all possible ways with the raising ($n<0$) modes.

We now wish to expand the Virasoro generators in terms of the mode operators. Insert the Laurent expansion for X^μ into the energy-momentum tensor (2.4.4) and collect terms with a given power of z, giving

$$L_m \sim \frac{1}{2}\sum_{n=-\infty}^{\infty}\alpha_{m-n}^\mu\alpha_{\mu n}\,. \tag{2.7.6}$$

The $\sim$ indicates that we have ignored the ordering of operators. For $m\neq 0$, the expansion (2.7.6) is well defined and correct as it stands — the mode operators in each term commute and so the ordering does not matter. For $m=0$, put the lowering operators on the right and introduce a normal ordering constant,

$$L_0 = \frac{\alpha' p^2}{4} + \sum_{n=1}^{\infty}(\alpha_{-n}^\mu\alpha_{\mu n}) + a^X\,. \tag{2.7.7}$$

We encountered this same issue in section 1.3, where we treated it in a heuristic way. Now the left-hand side has a finite and unambiguous definition in terms of the Laurent coefficients of the : :-ordered energy-momentum tensor, so the normal ordering constant is finite and calculable. There are several ways to determine it. The simplest uses the Virasoro algebra,

$$2L_0|0;0\rangle = (L_1L_{-1} - L_{-1}L_1)|0;0\rangle = 0\,, \tag{2.7.8}$$

and so

$$a^X = 0\,. \tag{2.7.9}$$

Here we have used the known form of L_1 and L_{-1}: every term of each contains either a lowering operator or p^μ and so annihilates $|0;0\rangle$.

We determined the central charge in the Virasoro algebra from the OPE. It can also be determined directly from the expression for the Virasoro generators in terms of the mode operators, though some care is needed. This is left as exercise 2.11.

Let us introduce a new notation. The symbol $\circ\,\circ$ will denote *creation–annihilation normal ordering*, placing all lowering operators to the right of all raising operators, with a minus sign whenever anticommuting operators are switched. For the purposes of this definition, we will include p^μ with the lowering operators and x^μ with the raising operators. In this notation we can write

$$L_m = \frac{1}{2}\sum_{n=-\infty}^{\infty}{\circ}\alpha_{m-n}^\mu\alpha_{\mu n}{\circ} \tag{2.7.10}$$

since $a^X = 0$.

We have now introduced two forms of normal ordering, *conformal normal ordering* : : (which is what we will mean if we just refer to 'normal ordering') and creation–annihilation normal ordering $\circ\,\circ$. The former is useful because it produces operators whose OPEs and conformal

transformation properties are simple. The latter, which is probably more familiar to the reader, is useful for working out the matrix elements of the operators. Let us work out the relation between them. We start by comparing the time-ordered and creation–annihilation-ordered products. For the product $X^\mu(z,\bar z)X^\nu(z',\bar z')$ with $|z| > |z'|$, insert the mode expansions and move the lowering operators in $X^\mu(z,\bar z)$ to the right. Keeping track of the commutators gives

$$\begin{aligned} X^\mu(z,\bar z)X^\nu(z',\bar z') &= {:}X^\mu(z,\bar z)X^\nu(z',\bar z'){:} \\ &\quad + \frac{\alpha'}{2}\eta^{\mu\nu}\left[-\ln|z|^2 + \sum_{m=1}^{\infty}\frac{1}{m}\left(\frac{z'^m}{z^m}+\frac{\bar z'^m}{\bar z^m}\right)\right] \\ &= {:}X^\mu(z,\bar z)X^\nu(z',\bar z'){:} - \frac{\alpha'}{2}\eta^{\mu\nu}\ln|z-z'|^2 . \end{aligned} \tag{2.7.11}$$

Since $|z| > |z'|$, the left-hand side here is time-ordered and the sum converges. The definition (2.1.21) gave the relation between the time-ordered product and the conformal-normal-ordered product (eq. (2.1.21) was in path integral language, so the product on its right becomes time-ordered in operator formalism). It is the same as the relation (2.7.11), so that

$${:}X^\mu(z,\bar z)X^\nu(z',\bar z'){:} = :X^\mu(z,\bar z)X^\nu(z',\bar z'): . \tag{2.7.12}$$

Such a relation does not hold in general CFTs, and in fact the somewhat arbitrary grouping of p^μ with the lowering operators was done in part to give this simple result. It also does not hold for operators conformal-normal-ordered in terms of w rather than z, for example. From eq. (2.7.12) one can write the mode expansion (2.7.10) at once, giving a second derivation of $a^X = 0$.

Creation–annihilation normal-ordered products of more than two X^μs have the same combinatoric properties as for : : ordering. That is, they are obtained from the time-ordered product by summing over all subtractions — this is *Wick's theorem* from field theory. Also, to convert operators normal-ordered in one form into a different normal ordering, one sums over all subtractions using the *difference* of the two-point functions. That is, if we have two kinds of ordering,

$$[X^\mu(z,\bar z)X^\nu(z',\bar z')]_1 = [X^\mu(z,\bar z)X^\nu(z',\bar z')]_2 + \eta^{\mu\nu}\Delta(z,\bar z,z',\bar z') , \tag{2.7.13}$$

then for a general operator $\mathscr{F}$,

$$[\mathscr{F}]_1 = \exp\left(\frac{1}{2}\int d^2z\, d^2z'\, \Delta(z,\bar z,z',\bar z')\frac{\delta}{\delta X^\mu(z,\bar z)}\frac{\delta}{\delta X_\mu(z',\bar z')}\right)[\mathscr{F}]_2 . \tag{2.7.14}$$

For the linear dilaton CFT, the Laurent expansion and commutators

are unchanged, while the Virasoro generators contain the extra term

$$L_m = \frac{1}{2}\sum_{n=-\infty}^{\infty}{:}\alpha^\mu_{m-n}\alpha_{\mu n}{:} + i\left(\frac{\alpha'}{2}\right)^{1/2}(m+1)V^\mu\alpha_{\mu m} . \tag{2.7.15}$$

bc CFT

The fields b and c have the Laurent expansions

$$b(z) = \sum_{m=-\infty}^{\infty}\frac{b_m}{z^{m+\lambda}} , \quad c(z) = \sum_{m=-\infty}^{\infty}\frac{c_m}{z^{m+1-\lambda}} . \tag{2.7.16}$$

To be precise, these are only Laurent expansions if λ is an integer, which we will assume for now. The half-integer case is also of interest, but will be dealt with in detail in chapter 10. The OPE gives the anticommutators

$$\{b_m, c_n\} = \delta_{m,-n} . \tag{2.7.17}$$

Consider first the states that are annihilated by all of the $n > 0$ operators. The b_0, c_0 oscillator algebra generates two such ground states $|\downarrow\rangle$ and $|\uparrow\rangle$, with the properties

$$b_0|\downarrow\rangle = 0 , \quad b_0|\uparrow\rangle = |\downarrow\rangle , \tag{2.7.18a}$$

$$c_0|\downarrow\rangle = |\uparrow\rangle , \quad c_0|\uparrow\rangle = 0 , \tag{2.7.18b}$$

$$b_n|\downarrow\rangle = b_n|\uparrow\rangle = c_n|\downarrow\rangle = c_n|\uparrow\rangle = 0 , \quad n > 0 . \tag{2.7.18c}$$

The general state is obtained by acting on these states with the $n < 0$ modes at most once each (because these anticommute). For reasons to appear later it is conventional to group b_0 with the lowering operators and c_0 with the raising operators, so we will single out $|\downarrow\rangle$ as the ghost vacuum $|0\rangle$. In string theory we will have a holomorphic bc theory and an antiholomorphic $\tilde b\tilde c$ theory, each with $\lambda = 2$. The closed string spectrum thus includes a product of two copies of the above.

The Virasoro generators are

$$L_m \sim \sum_{n=-\infty}^{\infty}(m\lambda - n){:}b_n c_{m-n}{:} + \delta_{m,0}a^{\mathrm{g}} . \tag{2.7.19}$$

The ordering constant can be determined as in eq. (2.7.8), which gives

$$\begin{aligned} 2L_0|\downarrow\rangle &= (L_1L_{-1} - L_{-1}L_1)|\downarrow\rangle \\ &= (\lambda b_0 c_1)[(1-\lambda)b_{-1}c_0]|\downarrow\rangle = \lambda(1-\lambda)|\downarrow\rangle . \end{aligned} \tag{2.7.20}$$

Thus, $a^{\mathrm{g}} = \frac{1}{2}\lambda(1-\lambda)$ and

$$L_m = \sum_{n=-\infty}^{\infty}(m\lambda - n){:}b_n c_{m-n}{:} + \frac{\lambda(1-\lambda)}{2}\delta_{m,0} . \tag{2.7.21}$$

The constant can also be obtained by working out the relation between : : and ⁚ ⁚ for the CFT (exercises 2.13, 2.14).

For the ghost number current (2.5.14), $j = -:bc:$, the charge is

$$N^g = \frac{1}{2\pi i}\int_0^{2\pi} dw\, j_w = \sum_{n=1}^{\infty}(c_{-n}b_n - b_{-n}c_n) + c_0 b_0 - \frac{1}{2} . \tag{2.7.22}$$

It satisfies

$$[N^g, b_m] = -b_m , \quad [N^g, c_m] = c_m , \tag{2.7.23}$$

and so counts the number of c minus the number of b excitations. The ground states have ghost number $\pm\frac{1}{2}$:

$$N^g|\downarrow\rangle = -\frac{1}{2}|\downarrow\rangle , \quad N^g|\uparrow\rangle = \frac{1}{2}|\uparrow\rangle . \tag{2.7.24}$$

This depends on the value of the ordering constant, determined in exercise 2.13, but one might guess it on the grounds that the average ghost number of the ground states should be zero: the ghost number changes sign under $b \leftrightarrow c$.

The $\beta\gamma$ theory is similar; we leave the details until we need them in chapter 10.

Open strings

In the open string, the Neumann boundary condition becomes $\partial_z X^\mu = \partial_{\bar z} X^\mu$ on the real axis. There is only one set of modes, the boundary condition requiring $\alpha_m^\mu = \tilde\alpha_m^\mu$ in the expansions (2.7.1). The spacetime momentum integral (2.7.3) runs only over a semi-circle, so the normalization is now

$$\alpha_0^\mu = (2\alpha')^{1/2} p^\mu . \tag{2.7.25}$$

The expansion for X^μ is then

$$X^\mu(z,\bar z) = x^\mu - i\alpha' p^\mu \ln|z|^2 + i\left(\frac{\alpha'}{2}\right)^{1/2}\sum_{\substack{m=-\infty\\ m\neq 0}}^{\infty}\frac{\alpha_m^\mu}{m}(z^{-m} + \bar z^{-m}) . \tag{2.7.26}$$

Also,

$$L_0 = \alpha' p^2 + \sum_{n=1}^{\infty}\alpha_{-n}^\mu \cdot \alpha_{\mu n} . \tag{2.7.27}$$

The commutators are as before,

$$[\alpha_m^\mu, \alpha_n^\nu] = m\delta_{m,-n}\eta^{\mu\nu} , \quad [x^\mu, p^\nu] = i\eta^{\mu\nu} . \tag{2.7.28}$$

For the bc theory, the boundary conditions that will be relevant to the string are

$$c(z) = \tilde c(\bar z) , \quad b(z) = \tilde b(\bar z) , \quad \mathrm{Im}\, z = 0 , \tag{2.7.29}$$

written in terms of the z-coordinate where the boundary is the real axis. We can then use the doubling trick to write the holomorphic and antiholomorphic fields in the upper half-plane in terms of holomorphic fields in the whole plane,

$$c(z) \equiv \tilde c(\bar z') , \quad b(z) \equiv \tilde b(\bar z') , \quad \mathrm{Im}(z) \le 0 , \ z' = \bar z . \tag{2.7.30}$$

The open string thus has a single set of Laurent modes for each of b and c.

2.8 Vertex operators

In quantum field theory, there is on the one hand the space of states of the theory and on the other hand the set of local operators. In conformal field theory there is a simple and useful isomorphism between these, with the CFT quantized on a circle. Consider the semi-infinite cylinder in the w-coordinate,

$$0 \le \mathrm{Re}\, w \le 2\pi , \quad w \sim w + 2\pi , \quad \mathrm{Im}\, w \le 0 , \tag{2.8.1}$$

which maps into the unit disk in the coordinate $z = \exp(-iw)$. This is shown in figure 2.6. To define the path integral in the w-coordinate one must in particular specify the boundary condition as $\mathrm{Im}\, w \to -\infty$. That is, one must specify the initial state. In the z-coordinate, $\mathrm{Im}\, w = -\infty$ maps to the origin, so this is equivalent to specifying the behavior of the fields at that point. In effect, this defines a local operator at the origin, known as the *vertex operator* associated with the state. Going in the other direction, the path integral on the disk with an operator $\mathscr{A}$ at the origin maps to the path integral on the cylinder with a specified initial state $|\mathscr{A}\rangle$.

For free field theories one can easily work out the detailed form of this isomorphism. Suppose one has a conserved charge Q acting on the state $|\mathscr{A}\rangle$ as in figure 2.6(a). One can find the corresponding local operator by using the OPE to evaluate the contour integral in figure 2.6(b). Let us use this to identify the state $|1\rangle$ corresponding to the unit operator. With no operator at the origin, ∂X^μ and $\bar\partial X^\mu$ are (anti)holomorphic inside the Q contour in figure 2.6(b). The contour integrals (2.7.2) defining α_m^μ and $\tilde\alpha_m^\mu$ for $m \ge 0$ then have no poles and so vanish. Thus $|1\rangle$ is annihilated by these modes. This identifies it as the ground state,

$$|1\rangle = |0;0\rangle , \tag{2.8.2}$$

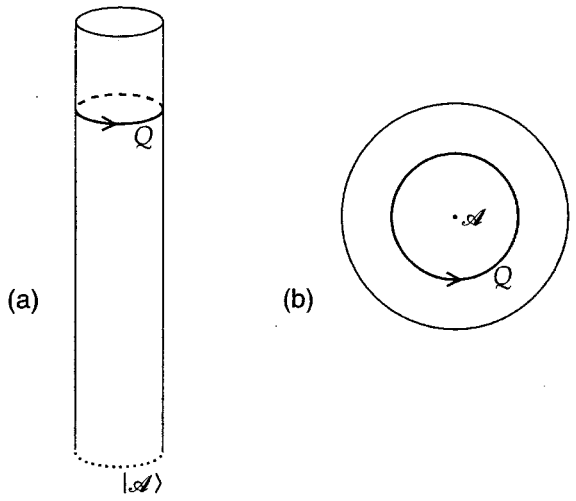


Fig. 2.6. (a) A semi-infinite cylinder in w, with initial state $|\mathscr{A}\rangle$ and charge Q. (b) The conformally equivalent unit disk with local operator $\mathscr{A}$ and contour integral for Q.

with the normalization chosen for convenience. Now consider, for example, the state $\alpha_{-m}^{\mu}|1\rangle$ with m positive. Passing to figure 2.6(b) with $Q = \alpha_{-m}^{\mu}$ surrounding the unit operator, the fields are holomorphic inside the contour, and so we can evaluate

$$\alpha_{-m}^{\mu} = \left(\frac{2}{\alpha'}\right)^{1/2} \oint \frac{dz}{2\pi} z^{-m} \partial X^{\mu}(z) \rightarrow \left(\frac{2}{\alpha'}\right)^{1/2} \frac{i}{(m-1)!} \partial^m X^{\mu}(0) \qquad (2.8.3)$$

for $m \geq 1$. Thus

$$\alpha_{-m}^{\mu}|1\rangle \cong \left(\frac{2}{\alpha'}\right)^{1/2} \frac{i}{(m-1)!} \partial^m X^{\mu}(0) , \quad m \geq 1 . \qquad (2.8.4)$$

Similarly

$$\tilde{\alpha}_{-m}^{\mu}|1\rangle \cong \left(\frac{2}{\alpha'}\right)^{1/2} \frac{i}{(m-1)!} \bar{\partial}^m X^{\mu}(0) , \quad m \geq 1 . \qquad (2.8.5)$$

This correspondence continues to hold when α_{-m}^{μ} or $\tilde{\alpha}_{-m}^{\mu}$ acts on a general state $|\mathscr{A}\rangle$. If $:\mathscr{A}(0,0):$ is any normal-ordered operator, there may be singularities in the operator product of $\partial X^{\mu}(z)$ with $:\mathscr{A}(0,0):$, but it is not hard to check that

$$\alpha_{-m}^{\mu} :\mathscr{A}(0,0): = :\alpha_{-m}^{\mu} \mathscr{A}(0,0): \qquad (2.8.6)$$

for $m > 0$ because the contour integral of the contractions will never have a single pole. One can then carry out the same manipulation (2.8.3) inside the normal ordering, and so the state with several α oscillators excited comes out as the : : normal-ordered product of the corresponding operators,

$$\alpha_{-m}^{\mu} \rightarrow i \left(\frac{2}{\alpha'}\right)^{1/2} \frac{1}{(m-1)!} \partial^m X^{\mu}(0) , \quad m \geq 1 , \qquad (2.8.7a)$$

$$\tilde{\alpha}_{-m}^{\mu} \rightarrow i \left(\frac{2}{\alpha'}\right)^{1/2} \frac{1}{(m-1)!} \bar{\partial}^m X^{\mu}(0) , \quad m \geq 1 . \qquad (2.8.7b)$$

Similarly,

$$x_0^{\mu} \rightarrow X^{\mu}(0,0) . \qquad (2.8.8)$$

Any state can be obtained from $|1\rangle$ by acting with the operators on the left-hand sides of eqs. (2.8.7) and (2.8.8). The operator corresponding to this state is then given by the : : normal-ordered product of the corresponding local operators on the right-hand sides. For example,

$$|0;k\rangle \cong :e^{ik\cdot X(0,0)}: . \qquad (2.8.9)$$

This is easy to understand: under the translation $X^{\mu} \rightarrow X^{\mu} + a^{\mu}$, both the state and the corresponding operator have the same transformation, being multiplied by $\exp(ik \cdot a)$.

The same method applies to the bc theory. For clarity we specialize to the case $\lambda = 2$ which is of interest for bosonic string theory, leaving the generalization to the reader. From the Laurent expansions (2.7.16) and the contour argument it follows that

$$b_m|1\rangle = 0 , \quad m \geq -1 , \qquad c_m|1\rangle = 0 , \quad m \geq 2 . \qquad (2.8.10)$$

Note that due to the shifts in the exponents of the Laurent expansion coming from the conformal weights of b and c, the unit operator no longer maps to a ground state. Rather, relations (2.8.10) determine that

$$|1\rangle = b_{-1}|{\downarrow}\rangle . \qquad (2.8.11)$$

The translation of the raising operators is straightforward,

$$b_{-m} \rightarrow \frac{1}{(m-2)!} \partial^{m-2} b(0) , \quad m \geq 2 , \qquad (2.8.12a)$$

$$c_{-m} \rightarrow \frac{1}{(m+1)!} \partial^{m+1} c(0) , \quad m \geq -1 . \qquad (2.8.12b)$$

Notice that the ghost number $-\frac{3}{2}$ state $b_{-1}|{\downarrow}\rangle$ maps to the ghost number 0 unit operator, and the ghost number $-\frac{1}{2}$ state $|{\downarrow}\rangle$ to the ghost number 1 operator c. The difference arises from the nontensor

property (2.5.17) of the ghost number current,

$$(\partial_z w) j_w(w) = j_z(z) + q_0 \frac{\partial_z^2 w}{\partial_z w} = j_z(z) - \frac{q_0}{z} , \tag{2.8.13}$$

where $q_0 = \lambda - \frac{1}{2} = \frac{3}{2}$. The ghost number of states is conventionally defined by the cylindrical frame expression N^g, eq. (2.7.22), while the contour argument of figure 2.6 relates the ghost number of the *vertex operator* to the radial frame charge

$$Q^g \equiv \frac{1}{2\pi i} \oint dz\, j_z = N^g + q_0 . \tag{2.8.14}$$

This applies to other charges as well; for a tensor current the charges of a state and the corresponding operator are equal.

Most of the above can be extended to the $\beta\gamma$ theory, but there are certain complications in the superstring so we defer this to chapter 10.

All the ideas of this section extend to the open string. The semi-infinite strip

$$0 \le \mathrm{Re}\, w \le \pi , \quad \mathrm{Im}\, w \le 0 , \tag{2.8.15}$$

maps to a half-disk, the intersection of the upper half-plane and the unit circle, under $z = -\exp(-iw)$. The initial state again maps to the origin, which is on the boundary, so there is an isomorphism

$$\text{local operators on the boundary} \leftrightarrow \text{states on the interval} . \tag{2.8.16}$$

The details, which are parallel to the above, are left for the reader. Again the doubling trick, extending the fields into the lower half-plane by holomorphicity, is useful in making the contour arguments.

Path integral derivation

The state–operator isomorphism is an important but unfamiliar idea and so it is useful to give also a more explicit path integral derivation. Consider a unit disk in the z-plane, with local operator $\mathscr{A}$ at the origin and with the path integral fields ϕ fixed to some specific boundary values ϕ_b on the unit circle. The path integral over the fields ϕ_i on the interior of the disk with ϕ_b held fixed produces some functional $\Psi_{\mathscr{A}}[\phi_b]$,

$$\Psi_{\mathscr{A}}[\phi_b] = \int [d\phi_i]_{\phi_b} \exp(-S[\phi_i]) \mathscr{A}(0) . \tag{2.8.17}$$

A functional of the fields is the Schrödinger representation of a state, so this is a mapping from operators to states. The Schrödinger representation, which assigns a complex amplitude to each field configuration, has many uses. It is often omitted from field theory texts, the Fock space representation being emphasized.

To go the other way, start with some state $\Psi[\phi_b]$. Consider a path integral over the annular region $1 \ge |z| \ge r$, with the fields ϕ_b on the outer circle fixed and the fields ϕ'_b integrated over the inner circle as follows:

$$\int [d\phi_i]_{\phi_b,\phi'_b} [d\phi'_b] \exp(-S[\phi_i])\, r^{L_0+\tilde{L}_0} \Psi[\phi'_b] . \tag{2.8.18}$$

That is, the integral is weighted by the state (functional) $r^{L_0+\tilde{L}_0}\Psi[\phi'_b]$. Now, the path integral over the annulus just corresponds to propagating from $|z| = r$ to $|z| = 1$, which is equivalent to acting with the operator $r^{-L_0-\tilde{L}_0}$. This undoes the operator acting on Ψ, so the path integral (2.8.18) is again $\Psi[\phi_b]$. Now take the limit as $r \to 0$. The annulus becomes a disk, and the limit of the path integral over the inner circle can be thought of as defining some local operator at the origin. By construction the path integral on the disk with this operator reproduces $\Psi[\phi_b]$ on the boundary.

Let us work this out explicitly for a single free scalar field X. On the unit circle expand

$$X_b(\theta) = \sum_{n=-\infty}^{\infty} X_n e^{in\theta} , \quad X_n^* = X_{-n} . \tag{2.8.19}$$

The boundary state $\Psi[X_b]$ can thus be regarded as a function of all of the X_n. Let us first identify the state corresponding to the unit operator, given by the path integral without insertion,

$$\Psi_1[X_b] = \int [dX_i]_{X_b} \exp\left(-\frac{1}{2\pi\alpha'} \int d^2z\, \partial X \bar{\partial} X\right) . \tag{2.8.20}$$

Evaluate this by the usual Gaussian method. Separate X_i into a classical part and a fluctuation,

$$X_i = X_{cl} + X'_i , \tag{2.8.21a}$$

$$X_{cl}(z,\bar{z}) = X_0 + \sum_{n=1}^{\infty} (z^n X_n + \bar{z}^n X_{-n}) . \tag{2.8.21b}$$

With this definition, X_{cl} satisfies the equation of motion and X'_i vanishes on the boundary. The path integral then separates,

$$\Psi_1[X_b] = \exp(-S_{cl}) \int [dX'_i]_{X_b=0} \exp\left(-\frac{1}{2\pi\alpha'} \int d^2z\, \partial X' \bar{\partial} X'\right) , \tag{2.8.22}$$

with

$$\begin{aligned} S_{cl} &= \frac{1}{2\pi\alpha'} \sum_{m,n=1}^{\infty} mn X_m X_{-n} \int_{|z|<1} d^2z\, z^{m-1} \bar{z}^{n-1} \\ &= \frac{1}{\alpha'} \sum_{m=1}^{\infty} m X_m X_{-m} . \end{aligned} \tag{2.8.23}$$

The X_i^l integral is a constant, independent of the boundary condition, so

$$\Psi_1[X_b] \propto \exp\left(-\frac{1}{\alpha'}\sum_{m=1}^{\infty} m X_m X_{-m}\right) . \tag{2.8.24}$$

This is a Gaussian, and is in fact the ground state. To see this, write the raising and lowering operators in the Schrödinger basis,

$$\alpha_n = -\frac{in}{(2\alpha')^{1/2}} X_{-n} - i\left(\frac{\alpha'}{2}\right)^{1/2}\frac{\partial}{\partial X_n} , \tag{2.8.25a}$$

$$\tilde{\alpha}_n = -\frac{in}{(2\alpha')^{1/2}} X_n - i\left(\frac{\alpha'}{2}\right)^{1/2}\frac{\partial}{\partial X_{-n}} , \tag{2.8.25b}$$

as follows from the Laurent expansion at $|z| = 1$ and the mode algebra. Acting on (2.8.24), we find

$$\alpha_n \Psi_1[X_b] = \tilde{\alpha}_n \Psi_1[X_b] = 0 , \quad n \geq 0 , \tag{2.8.26}$$

so this is indeed the ground state. Thus,

$$|1\rangle \propto |0;0\rangle . \tag{2.8.27}$$

Rather than try to keep track of the overall normalization of the path integral at this point, we define $|1\rangle = |0;0\rangle$.

Another easy calculation is the state corresponding to $\partial^k X$; this just adds a factor of $\partial^k X_{cl}(0) = k! X_k$ to the result, so

$$|\partial^k X\rangle = k! X_k \Psi_1 = -i\left(\frac{\alpha'}{2}\right)^{1/2} (k-1)! \alpha_{-k} |0;0\rangle , \tag{2.8.28}$$

and similarly for $\bar{\partial}^k X$. This extends to all products of exponentials and derivatives of X. The conformal normal ordering just cancels the effect of the X_i^l path integral.

2.9 More on states and operators

The OPE

In this section we make additional applications of the state–operator correspondence. The first is a simple and general derivation of the OPE, as shown in figure 2.7. Consider the product

$$\mathscr{A}_i(z,\bar{z})\mathscr{A}_j(0,0) , \tag{2.9.1}$$

where $|z| < 1$. We can divide the path integral into an integral over fields ϕ_i on the interior of the unit circle, an integral over fields ϕ_b on the unit circle, and an integral over fields ϕ_e on the exterior of the unit circle. This sort of cutting open of path integrals is discussed in detail in the

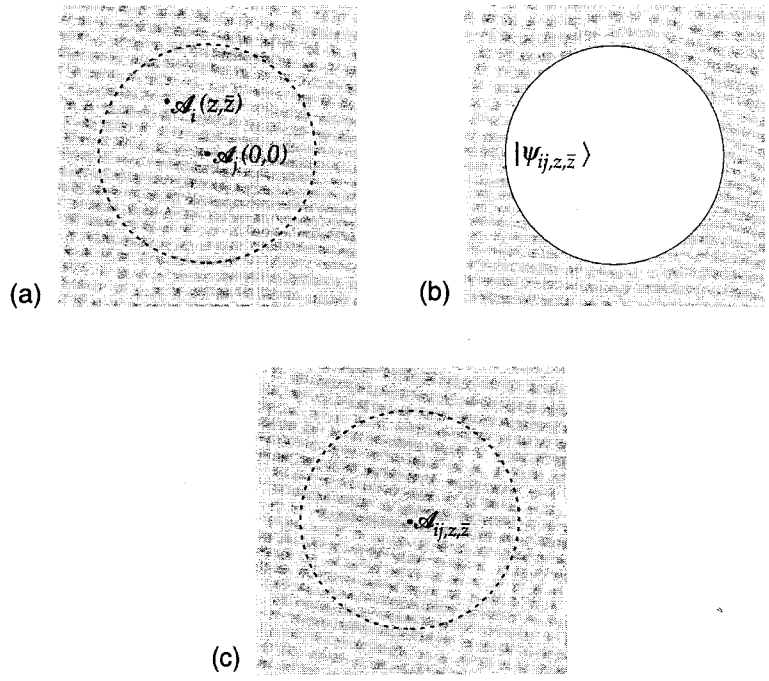


Fig. 2.7. (a) World-sheet with two local operators. (b) Integration over fields on the interior of the disk produces boundary state $|\psi_{ij,z,\bar{z}}\rangle$. (c) Sewing in a disk with the corresponding local operator. Expanding in operators of definite weight gives the OPE.

appendix. As in the discussion at the end of the previous section, the integral over ϕ_i leaves some functional of ϕ_b; call it $\Psi_{ij,z,\bar{z}}[\phi_b]$. By the state–operator correspondence this is equivalent to gluing back the disk with the appropriate local operator $\mathscr{A}_{ij,z,\bar{z}}(0,0)$ and so we are done! To put this in standard form expand in a complete set of L_0, $\tilde{L}_0$ eigenstates,

$$\mathscr{A}_{ij,z,\bar{z}} = \sum_k z^{h_k - h_i - h_j} \bar{z}^{\tilde{h}_k - \tilde{h}_i - \tilde{h}_j} C^k_{ij} \mathscr{A}_k , \tag{2.9.2}$$

with the z- and $\bar{z}$-dependence determined by the conformal weights as in eq. (2.4.20). The convergence of the OPE is just the usual convergence of a complete set in quantum mechanics. The construction is possible as long as there are no other operators with $|z'| \leq |z|$, so that we can cut on a circle of radius $|z| + \epsilon$.

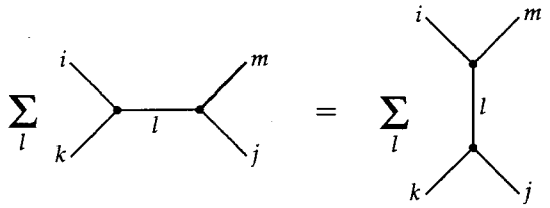


Fig. 2.8. Schematic picture of OPE associativity.

For three operators

$$\mathscr{A}_i(0,0)\mathscr{A}_j(1,1)\mathscr{A}_k(z,\bar{z})\,, \tag{2.9.3}$$

the regions of convergence of the $z \to 0$ and $z \to 1$ OPEs, $|z| < 1$ and $|1 - z| < 1$, overlap. The coefficient of $\mathscr{A}_l$ in the triple product can then be written as a sum involving $c^l_{ik}c^m_{lj}$ or as a sum involving $c^l_{jk}c^m_{li}$. Associativity requires these sums to be equal. This is represented schematically in figure 2.8.

The Virasoro algebra and highest weight states

Consider now the argument of figure 2.6 for the Virasoro generators:

$$\begin{aligned} L_m|\mathscr{A}\rangle &\cong \oint \frac{dz}{2\pi i} z^{m+1} T(z)\mathscr{A}(0,0) \\ &\cong L_m \cdot \mathscr{A}(0,0)\,. \end{aligned} \tag{2.9.4}$$

Here we have introduced a new idea and notation. Given the state–operator isomorphism, every operator acting on the Hilbert space has an image acting on the space of local operators. In other words, form the corresponding contour around the operator and evaluate the resulting local operator using the OPE. Corresponding to the L_m that map states to states are the $L_m\cdot$ that map local operators to local operators. In general there will be local operators at various positions z_i, and there will be a different copy of the Virasoro algebra for each one, obtained by a Laurent expansion centered on the position of the operator. In some geometries there may also be a standard set of generators, such as the ones centered on $z = 0$ for the cylinder. The '$\cdot$' serves as a reminder that we are talking about the Laurent coefficients around that particular operator.

In this notation the $T\mathscr{A}$ OPE is

$$T(z)\mathscr{A}(0,0) = \sum_{m=-\infty}^{\infty} z^{-m-2} L_m \cdot \mathscr{A}(0,0)\,. \tag{2.9.5}$$

To relate to the earlier notation (2.4.11), $\mathscr{A}^{(n)} = L_{n-1}\cdot\mathscr{A}$ for $n \geq 0$, and the conformal transformation of $\mathscr{A}$ is

$$\delta\mathscr{A}(z,\bar{z}) = -\epsilon \sum_{n=0}^{\infty} \frac{1}{n!}\left[\partial^n v(z) L_{n-1} + (\partial^n v(z))^* \tilde{L}_{n-1}\right]\cdot\mathscr{A}(z,\bar{z})\,. \tag{2.9.6}$$

Using the general form (2.4.14) for the $T\mathscr{A}$ OPE, we have the very useful result

$$L_{-1}\cdot\mathscr{A} = \partial\mathscr{A}\,, \quad \tilde{L}_{-1}\cdot\mathscr{A} = \bar{\partial}\mathscr{A}\,, \tag{2.9.7a}$$

$$L_0\cdot\mathscr{A} = h\mathscr{A}\,, \quad \tilde{L}_0\cdot\mathscr{A} = \tilde{h}\mathscr{A}\,. \tag{2.9.7b}$$

The OPE (2.9.5) implies that a primary field $\mathscr{O}$ of weight $(h,\tilde{h})$ corresponds to a state satisfying

$$L_0|\mathscr{O}\rangle = h|\mathscr{O}\rangle\,, \quad \tilde{L}_0|\mathscr{O}\rangle = \tilde{h}|\mathscr{O}\rangle\,, \tag{2.9.8a}$$

$$L_m|\mathscr{O}\rangle = \tilde{L}_m|\mathscr{O}\rangle = 0\,, \quad m > 0\,. \tag{2.9.8b}$$

Such a state is conventionally known as a *highest weight state*. For the CFTs of interest L_0 and $\tilde{L}_0$ are bounded below. By acting repeatedly with lowering operators on any state we eventually reach a state annihilated by further lowering operators, a highest weight state.

An interesting special case is the unit operator. The operator product $T\,1$ is nonsingular, so it follows in any CFT that

$$L_m|1\rangle = \tilde{L}_m|1\rangle = 0\,, \quad m \geq -1\,. \tag{2.9.9}$$

As noted in section 2.4, the operators L_0 and $L_{\pm1}$ form a closed algebra, as do $\tilde{L}_0$ and $\tilde{L}_{\pm1}$. The full algebra is

$$SL(2,\mathbf{R}) \times SL(2,\mathbf{R}) = SL(2,\mathbf{C})\,. \tag{2.9.10}$$

Thus $|1\rangle$ is also called the $SL(2,\mathbf{C})$-invariant state. It is the only such state because the relations (2.9.7) imply that the operator $\mathscr{A}$ corresponding to any $SL(2,\mathbf{C})$-invariant state is independent of position. It must then be a c-number, as explained after eq. (2.4.24).

Unitary CFTs

Highest weight states play a special role in string theory, and also in the theory of representations of the Virasoro algebra. We will return to this at various points, but for now we derive a few important general results that hold in *unitary* CFTs. A unitary CFT is one that has a positive inner product $\langle\ |\ \rangle$ such that

$$L_m^\dagger = L_{-m}\,, \quad \tilde{L}_m^\dagger = \tilde{L}_{-m}\,. \tag{2.9.11}$$

Recall that the inner product defines the adjoint via $\langle\alpha|A\beta\rangle = \langle A^\dagger\alpha|\beta\rangle$.

For example, the X^μ CFT is unitary for *spacelike* μ if we take the inner product

$$\langle 0;k|0;k'\rangle = 2\pi\delta(k-k') \tag{2.9.12}$$

for the ground state and define

$$\alpha_m^\dagger = \alpha_{-m}\,, \quad \tilde{\alpha}_m^\dagger = \tilde{\alpha}_{-m}\,. \tag{2.9.13}$$

This implicitly defines the inner products of all higher states. This CFT is not unitary for $\mu = 0$ because of the opposite sign in the commutator there.

The first constraint is that any state in a unitary CFT must have $h, \tilde{h} \geq 0$. If this is true for highest weight states it is true for all states. For a highest weight state the Virasoro algebra gives

$$2h_{\mathcal{O}}\langle\mathcal{O}|\mathcal{O}\rangle = 2\langle\mathcal{O}|L_0|\mathcal{O}\rangle = \langle\mathcal{O}|[L_1, L_{-1}]|\mathcal{O}\rangle = \|L_{-1}|\mathcal{O}\rangle\|^2 \geq 0\,, \tag{2.9.14}$$

so $h_{\mathcal{O}} \geq 0$. It also follows that if $h_{\mathcal{O}} = 0$ then

$$L_{-1}\cdot\mathcal{O} = \tilde{L}_{-1}\cdot\mathcal{O} = 0\,, \tag{2.9.15}$$

and so $\mathcal{O}$ is independent of position. As noted before, $\mathcal{O}$ must then be a c-number. That is, the unit operator is the only (0,0) operator in a unitary CFT. Curiously, X^μ itself is the one notable exception: the corresponding state $x^\mu|0;0\rangle$ is nonnormalizable because of the infinite range of X^μ, and equation (2.9.14) no longer holds. The general theorems are of the most use for the CFTs corresponding to compactified dimensions, where this kind of exception cannot occur.

In a similar way, one finds that an operator in a unitary CFT is holomorphic if and only if $\tilde{h} = 0$, and antiholomorphic if and only if $h = 0$; this is an important result so we repeat it,

$$\partial\mathscr{A} = 0 \;\Leftrightarrow\; h = 0\,, \quad \bar{\partial}\mathscr{A} = 0 \;\Leftrightarrow\; \tilde{h} = 0\,. \tag{2.9.16}$$

Finally, using the above argument with the commutator $[L_n, L_{-n}]$, one can show that $c, \tilde{c} \geq 0$ in a unitary CFT. In fact, the only CFT with $c = 0$ is the trivial one, $L_n = 0$.

Zero-point energies

The state–operator mapping gives a simple alternative derivation of the various normal ordering constants. In any CFT, we know that $L_0|1\rangle = 0$, and this determines the additive normalization of L_0. In the X CFT, $|1\rangle$ is the ground state $|0;0\rangle$, so a^X vanishes. In the bc theory, $|1\rangle$ is the excited state $b_{-1}|\downarrow\rangle$, so the weight of $|\downarrow\rangle$ is -1 in agreement with the earlier result (2.7.21) for $\lambda = 2$.

This also provides one physical interpretation of the central charge. In a unitary CFT the ground state is $|1\rangle$ with $L_0 = \tilde{L}_0 = 0$. The conformal transformation (2.6.10) between the radial and time-translation generators then implies that

$$E = -\frac{c+\tilde{c}}{24} \tag{2.9.17}$$

for the ground state. This is a Casimir energy, coming from the finite size of the system, and it depends only on the central charge. On dimensional grounds this energy is inverse to the size ℓ of the system, which is 2π here, so the general result would be

$$E = -\frac{\pi(c+\tilde{c})}{12\ell}\,. \tag{2.9.18}$$

We have now given three honest ways of calculating normal ordering constants: from the Virasoro algebra as in eq. (2.7.8), by relating the two forms of normal ordering as in eq. (2.7.11), and from the state–operator mapping above. Nevertheless the idea of adding up zero-point energies is intuitive and a useful mnemonic, so we give a prescription that can be checked by one of the more honest methods:

1. Add the zero-point energies, $\frac{1}{2}\omega$ for each bosonic mode and $-\frac{1}{2}\omega$ for each fermionic (anticommuting) mode.

2. One encounters divergent sums of the form $\sum_{n=1}^{\infty}(n-\theta)$, the θ arising when one considers nontrivial periodicity conditions. Define this to be

$$\sum_{n=1}^{\infty}(n-\theta) = \frac{1}{24} - \frac{1}{8}(2\theta-1)^2\,. \tag{2.9.19}$$

   This is the value one obtains as in eq. (1.3.32) by regulating and discarding the quadratically divergent part.

3. The above gives the normal ordering constant for the w-frame generator T_0, eq. (2.6.8). For L_0 we must add the nontensor correction $\frac{1}{24}c$.

For the free boson, the modes are integer-valued so we get one-half of the sum (2.9.19) for $\theta = 0$ after step 2, which is $-\frac{1}{24}$. This is just offset by the correction in step 3, giving $a^X = 0$. For the ghosts we similarly get $\frac{2}{24} - \frac{26}{24} = -1$.

Exercises

2.1 Verify that

$$\partial\bar{\partial}\ln|z|^2 = \partial\frac{1}{\bar{z}} = \bar{\partial}\frac{1}{z} = 2\pi\delta^2(z,\bar{z})$$

(a) by use of the divergence theorem (2.1.9);
(b) by regulating the singularity and then taking a limit.

2.2 Work out explicitly the expression (2.2.5) for the normal-ordered product of four X^μ fields. Show that it is a harmonic function of each of the positions.

2.3 The expectation value of a product of exponential operators on the plane is

$$\left\langle \prod_{i=1}^{n} :e^{ik_i\cdot X(z_i,\bar{z}_i)}: \right\rangle = iC^X(2\pi)^D\delta^D\left(\textstyle\sum_{i=1}^n k_i\right)\prod_{\substack{i,j=1\\ i<j}}^{n}|z_{ij}|^{\alpha' k_i\cdot k_j} ,$$

with C^X a constant. This can be obtained as a limit of the expectation value (6.2.17) on the sphere, which we will obtain by several methods in chapter 6.
(a) Show that the leading behavior as one vertex operator approaches another is in agreement with the OPE (2.2.14).
(b) As implied by eq. (2.2.13), the expectation value is $|z_{12}|^{\alpha' k_1\cdot k_2}$ times a function that is smooth as $z_1 \to z_2$. For $n = 3$, work out the explicit expansion of this smooth function in powers of z_{12} and $\bar{z}_{12}$. From the ratio of the large-order terms find the radius of convergence.
(c) Give a general proof that the free-field OPE (2.2.4) is convergent inside the dashed circle in figure 2.1.

2.4 Extend the OPE (2.2.14) up to and including subleading terms of order z^2, $z\bar{z}$, and $\bar{z}^2$. Show that the $z,\bar{z}$-dependence of each term is in agreement with the general form (2.4.20). Give also the symmetric form as in eq. (2.2.15) and discuss the properties of the coefficients under interchange of the two operators.

2.5 Derive the classical Noether theorem in the form usually found in textbooks. That is, assume that

$$S[\phi] = \int d^d\sigma\, \mathscr{L}(\phi(\sigma), \partial_a\phi(\sigma))$$

and ignore the variation of the measure. Invariance of the action implies that the variation of the Lagrangian density is a total derivative,

$$\delta\mathscr{L} = \epsilon\partial_a\mathscr{K}^a$$

under a symmetry transformation (2.3.1). Then the classical result, with conventional string normalization, is

$$j^a = 2\pi i\left(\frac{\partial\mathscr{L}}{\partial\phi_{\alpha,a}}\epsilon^{-1}\delta\phi_\alpha - \mathscr{K}^a\right) .$$

Derive the Lagrangian equations of motion and show that they imply that this current is conserved. Show that it agrees with eq. (2.3.4).

2.6 Consider the flat Euclidean metric δ_{ab}, in d dimensions. An infinitesimal coordinate transformation $\delta\sigma^a = v^a(\sigma)$ changes the metric by $\delta g_{ab} = -\partial_a v_b - \partial_b v_a$. Determine the most general $v^a(\sigma)$ that leaves the metric invariant up to a local rescaling. There are $\frac{1}{2}(d+1)(d+2)$ independent solutions in $d > 2$, and an infinite number in $d = 2$. [Although the result is given in many places in the literature, the author has not found a simple derivation.]

2.7 (a) By computing the relevant OPEs, confirm the weights stated in eq. (2.4.17) and determine which operators are tensors.
(b) Do this for the same operators in the linear dilaton theory.

2.8 What is the weight of $f_{\mu\nu}:\partial X^\mu\bar{\partial}X^\nu e^{ik\cdot X}:$? What are the conditions on $f_{\mu\nu}$ and k_μ in order for it to be a tensor?

2.9 Derive the central charges for the linear dilaton, bc, and $\beta\gamma$ CFTs by working out the TT OPEs.

2.10 Consider now a patch of world-sheet with boundary. For convenience suppose that the patch lies in the upper half-z-plane, with the real axis being the boundary. Show that expectation values of a normal-ordered operator, say $:\partial X^\mu(z)\bar{\partial}X_\mu(\bar{z}):$, although finite in the interior, diverge as z approaches the boundary (represent the effect of the boundary by an image charge). Define *boundary normal ordering* $*\ *$, which is the same as $:\ :$ except that the contraction includes the image charge piece as well. Operators on the boundary are finite if they are boundary normal ordered.

In a general CFT on a manifold with boundary, the interior operators and the boundary operators are independent. Label the bases $\mathscr{A}_i$ and $\mathscr{B}_r$ respectively, and define the former by $:\ :$ and the latter by $*\ *$. Each set has its own closed OPE: $\mathscr{A}\mathscr{A} \to \mathscr{A}$ and $\mathscr{B}\mathscr{B} \to \mathscr{B}$. In addition, the sets are related: $\mathscr{A}_i$, as it approaches the boundary, can be expanded in terms of the $\mathscr{B}_r$. Find the leading behaviors of

$$*e^{ik_1\cdot X(y_1)}**e^{ik_2\cdot X(y_2)}* , \quad y_1 \to y_2 \ (y \text{ real}) ,$$
$$:e^{ik\cdot X(z,\bar{z})}: , \quad \mathrm{Im}(z) \to 0 .$$

The identity (2.7.14) is useful for the latter.

2.11 Evaluate the central charge in the Virasoro algebra for X^μ by calcu-

lating

$$L_m(L_{-m}|0;0\rangle) - L_{-m}(L_m|0;0\rangle) .$$

2.12 Use the OPE and the contour results (2.6.14) and (2.6.15) to derive the commutators (2.7.5) and (2.7.17).

2.13 (a) Show that

$$:b(z)c(z'): - {}^{\circ}_{\circ}b(z)c(z'){}^{\circ}_{\circ} = \frac{(z/z')^{1-\lambda} - 1}{z - z'}$$

by the method of eq. (2.7.11).
(b) Use this to determine the ordering constant in N^{g}, eq. (2.7.22). You also need the conformal transformation (2.8.14).
(c) Show that one obtains the same value for N^{g} by a heuristic treatment of the ordering similar to that in section 2.9.

2.14 (a) Determine the ordering constant in L_0^{g}, eq. (2.7.21), using part (a) of the previous exercise.
(b) Determine it using the heuristic rules in section 2.9.

2.15 Work out the state–operator mapping for the open string X^μ CFT.

2.16 Let $\mathcal{O}$ be a tensor field of weight $(h, \tilde{h})$. Evaluate the commutators $[L_m, \mathcal{O}(z, \bar{z})]$ using the contour argument of figure 2.4. Show in particular that the commutator of a (1,1) tensor is a total derivative.

2.17 Extend the argument of eq. (2.9.14) to show that $c \geq 0$.

3

The Polyakov path integral

We now begin a systematic study of string theory using the Polyakov path integral and conformal field theory. After an introduction to the path integral picture, we discuss gauge fixing, the Weyl anomaly, and vertex operators. We then generalize to strings propagating in curved spacetime.

3.1 Sums over world-sheets

The Feynman path integral is one way to represent a quantum theory, and it is a very natural method for describing interactions in string theory. In path integral quantization, amplitudes are given by summing over all possible histories interpolating between the initial and final states. Each history is weighted by

$$\exp(iS_{\mathrm{cl}}/\hbar) , \tag{3.1.1}$$

with S_{cl} the classical action for the given history. Thus, one would define an amplitude in string theory by summing over all world-sheets connecting given initial and final curves, as in figure 3.1(a) for the open string and figure 3.1(b) for the closed string. The analogous sum for the relativistic point particle produces the free propagator (exercise 5.1).

One can imagine a number of ways in which strings might interact. One would be a contact interaction, an energy when two strings intersect. Another would be a long-range force mediated by some quantum field. However, it is not possible to *add* such interactions to string theory in a way that is consistent with the symmetries; we will get some idea later of why this is so. Rather, the only interactions that are allowed are those that are already implicit in the sum over world-sheets. Consider, for example, the world-sheets shown in figure 3.2. Figure 3.2(a) looks like a quantum correction to the open string amplitude of figure 3.1(a), involving intermediate states with two open strings. Figure 3.2(b) has three external

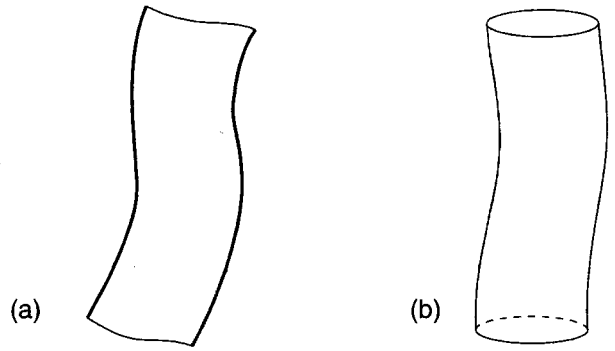


Fig. 3.1. (a) An open string world-sheet with the topology of a strip. The heavier curves are the world-lines of string endpoints. (b) A closed string world-sheet with the topology of a cylinder.

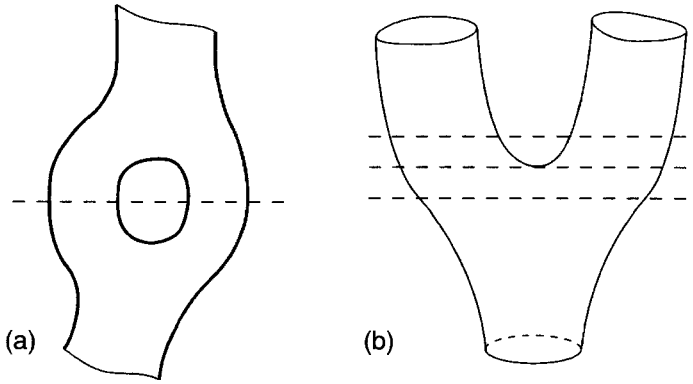


Fig. 3.2. (a) Quantum correction to open string propagation. (b) Decay of one closed string into two. The dashed lines are time-slices.

closed strings and represents one string decaying into two. It turns out that this is the correct way to introduce interactions into string theory. We will see that these interactions produce a consistent theory, finite and unitary, which includes gravity.

It is interesting to consider the process of figure 3.2(b) as seen at successive times as shown in figure 3.3. A closed string splits into two, or in the reverse process two join into one. This is the basic closed string interaction. In closed string theory, all particles are obtained as various states of excitation of the string, and all interactions (gauge, gravitational,

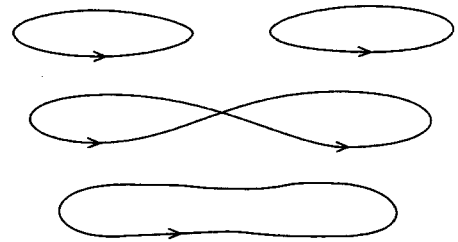


Fig. 3.3. Successive time-slices from figure 3.2(b). The arrows indicate the orientation of the string.

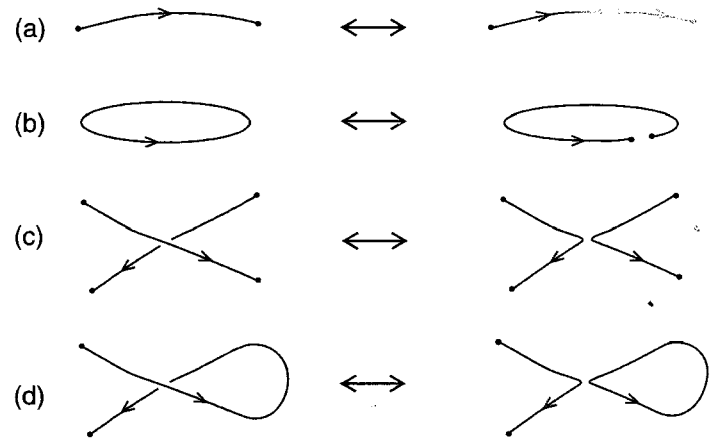


Fig. 3.4. Processes involving open strings: (a) open $\leftrightarrow$ open + open; (b) closed $\leftrightarrow$ open; (c) open + open $\leftrightarrow$ open + open; (d) open $\leftrightarrow$ open + closed.

and Yukawa, for example) arise from the single process of figure 3.3. For theories with open strings there are several additional processes that can occur, as shown in figure 3.4. With a given slicing, the interactions of figures 3.3 and 3.4 appear to occur at a definite spacetime point, but this is an illusion. A different slicing, in a boosted Lorentz frame, will put the apparent interaction at a different point; there is no distinguished point in spacetime. The interaction arises only from the global topology of the world-sheet, while the local properties of the world-sheet are the same as they were in the free case. As discussed in the introduction, it is this smearing out of the interaction that cuts off the short-distance divergences of gravity.

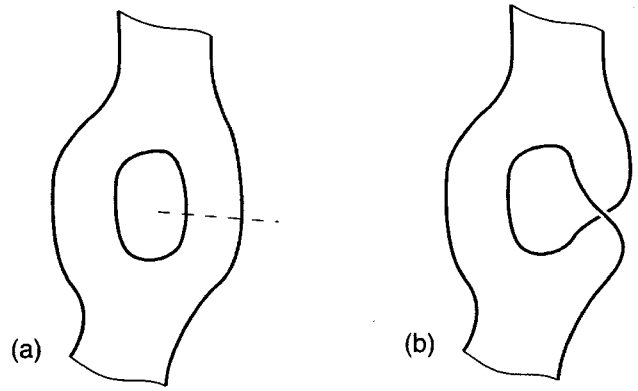


Fig. 3.5. Oriented (a) and unoriented (b) contributions to the two-open-string amplitude.

There are actually several different string theories, depending on which topologies we include in the sum over world-sheets. First note that the world-sheets we have drawn have two kinds of boundary, the source boundaries corresponding to the initial and final string configurations, and the endpoint boundaries corresponding to the world-lines of the endpoints of open strings. Ignore for now the source boundaries; we will discuss the details of the sources later in this chapter. So in the following discussion 'boundary' means endpoint boundary. There are then four ways to define the sum over world-sheets, which correspond to the four free string theories listed at the end of section 1.4:

1. Closed oriented: All oriented world-sheets without boundary.
2. Closed unoriented: All world-sheets without boundary.
3. Closed plus open oriented: All oriented world-sheets with any number of boundaries.
4. Closed plus open unoriented: All world-sheets with any number of boundaries.

There are two things that we should expand upon here. The first is the connection between the inclusion of unoriented world-sheets and the $\Omega = +1$ projection described in section 1.4. Figure 3.5 shows two world-sheets that are included in the unoriented theory with open strings. The unoriented world-sheet of figure 3.5(b) is equivalent to cutting the oriented surface of figure 3.5(a) along the dashed line, which we will parameterize

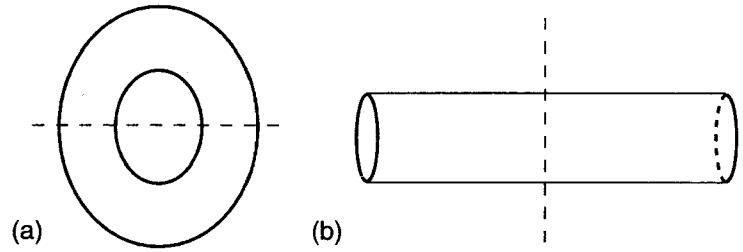


Fig. 3.6. (a) An annulus: the intermediate state is two open strings. (b) The same topology drawn as a cylinder: the intermediate state is one closed string.

by $0 \leq \sigma \leq \pi$, and requiring the fields on the cut edges to satisfy

$$X^{\mu}_{\text{upper}}(\sigma) = X^{\mu}_{\text{lower}}(\pi - \sigma) \,. \tag{3.1.2}$$

This in turn is equivalent to acting with the operator Ω on the open string state. Adding these two surfaces with appropriate weight then amounts to inserting the operator $\frac{1}{2}(1 + \Omega)$, which projects onto states of $\Omega = +1$. Summing over all oriented and unoriented surfaces inserts projection operators on all intermediate states, reducing the spectrum to the $\Omega = +1$ sector.

Second, the list above does not include any theory with open strings but not closed. Let us explain in more detail why this is the case. Consider a world-sheet with the topology of an annulus, figure 3.6(a). For convenience we have drawn only the vacuum amplitude, but by attaching appropriate sources the same argument will also apply to scattering amplitudes. We see from the dashed line that this is a process with an intermediate state of two open strings, and so world-sheets of this topology must be present in any theory with open strings. In figure 3.6(b) we have drawn the same topology as a cylinder and cut it open on an intermediate state of a single closed string. Thus, even if we start with open strings only, the sum over world-sheets will necessarily include processes in which scattering of open strings produces closed strings. As we develop the amplitudes we will see how this works in detail.

The need for closed strings in open string theory can also be seen in another way. Consider the interactions in figures 3.4(a) and (b), but time reversed: respectively two open strings joining into one, and an open to closed string transition. Near the interaction point these are the same, two endpoints coalescing. To have the first interaction without the second would require some nonlocal constraint on the dynamics of the string; this would surely be inconsistent. One can make the same argument with figures 3.4(c) and (d), where the interaction is locally the reconnection of

a pair of strings. So if any open string interaction is allowed, then so is some process in which closed strings are produced from open strings.

3.2 The Polyakov path integral

We now begin to develop the sum over world-sheets, integrating over the fields of the Polyakov formalism. We make one change from chapter 1: the Minkowskian world-sheet metric γ_{ab} is replaced with a Euclidean world-sheet metric $g_{ab}(\sigma^1,\sigma^2)$, of signature $(+,+)$. The integral runs over all Euclidean metrics and over all embeddings $X^\mu(\sigma^1,\sigma^2)$ of the world-sheet in Minkowski spacetime:

$$\int [dX\, dg]\, \exp(-S)\,. \tag{3.2.1}$$

The Euclidean action is

$$S = S_X + \lambda\chi\,, \tag{3.2.2}$$

with

$$S_X = \frac{1}{4\pi\alpha'}\int_M d^2\sigma\, g^{1/2} g^{ab}\partial_a X^\mu \partial_b X_\mu\,, \tag{3.2.3a}$$

$$\chi = \frac{1}{4\pi}\int_M d^2\sigma\, g^{1/2} R + \frac{1}{2\pi}\int_{\partial M} ds\, k\,. \tag{3.2.3b}$$

The geodesic curvature k was defined in exercise 1.3; note that the boundary now is always spacelike.

The advantage of the Euclidean path integral is that the integral over metrics is better defined. The topologically nontrivial world-sheets we have been describing can have nonsingular Euclidean metrics but not Minkowskian ones. We will take the Euclidean theory (3.2.1) as our starting point, and we will show how to give it a precise definition and that it defines a consistent spacetime theory with the advertised properties. However, let us give a brief formal argument that it is equivalent to the Minkowski theory with which we began.

We start again with the example of the point particle, where the path integral

$$\int [d\eta\, dX] \exp\left[\frac{i}{2}\int d\tau\, \left(\eta^{-1}\dot{X}^\mu \dot{X}_\mu - \eta m^2\right)\right] \tag{3.2.4}$$

is oscillatory. The path integral over η and X^μ is a product of ordinary integrals, and so we can deform contours just as in ordinary integration. If we take

$$\eta(\tau) \to e^{-i\theta}\eta(\tau)\,, \quad X^0(\tau) \to e^{-i\theta} X^0(\tau)\,, \tag{3.2.5}$$

for infinitesimal θ, all terms in the exponent acquire a negative real part and so this acts as a convergence factor. Now we can perform a contour rotation in field space, all the way to $\eta = -ie$, $X^0 = -iX^D$. The integral becomes

$$\int [de\, dX] \exp\left[-\frac{1}{2}\int d\tau \left(e^{-1}\sum_{\mu=1}^{D}\dot{X}^\mu\dot{X}^\mu + em^2\right)\right]. \tag{3.2.6}$$

This is the Euclidean point-particle analog of the path integral (3.2.1). We have just made a contour rotation, so the Euclidean path integral gives the same amplitude as the Minkowski one.

The same procedure works for the Polyakov action if we write the metric in terms of a tetrad, $\gamma_{ab} = -e^0_a e^0_b + e^1_a e^1_b$, and make the same rotation on e^0_a. This provides a formal justification for the equivalence of the Minkowski and Euclidean path integrals.[1] It has been shown by explicit calculation that they define the same amplitudes, respectively in the light-cone and conformal gauges. We have in the action (3.2.3a) left the rotation of X^0 implicit, as in the previous chapter, so as to emphasize the Minkowski nature of spacetime. Written in terms of X^0, the equations of motion and the OPE are covariant, with metric $\eta^{\mu\nu}$.

We have noted that χ is locally a total derivative in two dimensions and therefore depends only on the topology of the world-sheet — it is the *Euler number* of the world-sheet. Thus, the $e^{-\lambda\chi}$ factor in the path integral affects only the relative weighting of different topologies in the sum over world-sheets. In particular, if one adds an extra strip to a world-sheet, as we have done in going from figure 3.1(a) to figure 3.2(a), the Euler number decreases by one and the path integral is weighted by an extra factor of e^λ. Since adding a strip corresponds to emitting and reabsorbing a virtual open string, the amplitude for emitting an open string from any process is proportional to $e^{\lambda/2}$. Adding a handle to any world-sheet reduces the Euler number by 2 and adds a factor $e^{2\lambda}$. Since this corresponds to emitting and reabsorbing a closed string, the amplitude for emitting a closed string is proportional to e^λ. The Euler term in the action thus controls the coupling constants in the string theory, with

$$g_o^2 \sim g_c \sim e^\lambda\,. \tag{3.2.7}$$

By the way, λ might seem to be a free parameter in the theory, contradicting the statement in the introduction that there are no such parameters

[1] In more than two dimensions, things are not so simple because the Hilbert action behaves in a more complicated way under the rotation. No simple rotation damps the path integral. In particular, the meaning of the Euclidean path integral for four-dimensional gravity is very uncertain.

in string theory. This is an important point, and we will resolve it in section 3.7.

As an aside, the counting of couplings extends in a simple way to world-sheets with stringy sources. The sources for closed strings are closed boundary loops, while the open string sources are boundary segments. For convenience we will require the open string source boundaries to meet the endpoint boundaries at right angles in the world-sheet metric. We have to be careful because the boundary curvature k diverges at the corners: the Euler number is

$$\chi = \tilde{\chi} + \frac{1}{4} n_c \,, \tag{3.2.8}$$

where $\tilde{\chi}$ includes only the integral over the smooth segments of boundary and n_c is the number of corners. The correct weight for the path integral is

$$\exp(-\lambda\tilde{\chi}) = \exp(-\lambda\chi + \lambda n_c/4) \,. \tag{3.2.9}$$

This follows from unitarity: the λ-dependence must be unchanged if we cut through a world-sheet, leaving fewer internal lines and more external sources. This will be the case for the weight (3.2.9) because the surface integral in $\tilde{\chi}$ cancels on the cut edges. Adding a closed string source amounts to cutting a hole in the world-sheet and decreases χ by 1. Adding an open string source leaves χ unchanged and increases n_c by 2. Thus we reach the same conclusion (3.2.7) about the amplitude for emission of closed or open strings.

3.3 Gauge fixing

The path integral (3.2.1) is not quite right. It contains an enormous overcounting, because configurations (X, g) and (X', g') that are related to one another by the local diff×Weyl symmetry represent the same physical configuration. We need, in effect, to divide by the volume of this local symmetry group,

$$\int \frac{[dX\, dg]}{V_{\text{diff}\times\text{Weyl}}} \exp(-S) \equiv Z \,. \tag{3.3.1}$$

We will carry this out by gauge-fixing, integrating over a slice that cuts through each gauge equivalence class once and obtaining the correct measure on the slice by the Faddeev–Popov method.

In chapter 1 we imposed light-cone gauge. This is useful for some purposes but hides some of the symmetry of the theory, so we now make a different choice. Note that the metric, being symmetric, has three independent components, and that there are three gauge functions, the

two coordinates and the local scale of the metric. Thus there is just enough gauge freedom to eliminate the integration over the metric, fixing it at some specific functional form which we will call the *fiducial* metric,

$$g_{ab}(\sigma) \to \hat{g}_{ab}(\sigma) \,. \tag{3.3.2}$$

A simple choice is the flat or *unit gauge* metric

$$\hat{g}_{ab}(\sigma) = \delta_{ab} \,. \tag{3.3.3}$$

One sometimes wishes to consider the effect of the diff group alone. In this case, one can bring an arbitrary metric to within a Weyl transformation of the unit form. This is the *conformal gauge,*

$$\hat{g}_{ab}(\sigma) = \exp[2\omega(\sigma)]\delta_{ab} \,. \tag{3.3.4}$$

Let us see explicitly that this counting of fields and gauge symmetries works — that any metric can be brought to the flat form at least locally, in a given neighborhood on the world-sheet. First, make a Weyl transformation to set to zero the Ricci scalar built from g_{ab}. The Weyl transformation of the Ricci scalar is

$$g'^{1/2}R' = g^{1/2}(R - 2\nabla^2\omega) \,. \tag{3.3.5}$$

To set $R' = 0$ requires solving $2\nabla^2\omega = R$ as an equation for ω. This is always possible, at least locally, by a standard argument from electrostatics. Now, in two dimensions the vanishing of the Ricci scalar implies the vanishing of the entire Riemann tensor, because the symmetries of the Riemann tensor imply that

$$R_{abcd} = \frac{1}{2}(g_{ac}g_{bd} - g_{ad}g_{bc})R \,. \tag{3.3.6}$$

So the metric is flat and coordinate-equivalent to the unit metric (3.3.3).

Actually, locally there is a little bit of extra gauge freedom. There are diff×Weyl transformations that leave the metric in unit gauge and so are not fixed by the choice of metric. It is convenient here to introduce a complex coordinate $z = \sigma^1 + i\sigma^2$ as in chapter 2, where the flat metric is $ds^2 = dz d\bar{z}$. Consider a coordinate transformation such that z' is a holomorphic function of z,

$$z' \equiv \sigma'^1 + i\sigma'^2 = f(z) \,, \tag{3.3.7}$$

combined with a Weyl transformation. The new metric is

$$ds'^2 = \exp(2\omega)|\partial_z f|^{-2} dz' d\bar{z}' \,. \tag{3.3.8}$$

Then for

$$\omega = \ln|\partial_z f| \tag{3.3.9}$$

the metric is invariant. This extra gauge freedom does not conflict with our earlier counting (3 = 3) because the mismatch, the set of holomorphic reparameterizations, is of measure zero compared to all reparameterizations.

When we consider the full world-sheet, most or all of this extra freedom will be eliminated by the boundary conditions. There are two issues here. First, in the discussion of Noether's theorem and the Ward identities, we were careful to emphasize below eq. (2.3.10) that these results depend only on the symmetry being defined locally on the world-sheet. The transformation (3.3.7), (3.3.9) will therefore give rise to a conserved current and to Ward identities. This is precisely the conformal invariance studied in chapter 2. We see that it arises as the subgroup of diff×Weyl that leaves invariant the unit metric.

The second issue is what happens to our counting of metric and gauge degrees of freedom when we do consider the full world-sheet. We will take this up in chapter 5, where we will see that globally there is a small mismatch between the two, a finite number of parameters. Until then, we will be concerned only with local properties of the world-sheet theory.

The Faddeev–Popov determinant

After fixing the metric, the functional integral runs along a slice parameterized by X^μ alone. In order to obtain the correct measure, we follow the Faddeev–Popov procedure. The steps are the same as were used to obtain the gauge-fixed measure in Yang-Mills theory, but we will give a self-contained discussion. The idea, illustrated in figure 3.7, is to separate the path integral into an integral over the gauge group times an integral along the gauge slice, and to divide out the former. The Faddeev–Popov determinant is the Jacobian for this change of variables.

Let ζ be shorthand for a combined coordinate and Weyl transformation,

$$\zeta : g \to g^\zeta , \quad g^\zeta_{ab}(\sigma') = \exp[2\omega(\sigma)] \frac{\partial \sigma^c}{\partial \sigma'^a} \frac{\partial \sigma^d}{\partial \sigma'^b} g_{cd}(\sigma) . \tag{3.3.10}$$

Following a standard route, we define the Faddeev–Popov measure Δ_{FP} by

$$1 = \Delta_{FP}(g) \int [d\zeta]\, \delta(g - \hat{g}^\zeta) , \tag{3.3.11}$$

where again $\hat{g}_{ab}$ is the fiducial metric. In (3.3.11), $[d\zeta]$ is a gauge-invariant measure on the diff×Weyl group. We will not need the explicit form of this measure; we will discuss the question of its existence in the next section. The delta function is actually a delta functional, requiring $g_{ab}(\sigma) = \hat{g}^\zeta_{ab}(\sigma)$ at every point.

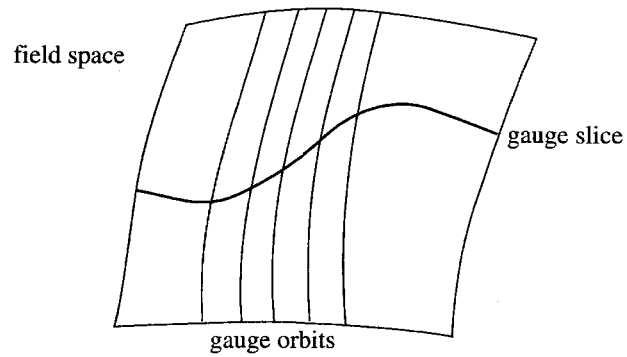


Fig. 3.7. Schematic picture of the Faddeev–Popov procedure. The gauge orbits are families of gauge-equivalent configurations. Integrating over the whole field space and dividing by the orbit volume is equivalent to integrating over a slice that intersects each orbit once, with appropriate Jacobian.

Inserting (3.3.11) into the functional integral (3.3.1), the latter becomes

$$Z[\hat{g}] = \int \frac{[d\zeta\, dX\, dg]}{V_{\text{diff}\times\text{Weyl}}} \Delta_{FP}(g) \delta(g - \hat{g}^\zeta) \exp(-S[X, g]) . \tag{3.3.12}$$

For later reference, we denote explicitly the dependence of Z on the choice of fiducial metric. Carry out the integration over g_{ab}, and also rename the dummy variable $X \to X^\zeta$, to obtain

$$Z[\hat{g}] = \int \frac{[d\zeta\, dX^\zeta]}{V_{\text{diff}\times\text{Weyl}}} \Delta_{FP}(\hat{g}^\zeta) \exp(-S[X^\zeta, \hat{g}^\zeta]) . \tag{3.3.13}$$

Now use the gauge invariance[2] of Δ_{FP}, of the functional integral measure $[dX]$, and of the action to obtain

$$Z[\hat{g}] = \int \frac{[d\zeta\, dX]}{V_{\text{diff}\times\text{Weyl}}} \Delta_{FP}(\hat{g}) \exp(-S[X, \hat{g}]) . \tag{3.3.14}$$

Finally, nothing in the integrand depends on ζ, so the integral over ζ just produces the volume of the gauge group and cancels the denominator,

[2] To show this invariance:

$$\Delta_{FP}(g^\zeta)^{-1} = \int [d\zeta']\, \delta(g^\zeta - \hat{g}^{\zeta'}) = \int [d\zeta']\, \delta(g - \hat{g}^{\zeta^{-1}\cdot\zeta'})$$
$$= \int [d\zeta'']\, \delta(g - \hat{g}^{\zeta''}) = \Delta_{FP}(g)^{-1} ,$$

where $\zeta'' = \zeta^{-1} \cdot \zeta'$. In the second equality we have used the gauge invariance of the delta function, and in the third the invariance of the measure.

leaving

$$Z[\hat{g}] = \int [dX]\, \Delta_{FP}(\hat{g}) \exp(-S[X,\hat{g}]) \,. \tag{3.3.15}$$

Thus, $\Delta_{FP}(\hat{g})$ is the correct measure on the slice.

To evaluate the expression (3.3.11) for the Faddeev–Popov measure, let us pretend that the gauge choice precisely fixes the gauge symmetry, so that for exactly one value of ζ the delta function $\delta(g - \hat{g}^\zeta)$ is nonzero; for $\delta(\hat{g} - \hat{g}^\zeta)$ this would be the identity. As noted above, there is a small global mismatch, but we will deal with this when the time comes — it does not affect the local properties of the world-sheet. For ζ near the identity, we can expand

$$\begin{aligned} \delta g_{ab} &= 2\delta\omega\, g_{ab} - \nabla_a \delta\sigma_b - \nabla_b \delta\sigma_a \\ &= (2\delta\omega - \nabla_c \delta\sigma^c) g_{ab} - 2(P_1 \delta\sigma)_{ab} \,. \end{aligned} \tag{3.3.16}$$

We have defined a differential operator P_1 that takes vectors into traceless symmetric 2-tensors,

$$(P_1 \delta\sigma)_{ab} = \frac{1}{2}(\nabla_a \delta\sigma_b + \nabla_b \delta\sigma_a - g_{ab} \nabla_c \delta\sigma^c) \,. \tag{3.3.17}$$

Near the identity, the inverse determinant (3.3.11) becomes

$$\begin{aligned} &\Delta_{FP}(\hat{g})^{-1} \\ &= \int [d\delta\omega\, d\delta\sigma]\, \delta\left[-(2\delta\omega - \hat{\nabla}\cdot\delta\sigma)\hat{g} + 2\hat{P}_1 \delta\sigma\right] \\ &= \int [d\delta\omega\, d\beta\, d\delta\sigma] \exp\left\{ 2\pi i \int d^2\sigma\, \hat{g}^{1/2} \beta^{ab} \left[-(2\delta\omega - \hat{\nabla}\cdot\delta\sigma)\hat{g} + 2\hat{P}_1 \delta\sigma\right]_{ab} \right\} \\ &= \int [d\beta'\, d\delta\sigma] \exp\left\{ 4\pi i \int d^2\sigma\, \hat{g}^{1/2} \beta'^{ab} (\hat{P}_1 \delta\sigma)_{ab} \right\} \,. \end{aligned} \tag{3.3.18}$$

A hat on a differential operator indicates that it contains the fiducial metric $\hat{g}$. In the second equality we have used the integral representation for the functional delta function, introducing the symmetric tensor field β^{ab}. In the third equality we have integrated over $\delta\omega$, which produces a delta functional forcing β^{ab} to be traceless; the functional integral $[d\beta']$ is then over *traceless* symmetric tensors. We now have a representation for $\Delta_{FP}(\hat{g})^{-1}$ as a functional integral over a vector field $\delta\sigma^a$ and over a traceless symmetric tensor β'^{ab}.

As discussed in the appendix, we can invert this path integral by replacing each bosonic field with a corresponding Grassmann *ghost* field.

$$\delta\sigma^a \to c^a \,, \tag{3.3.19a}$$

$$\beta'_{ab} \to b_{ab} \,, \tag{3.3.19b}$$

with b^{ab}, like β'^{ab}, being traceless. Thus,

$$\Delta_{FP}(\hat{g}) = \int [db\, dc] \exp(-S_g) \,, \tag{3.3.20}$$

where the ghost action S_g, with a convenient normalization for the fields, is

$$S_g = \frac{1}{2\pi} \int d^2\sigma\, \hat{g}^{1/2} b_{ab} \hat{\nabla}^a c^b = \frac{1}{2\pi} \int d^2\sigma\, \hat{g}^{1/2} b_{ab} (\hat{P}_1 c)^{ab} \,. \tag{3.3.21}$$

Locally on the world-sheet, the Polyakov path integral is now

$$Z[\hat{g}] = \int [dX\, db\, dc] \exp(-S_X - S_g) \,. \tag{3.3.22}$$

The action is quadratic in the fields, so we can evaluate the path integral in terms of determinants as discussed in the appendix. This step will be carried out in more detail in the chapter 5; the result here is

$$Z[\hat{g}] = (\det \hat{\nabla}^2)^{-D/2} \det \hat{P}_1 \,, \tag{3.3.23}$$

the first determinant coming from the X integration and the second from the ghost integration.

In conformal gauge, $\hat{g}_{ab}(\sigma) = \exp[2\omega(\sigma)]\delta_{ab}$, the ghost action (3.3.21) is

$$\begin{aligned} S_g &= \frac{1}{2\pi} \int d^2z \left(b_{zz} \nabla_{\bar{z}} c^z + b_{\bar{z}\bar{z}} \nabla_z c^{\bar{z}} \right) \\ &= \frac{1}{2\pi} \int d^2z \left(b_{zz} \partial_{\bar{z}} c^z + b_{\bar{z}\bar{z}} \partial_z c^{\bar{z}} \right) \,. \end{aligned} \tag{3.3.24}$$

Notice that $\omega(\sigma)$ does not appear in the final form. This is because the covariant $\bar{z}$-derivative of a tensor that has only z indices reduces to the simple derivative, and vice versa, as the reader can check by working out the connection tensor. We have positioned the indices, lower on b and upper on c, to take advantage of this. Since the action (3.3.24) is Weyl-invariant, it follows that b_{ab} and c^a must be neutral under the Weyl transformation (in contrast to b^{ab} and c_a, for example, which contain additional factors of the metric). Since the conformal transformation is a combination of a coordinate transformation (3.3.7) and a Weyl transformation (3.3.9), we can immediately read off the conformal transformations of b_{ab} and c^a from the tensor indices: this is a bc CFT with $(h_b, h_c) = (2, -1)$ and a $\tilde{b}\tilde{c}$ CFT with $(\tilde{h}_b, \tilde{h}_c) = (2, -1)$.

The derivation that we have given of the gauge-fixed path integral is sometimes referred to as the 'heuristic' derivation. The problem is that the starting point, the Polyakov functional integral (3.3.1), is ill defined due to the enormous gauge overcounting. The gauge-fixed functional integral (3.3.22), on the other hand, is quite well defined, and can be calculated explicitly. For practical purposes we should perhaps regard the gauge-fixed form as our true starting point, with the gauge-invariant one giving

an intuitive interpretation. In the gauge-fixed form, the essential content of the gauge symmetry is still present as *BRST invariance*, which we will take up in the next chapter.

In the open string we need also to consider a patch containing a segment of the boundary. It is convenient to regard functional integrals as being over fields on a fixed coordinate region, so the diff invariance is limited to coordinate changes that take the boundary into itself. The variation $\delta\sigma^a$ therefore has vanishing normal component,

$$n_a \delta\sigma^a = 0 \,. \tag{3.3.25}$$

This boundary condition is inherited by the corresponding ghost field,

$$n_a c^a = 0 \,, \tag{3.3.26}$$

so c^a is proportional to the tangent vector t^a. The equations of motion then provide a boundary condition on b_{ab}. They have a surface term

$$\int_{\partial M} ds\, n^a b_{ab} \delta c^b = 0 \,. \tag{3.3.27}$$

With the boundary condition on c^a, this implies

$$n_a t_b b^{ab} = 0 \,. \tag{3.3.28}$$

These are the boundary conditions that we used in section 2.7.

3.4 The Weyl anomaly

A key feature of string theory is that it is not consistent in all spacetime backgrounds, but only in those satisfying certain conditions. For the bosonic string theory in flat spacetime, the condition is $D = 26$. This was first discovered as a pathology of the scattering amplitudes. In the light-cone analysis, it arises as a loss of Lorentz invariance, as we briefly and rather heuristically explained in section 1.3. The underlying source of the restriction is an *anomaly* in the local world-sheet symmetries. In our present formalism, the anomaly is in the Weyl symmetry: $T^a{}_a$ vanishes classically, but not in the quantum theory.

The reader who has studied quantum field theory has already encountered a related anomaly, although it was likely not described in the same language. A global Weyl transformation, $\omega(\sigma) =$ constant, is equivalent to an overall rescaling of lengths. A number of familiar field theories in four dimensions are classically invariant under such a rescaling. These include massless scalar field theory with a ϕ^4 interaction, and non-Abelian gauge theory. The scale invariance is evident because the Lagrangian contains no dimensionful parameters — the coupling constant in each case is dimensionless. However, we know that due to divergences in the quantum

theory there is a nonzero renormalization group *beta function* in each theory, implying that the effective coupling constant is in fact a function of the length scale. Correspondingly, the trace of the energy-momentum tensor vanishes in the classical theory but is nonzero and proportional to the beta function when quantum effects are taken into account.

In the previous section, we ignored the possibility of anomalies. We need to check this: is the gauge-fixed path integral $Z[g]$ really independent of the choice of fiducial metric,

$$Z[g^\zeta] = Z[g] \;? \tag{3.4.1}$$

In section 4.2 we will extend this to more general changes of gauge. For convenience, we omit the hat on g henceforth. We will also be interested in path integrals with additional insertions,

$$\langle \ldots \rangle_g \equiv \int [dX\, db\, dc]\, \exp(-S[X,b,c,g]) \ldots \,. \tag{3.4.2}$$

Then we also require

$$\langle \ldots \rangle_{g^\zeta} = \langle \ldots \rangle_g \,. \tag{3.4.3}$$

We are not at the present time interested in the details of the insertions themselves — that will be the subject of the next two sections — so we will restrict attention to gauge transformations ζ vanishing near the positions of the fields in '...'. That is, we are asking that Weyl invariance hold as an operator equation.

It is easy to preserve the diff and Poincaré invariances in the quantum theory. For example, one may define the gauge-fixed path integral using a Pauli–Villars regulator as is done for the harmonic oscillator in section A.1, dividing by the path integral for a massive regulator field. The massive regulator field can be coupled to the metric in a diff- and Poincaré-invariant way. However, the diff- and Poincaré-invariant mass term $\mu^2 \int d^2\sigma\, g^{1/2}\, Y^\mu Y_\mu$ for a regulator field Y^μ is not Weyl-invariant. We must therefore check Weyl invariance by hand: does the invariance (3.4.3) hold for Weyl transformations? Any Weyl transformation can be obtained as the result of repeated infinitesimal transformations, so it is sufficient to study the latter.

The energy-momentum tensor T^{ab} is the infinitesimal variation of the path integral with respect to the metric,

$$\delta \langle \ldots \rangle_g = -\frac{1}{4\pi} \int d^2\sigma\, g(\sigma)^{1/2} \delta g_{ab}(\sigma) \left\langle T^{ab}(\sigma) \ldots \right\rangle_g \,. \tag{3.4.4}$$

Classically, T^{ab} comes entirely from the variation of the action, and this coincides with the Noether definition

$$T^{ab}(\sigma) \overset{\text{classical}}{=} \frac{4\pi}{g(\sigma)^{1/2}} \frac{\delta}{\delta g_{ab}(\sigma)} S \,. \tag{3.4.5}$$

It also reduces to the earlier definition in the limit of a flat world-sheet. If one takes δg_{ab} to have the form of a coordinate transformation, the diff invariance of the path integral implies the conservation of T^{ab}.

In the following analysis, we will at first neglect possible boundary terms. For a Weyl transformation, the definition (3.4.4) of T^{ab} means that

$$\delta_W \langle \dots \rangle_g = -\frac{1}{2\pi} \int d^2\sigma\, g(\sigma)^{1/2} \delta\omega(\sigma) \langle T^a{}_a(\sigma) \dots \rangle_g \,, \tag{3.4.6}$$

so Weyl invariance with the general insertions '...' can be phrased as the operator statement that the energy-momentum tensor is traceless:

$$T^a{}_a \stackrel{?}{=} 0 \,. \tag{3.4.7}$$

The classical action is Weyl-invariant, but in the quantum theory the trace might be nonzero because we have not managed to find a fully gauge-invariant regulator. The trace must be diff- and Poincaré-invariant, because we have preserved these symmetries, and it must vanish in the flat case because we know from the previous chapter that that theory is conformally invariant. This leaves only one possibility,

$$T^a{}_a = a_1 R \tag{3.4.8}$$

for some constant a_1, with R again the world-sheet Ricci scalar. Terms with more than two derivatives are forbidden for dimensional reasons. In units of world-sheet length, g_{ab} and X^μ are both dimensionless, so the constant a_1 in eq. (3.4.8) is dimensionless as well. Terms with more derivatives would have a coefficient with positive powers of the cutoff length used to define the path integral, and so vanish in the limit where the cutoff is taken to zero.

Calculation of the Weyl anomaly

The possible obstruction to a gauge-invariant theory has been reduced to the single number a_1. In fact, this number is proportional to something we have already calculated, the *central charge* c of the CFT on a flat world-sheet. Both a_1 and c are related to the two-point function of the energy-momentum tensor, though to different components. To get the precise relation between them we will need to use the diff invariance, as follows.

In complex coordinates,

$$T_{z\bar{z}} = \frac{a_1}{2} g_{z\bar{z}} R \,. \tag{3.4.9}$$

Taking the covariant derivative,

$$\nabla^{\bar{z}} T_{\bar{z}z} = \frac{a_1}{2} \nabla^{\bar{z}} (g_{z\bar{z}} R) = \frac{a_1}{2} \partial_z R \,, \tag{3.4.10}$$

where we have used the fact that the metric is covariantly constant. Then by conservation of T_{ab},

$$\nabla^z T_{zz} = -\nabla^{\bar{z}} T_{\bar{z}z} = -\frac{a_1}{2} \partial_z R \,. \tag{3.4.11}$$

To fix a_1 we compare the Weyl transformations on the left and right. The Weyl transformation (3.3.5) of the right is

$$a_1 \partial_z \nabla^2 \delta\omega \approx 4 a_1 \partial_z^2 \partial_{\bar{z}} \delta\omega \,, \tag{3.4.12}$$

where we have expanded around a flat world-sheet. To get the Weyl transformation of T_{zz} we use first the conformal transformation (2.4.24),

$$\delta T_{zz}(z) = -\frac{c}{12} \partial_z^3 v^z(z) - 2\partial_z v^z(z) T_{zz}(z) - v^z(z) \partial_z T_{zz}(z) \,. \tag{3.4.13}$$

From the discussion in section 3.3, this conformal transformation consists of a coordinate transformation $\delta z = v$ plus a Weyl transformation $2\delta\omega = \partial v + (\partial v)^*$. The last two terms in the variation are the coordinate transformation of the tensor, so the Weyl transformation, to leading order around flat space, is

$$\delta_W T_{zz} = -\frac{c}{6} \partial_z^2 \delta\omega \,. \tag{3.4.14}$$

Acting with $\partial^z = 2\partial_{\bar{z}}$ and comparing with the transformation (3.4.12) gives

$$c = -12 a_1 \,, \quad T^a{}_a = -\frac{c}{12} R \,. \tag{3.4.15}$$

Let us derive this again by a slightly longer route, along which we obtain a useful intermediate result. In conformal gauge,

$$R = -2 \exp(-2\omega) \partial_a \partial_a \omega \,, \tag{3.4.16a}$$

$$\nabla^2 = \exp(-2\omega) \partial_a \partial_a \,. \tag{3.4.16b}$$

By the contraction with two lowered indices we mean $\delta^{ab} \partial_a \partial_b$. The Weyl variation (3.4.6) of $Z[g]$ becomes

$$\delta_W Z[\exp(2\omega)\delta] = \frac{a_1}{\pi} Z[\exp(2\omega)\delta] \int d^2\sigma\, \delta\omega \partial_a \partial_a \omega \,, \tag{3.4.17}$$

where '$\exp(2\omega)\delta$' denotes conformal gauge. This integrates immediately to

$$Z[\exp(2\omega)\delta] = Z[\delta] \exp\left(-\frac{a_1}{2\pi} \int d^2\sigma\, \partial_a \omega \partial_a \omega \right) . \tag{3.4.18}$$

Since every metric is diff-equivalent to a conformal metric, eq. (3.4.18) actually determines the complete metric-dependence of $Z[g]$. We need to find the diff-invariant expression that reduces to eq. (3.4.18) in conformal gauge. Using the relations (3.4.16), one verifies that the desired expression

is

$$Z[g] = Z[\delta] \exp\left[\frac{a_1}{8\pi}\int d^2\sigma \int d^2\sigma' g^{1/2} R(\sigma) G(\sigma,\sigma') g^{1/2} R(\sigma')\right] , \quad (3.4.19)$$

where

$$g(\sigma)^{1/2}\nabla^2 G(\sigma,\sigma') = \delta^2(\sigma-\sigma') \quad (3.4.20)$$

defines the scalar Green's function. This is an interesting result: the path integral is completely determined by the anomaly equation.

Now expand $Z[g]$ around a flat background, $g_{ab} = \delta_{ab} + h_{ab}$, and keep terms of second order in $h_{z\bar z}$. The Ricci scalar is $4\partial_z^2 h_{z\bar z}$ to first order in $h_{z\bar z}$, so

$$\begin{aligned}\ln\frac{Z[\delta+h]}{Z[\delta]} &\approx \frac{a_1}{8\pi^2}\int d^2z \int d^2z' \,(\partial_z^2 \ln|z-z'|^2) h_{z\bar z}(z,\bar z)\partial_{z'}^2 h_{z\bar z}(z',\bar z') \\ &= -\frac{3a_1}{4\pi^2}\int d^2z \int d^2z' \,\frac{h_{z\bar z}(z,\bar z)h_{z\bar z}(z',\bar z')}{(z-z')^4} . \end{aligned} \quad (3.4.21)$$

We can also calculate this by using second order perturbation theory in the metric, which by eq. (3.4.4) gives

$$\ln\frac{Z[\delta+h]}{Z[\delta]} \approx \frac{1}{8\pi^2}\int d^2z \int d^2z' \, h_{z\bar z}(z,\bar z)h_{z\bar z}(z',\bar z')\langle T_{zz}(z)T_{zz}(z')\rangle_\delta . \quad (3.4.22)$$

Now use the standard TT OPE (2.4.25). All terms except the first involve operators of nonzero spin and so vanish by rotational invariance, leaving $\langle T_{zz}(z)T_{zz}(z')\rangle_1 = \frac{1}{2}c(z-z')^{-4}$. Comparison with the result (3.4.21) from the Weyl anomaly gives again eq. (3.4.15).

Discussion

The world-sheet theory consists of the X^μ, with central charge D, and the ghosts, whose central charge (2.5.12) is -26 for $\lambda = 2$, giving

$$c = c^X + c^{\rm g} = D - 26 . \quad (3.4.23)$$

The theory is Weyl-invariant only for $D = 26$, the same as the condition for Lorentz invariance found in chapter 1. When the Weyl anomaly is nonvanishing, different gauge choices are inequivalent, and as with anomalous gauge theories any choice leads to a pathology such as loss of covariance or loss of unitarity. In light-cone gauge, for example, choosing different spacetime axes requires a change of gauge, so the underlying Weyl anomaly is translated into a Lorentz anomaly.

There is another thing that one might try: ignore the Weyl invariance and treat only the diff invariance as a gauge symmetry. The metric would then have one real degree of freedom $\omega(\sigma)$, to be integrated over rather than gauge-fixed. This is particularly plausible because the quantum effects in eq. (3.4.18) have generated an action for ω which looks just like that for one of the X^μ; for $D < 26$ the sign is that of a spacelike dimension. However, the physics is a little exotic because the diff transformation (and therefore the energy-momentum tensor) of ω is different from that of the X^μ. In fact the theory is the linear dilaton CFT. So there is no $(D+1)$-dimensional Lorentz invariance and this theory is not useful for our immediate purpose, which is to understand physics in our part of the universe. It is an interesting model, however, and we will return to this point briefly in section 9.9.

We have found above that a flat world-sheet CFT can be coupled to a curved metric in a Weyl-invariant way if and only if the total central charge c is zero. There is a paradox here, however, because the same argument applies to the antiholomorphic central charge $\tilde c$, and we have seen from the example of the bc CFT that c need not equal $\tilde c$. The point is that we have implicitly assumed world-sheet diff invariance in eqs. (3.4.11) and (3.4.19). If $c \neq \tilde c$ there is no diff-invariant expression that agrees with both the $O(h_{zz}^2)$ and $O(h_{z\bar z}^2)$ calculations. The CFT then *cannot* be coupled to a curved metric in a diff-invariant way; there is a *diff* or *gravitational* anomaly. The paradox shows that $c = \tilde c$ is necessary for a CFT to be free of gravitational anomalies, and it can also be shown that this is sufficient.

Inclusion of boundaries

Let us first return to eq. (3.4.8) and consider the more general possibility

$$T^a{}_a = a_1 R + a_2 . \quad (3.4.24)$$

This is allowed by diff invariance but would be nonvanishing in the flat limit. Earlier we set a_2 to zero by the choice (2.3.15b) of T_{ab}, explained in the footnote below that equation. Equivalently, the existence of more than one conserved T_{ab} on the flat world-sheet means that there is some freedom in how we couple to a curved metric. In particular, we can add to the action

$$S_{\rm ct} = b\int d^2\sigma\, g^{1/2} , \quad (3.4.25)$$

which is a two-dimensional cosmological constant. This is not Weyl-invariant. It adds $2\pi b g_{ab}$ to T_{ab} and so the trace (3.4.24) becomes

$$T^a{}_a = a_1 R + (a_2 + 4\pi b) . \quad (3.4.26)$$

We can therefore make the second term vanish by setting $b = -a_2/4\pi$; this is what we have done implicitly in eq. (2.3.15b). The separate values of a_2 and b depend on definitions and are uninteresting.

No such local counterterm can remove a_1. For example, $\int d^2\sigma\, g^{1/2}R$ is Weyl-invariant, as we have already noted. So if a_1 is nonzero, there is a real anomaly in the Weyl symmetry.

We now consider the effects of boundaries. Including boundary terms, the most general possible variation is[3]

$$\delta_W \ln Z[g] = -\frac{1}{2\pi}\int_M d^2\sigma\, g^{1/2}(a_1 R + a_2)\delta\omega - \frac{1}{2\pi}\int_{\partial M} ds\,(a_3 + a_4 k + a_5 n^a \partial_a)\delta\omega\ , \tag{3.4.27}$$

where again $k = -t^a n_b \nabla_a t^b$ is the geodesic curvature of the boundary. The possible counterterms are

$$S_{ct} = \int_M d^2\sigma\, g^{1/2} b_1 + \int_{\partial M} ds\,(b_2 + b_3 k)\ . \tag{3.4.28}$$

The b_3 term is in addition to the geodesic curvature term in the Euler number, in eq. (3.2.3b). The Weyl variation of S_{ct}, being careful to include the dependence of ds and of the unit vectors t and n on the metric, is

$$\delta_W S_{ct} = 2\int_M d^2\sigma\, g^{1/2} b_1 \delta\omega + \int_{\partial M} ds\,(b_2 + b_3 n^a \partial_a)\delta\omega\ . \tag{3.4.29}$$

We see that b_1, b_2, and b_3 can be chosen to set a_2, a_3, and a_5 to zero, leaving a_1 and a_4 as potential anomalies. There is one more tool at our disposal, the *Wess–Zumino consistency condition*. Take the second functional variation

$$\begin{aligned}\delta_{W_1}(\delta_{W_2} \ln\langle\ldots\rangle_g) &= \frac{a_1}{\pi}\int_M d^2\sigma\, g^{1/2}\,\delta\omega_2 \nabla^2 \delta\omega_1 - \frac{a_4}{2\pi}\int_{\partial M} ds\,\delta\omega_2\, n^a \partial_a \delta\omega_1 \\ &= -\frac{a_1}{\pi}\int_M d^2\sigma\, g^{1/2}\,\partial_a\delta\omega_2 \partial^a \delta\omega_1 + \frac{2a_1 - a_4}{2\pi}\int_{\partial M} ds\, n^a \delta\omega_2\, \partial_a \delta\omega_1\ .\end{aligned} \tag{3.4.30}$$

This is a second (functional) derivative and so by definition must be symmetric in $\delta\omega_1$ and $\delta\omega_2$. The first term is symmetric but the second one is not, and so $a_4 = 2a_1$. It follows that boundaries, with open string boundary conditions, give rise to no new Weyl anomalies. Weyl invariance holds in the quantum theory if and only if $a_1 = -c/12$ vanishes. This is determined by the local physics in the *interior* of the world-sheet.

The Wess–Zumino consistency condition is a very powerful restriction on the form of possible anomalies. A second application enables us to show that the central charge is a constant. This followed above from

[3] We require all terms to be independent of the fields X^μ, as can be shown to hold in free field theory. When we consider the interacting case in section 3.7, the Weyl anomaly will take a more general form.

Lorentz invariance and dimensional analysis, but in more general CFTs one might imagine that

$$T^a{}_a(\sigma) = -\frac{\mathscr{C}(\sigma)}{12}R(\sigma)\ , \tag{3.4.31}$$

where $\mathscr{C}(\sigma)$ is some local operator. In particular, the trace would still vanish on a flat world-sheet. Repeating the calculation (3.4.30) with $-\mathscr{C}(\sigma)/12$ in place of a_1, there is an additional asymmetric term proportional to $\partial_a \mathscr{C}(\sigma)$ from the integration by parts. The consistency condition thus implies that expectation values involving $\mathscr{C}(\sigma)$ are position-independent, implying that $\mathscr{C}(\sigma)$ itself is just a numerical constant c.

3.5 Scattering amplitudes

The idea of a sum over all world-sheets bounded by given initial and final curves seems like a natural one. However, it is difficult to define this in a way that is consistent with the local world-sheet symmetries, and the resulting amplitudes are rather complicated. There is one special case where the amplitudes simplify. This is the limit where the string sources are taken to infinity. This corresponds to a scattering amplitude, an S-matrix element, with the incoming and outgoing strings specified. For most of this book we will confine ourselves to the S-matrix and similar 'on-shell' questions. This is not completely satisfying, because one would like to be able to discuss the state of the system at finite times, not just asymptotic processes. For now we focus on the S-matrix because it is well understood, while it is not clear what is the correct way to think about string theory off-shell. We will return to this issue briefly in the next section, and again in chapter 9.

So we are now considering a process like that shown in figure 3.8(a), where the sources are pulled off to infinity. We will give a heuristic argument for how these sources should be represented in the path integral. Away from the scattering process the strings propagate freely. Each of the incoming and outgoing legs is a long cylinder, which can be described with a complex coordinate w,

$$-2\pi t \le \mathrm{Im}\, w \le 0\ , \quad w \cong w + 2\pi\ . \tag{3.5.1}$$

The lower end of the cylinder, $\mathrm{Im}\, w = -2\pi t$, is the end with the external source, the upper end fits onto the rest of the world-sheet, and the circumference is the periodic $\mathrm{Re}\, w$ direction. The limit corresponding to the scattering process is $t \to \infty$. It may seem that we are confusing a long distance in spacetime with a long cylinder in the world-sheet coordinate, but this will turn out to be correct: we will see later that propagation

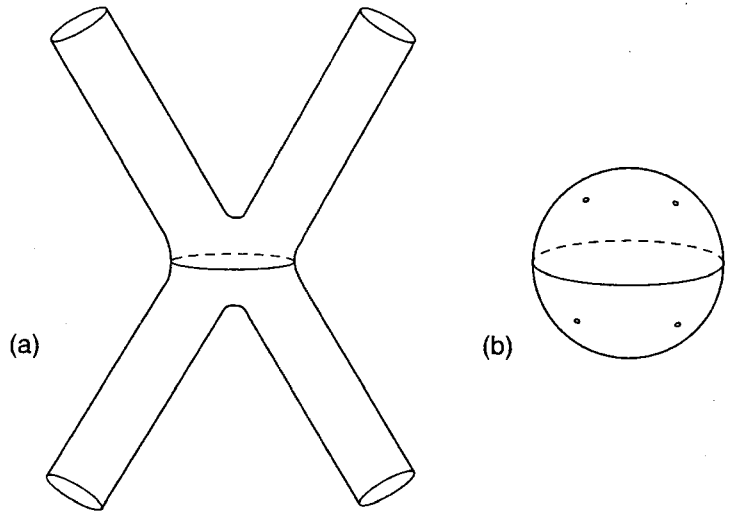


Fig. 3.8. (a) Scattering of four closed strings, with sources approaching $X^0 = \pm\infty$. (b) Conformally equivalent picture of the four-closed-string scattering: a sphere with small holes.

over long spacetime distances comes precisely from world-sheets where a cylinder is growing long in the sense above.

As we know from chapter 2, the cylinder has a conformally equivalent description in terms of the coordinate z,

$$z = \exp(-iw)\,, \quad \exp(-2\pi t) \leq |z| \leq 1\,. \tag{3.5.2}$$

In this picture, the long cylinder is mapped into the unit disk, where the external string state is a tiny circle in the center. In the geometry of the internal metric, the process of figure 3.8(a) now looks like figure 3.8(b). In the limit $t \to \infty$, the tiny circles shrink to points and the world-sheet reduces to a sphere with a point-like insertion for each external state.

The same idea holds for external open string states. The long strips in the process of figure 3.9(a) can be described as the region

$$-2\pi t \leq \operatorname{Im} w \leq 0\,, \quad 0 \leq \operatorname{Re} w \leq \pi\,, \tag{3.5.3}$$

where $\operatorname{Im} w = -2\pi t$ is the source and $\operatorname{Re} w = 0, \pi$ are endpoint boundaries. Under $z = -\exp(-iw)$, this maps into the intersection of the unit disk with the upper half-plane, with a tiny semi-circle cut out at the origin,

$$\exp(-2\pi t) \leq |z| \leq 1\,, \quad \operatorname{Im} z \geq 0\,. \tag{3.5.4}$$

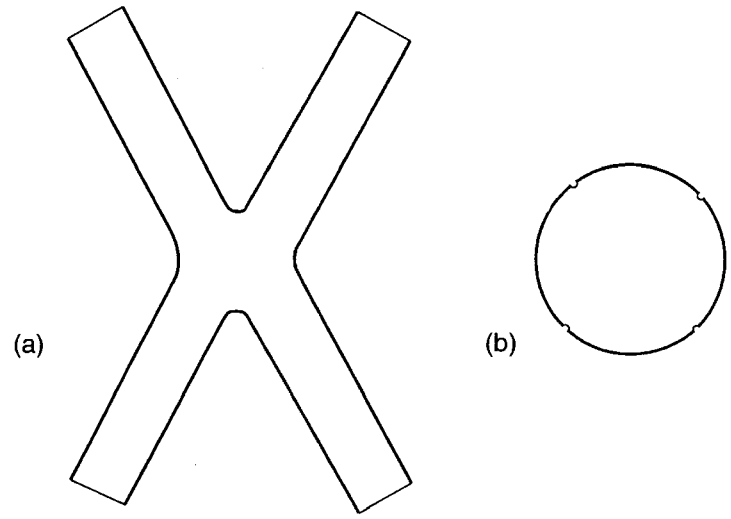


Fig. 3.9. (a) Scattering of four open strings, with sources approaching $X^0 = \pm\infty$. (b) Conformally equivalent picture: a disk with small dents.

The scattering process now looks like figure 3.9(b). In the limit $t \to \infty$, the sources shrink to points located on the (endpoint) boundary.

This is the state–operator mapping, which we have already seen. Each string source becomes a *local* disturbance on the world-sheet. To a given incoming or outgoing string, with D-momentum k^μ and internal state j, there corresponds a local *vertex operator* $\mathscr{V}_j(k)$ determined by the limiting process. Incoming and outgoing states are distinguished by the sign of k^0; for an incoming state $k^\mu = (E, \mathbf{k})$ and for an outgoing state $k^\mu = -(E, \mathbf{k})$. Figures 3.8 and 3.9 depict only the lowest order amplitudes, but the construction is clearly general: we may restrict attention to compact world-sheets, with no tubes going off to infinity but with a pointlike insertion in the interior for each external closed string and a pointlike insertion on the boundary for each external open string. An n-particle S-matrix element is then given by

$$S_{j_1 \ldots j_n}(k_1, \ldots, k_n) = \sum_{\substack{\text{compact} \\ \text{topologies}}} \int \frac{[dX\, dg]}{V_{\text{diff} \times \text{Weyl}}} \exp(-S_X - \lambda\chi) \prod_{i=1}^{n} \int d^2\sigma_i\, g(\sigma_i)^{1/2} \mathscr{V}_{j_i}(k_i, \sigma_i)\,. \tag{3.5.5}$$

To make the vertex operator insertions diff-invariant, we have integrated

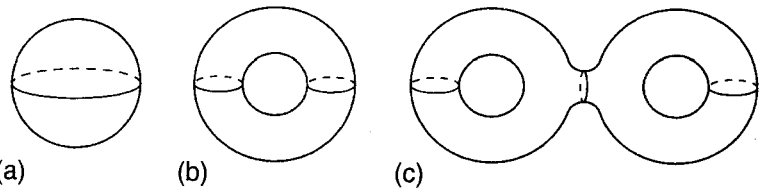


Fig. 3.10. Compact connected closed oriented surfaces of genus (a) 0, (b) 1, (c) 2.

them over the world-sheet. In the remainder of volume one, we will see that our guess is correct and that this does indeed define a sensible string S-matrix.

Depending on which of the four types of string theories we are considering, the sum over topologies may include unoriented world-sheets, and/or world-sheets with boundary. In general, the sum over topologies is not restricted to connected world-sheets; the total process may involve two or more widely separated sets of particles scattering independently. It is convenient to focus on the connected S-matrix, given by restricting the sum to connected surfaces.

To obtain the connected S-matrix, then, we must sum over all compact, connected topologies for the world-sheet. In two dimensions, the classification of these topologies is well known. Any compact, connected, oriented two-dimensional surface without boundary is topologically equivalent to a sphere with g handles; g is known as the *genus* of the surface. In figure 3.10 we show $g = 0$, 1, and 2. Boundaries can be added by cutting holes in a closed surface. Any compact, connected, oriented two-dimensional surface is topologically equivalent to a sphere with g handles and b holes. For example, $(g, b) = (0, 1)$ is a disk, $(0, 2)$ is an annulus, and $(0, 3)$ is a pair of pants.

To describe unoriented surfaces, it is useful to introduce the cross-cap: cut a hole in the surface and identify diametrically opposite points. In more detail, take a complex coordinate z, cut out a disk of slightly less than unit radius, and glue the opposite points together by defining pairs of points with $z' = -1/\bar{z}$ to be equivalent. The resulting surface is unoriented because the gluing is antiholomorphic. Also, unlike the case of a boundary, the cross-cap introduces no edge or other local feature: the edge introduced by the cutting is only the boundary of a coordinate patch. Any compact connected closed surface, oriented or unoriented, can be obtained by adding g handles and c cross-caps to the sphere; any compact connected surface can be obtained by adding g handles, b boundaries, and c cross-caps to the sphere. Actually, these descriptions

are somewhat redundant: in either the case with boundary or the case without, one obtains every topology exactly once by restricting the sum to let only one of g and c be nonzero. For example, $(g, b, c) = (0, 0, 1)$ is the projective plane, $(g, b, c) = (0, 1, 1)$ is the Möbius strip, $(g, b, c) = (0, 0, 2)$ is the Klein bottle, and a torus with cross-cap can be obtained either as $(g, b, c) = (0, 0, 3)$ or as $(g, b, c) = (1, 0, 1)$, trading two cross-caps for a handle. The Euler number is given by

$$\chi = 2 - 2g - b - c \,. \tag{3.5.6}$$

The number of distinct topologies is very small compared to the number of distinct Feynman graphs at a given order in field theory. For example, in the closed oriented theory, there is exactly one topology at each order of perturbation theory. In a field theory, the number of graphs grows factorially with the order. The single string graph contains all the field theory graphs. In various limits, where handles become long compared to their circumferences, we can approximate the handles by lines and the string graph in that limit is approximated by the corresponding Feynman graph.

3.6 Vertex operators

Using the state–operator mapping, the vertex operator for the closed string tachyon is

$$\begin{aligned} V_0 &= 2g_c \int d^2\sigma \, g^{1/2} \, e^{ik\cdot X} \\ &\to g_c \int d^2z \; :e^{ik\cdot X}: \,. \end{aligned} \tag{3.6.1}$$

We have now included a factor g_c in the mapping. This is the closed string coupling constant, from adding an extra string to the process: we use the normalization of the vertex operator as the definition of the coupling. In the second line we have gone to the flat world-sheet. The vertex operator must be (diff×Weyl)-invariant, so in particular it must be conformally invariant on the flat world-sheet. To offset the transformation of d^2z, the operator must be a tensor of weight $(1, 1)$. By a straightforward OPE calculation, $e^{ik\cdot X}$ is a tensor of weight $h = \tilde{h} = \alpha' k^2/4$, so the condition is

$$m^2 = -k^2 = -\frac{4}{\alpha'} \,. \tag{3.6.2}$$

This is precisely the mass found in the light-cone quantization.

Similarly, the tensor states at the first excited level of the closed string

have the flat world-sheet vertex operators

$$\frac{2g_c}{\alpha'}\int d^2z\ :\partial X^\mu \bar{\partial} X^\nu e^{ik\cdot X}: . \tag{3.6.3}$$

The normalization relative to the tachyon vertex operator follows from the state–operator mapping, and we will see in chapter 6 that the same relative normalization is obtained from unitarity of the S-matrix. The weight is

$$h = \tilde{h} = 1 + \frac{\alpha' k^2}{4} , \tag{3.6.4}$$

so these are massless, again agreeing with the light-cone quantization. There are further conditions in order that this be a tensor, as we will work out below.

Vertex operators in the Polyakov approach

We have in the state–operator mapping a systematic means of writing the flat world-sheet vertex operator for any state. Since the world-sheet can always be made locally flat, this is in principle all we need. For perspective, however, it is useful to study the curved world-sheet vertex operators in the Polyakov formalism. We will do this in the remainder of the section. The method we use will be somewhat clumsy and not as systematic as the state–operator mapping, but some of the results are useful. The reader need not work out all the details.

Any operator to be included in the Polyakov path integral (3.5.5) must respect the local diff×Weyl symmetry of the theory. We have not been specific in the first line of eq. (3.6.1) as to how the operator is to be defined on the curved world-sheet. As in the earlier discussion of the Weyl anomaly, it is convenient to use a method that preserves diff invariance automatically, and to check the Weyl transformation by hand. For most purposes *dimensional regularization* is used in the literature. However, we will not have extensive need for this and so will not introduce it; rather, we will generalize the normal ordering introduced earlier. Define a renormalized operator $[\mathscr{F}]_r$ by

$$[\mathscr{F}]_r = \exp\left(\frac{1}{2}\int d^2\sigma d^2\sigma' \Delta(\sigma,\sigma')\frac{\delta}{\delta X^\mu(\sigma)}\frac{\delta}{\delta X_\mu(\sigma')}\right)\mathscr{F} . \tag{3.6.5}$$

Here

$$\Delta(\sigma,\sigma') = \frac{\alpha'}{2}\ln d^2(\sigma,\sigma') , \tag{3.6.6}$$

where $d(\sigma,\sigma')$ is the geodesic distance between points σ and σ'. As with normal ordering, the expression (3.6.5) instructs us to sum over all ways of contracting pairs in $\mathscr{F}$ using $\Delta(\sigma,\sigma')$. On the flat world-sheet, $d^2(\sigma,\sigma') = |z-z'|^2$ and this reduces to conformal normal ordering, which we have studied in detail. On a curved world-sheet it cancels the singular part of the self-contractions of fields in $\mathscr{F}$. Diff invariance is obvious, but the contraction depends on the metric and so introduces the Weyl dependence

$$\delta_W[\mathscr{F}]_r = [\delta_W\mathscr{F}]_r + \frac{1}{2}\int d^2\sigma d^2\sigma' \delta_W\Delta(\sigma,\sigma')\frac{\delta}{\delta X^\mu(\sigma)}\frac{\delta}{\delta X_\mu(\sigma')}[\mathscr{F}]_r , \tag{3.6.7}$$

the first term being the explicit Weyl dependence in the operator.

A vertex operator for a state of momentum k^μ must transform under translations in the same way as the state, and so will be of the form $e^{ik\cdot X}$ times derivatives of X^μ. Operators with different numbers of derivatives will not mix under Weyl transformations, so we can start by considering zero derivatives, the operator (3.6.1). The Weyl variation comes from the explicit factor of $g^{1/2}$ as well as from the renormalization as given in eq. (3.6.7),

$$\delta_W V_0 = 2g_c\int d^2\sigma\, g^{1/2}\left(2\delta\omega(\sigma) - \frac{k^2}{2}\delta_W\Delta(\sigma,\sigma)\right)[e^{ik\cdot X(\sigma)}]_r . \tag{3.6.8}$$

At short distance,

$$d^2(\sigma,\sigma') \approx (\sigma-\sigma')^2\exp(2\omega(\sigma)) \tag{3.6.9}$$

and so

$$\Delta(\sigma,\sigma') \approx \alpha'\omega(\sigma) + \frac{\alpha'}{2}\ln(\sigma-\sigma')^2 . \tag{3.6.10}$$

The Weyl variation is nonsingular in the limit $\sigma' \to \sigma$,

$$\delta_W\Delta(\sigma,\sigma) = \alpha'\delta\omega(\sigma) . \tag{3.6.11}$$

The condition for the Weyl variation (3.6.8) to vanish is then $k^2 = 4/\alpha'$, as already deduced above.

Off-shell amplitudes?

Observe that a naive attempt to define off-shell amplitudes, by taking k^μ off the mass shell, is inconsistent with the local world-sheet symmetries. In position space this means that a local probe of the string world-sheet, given by inserting

$$\delta^D(X(\sigma) - x_0) = \int \frac{d^Dk}{(2\pi)^D}\exp[ik\cdot(X(\sigma) - x_0)] \tag{3.6.12}$$

into the path integral, is inconsistent because it involves all momenta.

From several points of view this is not surprising. First, the argument given above that we could use pointlike sources (vertex operators) for

string states required a limiting process that essentially restricts us to on-shell questions. Second, string theory contains gravity, and in general relativity simple off-shell amplitudes do not exist. This is because we have to specify the location of the probe, but the coordinates are unphysical and coordinate-invariant off-shell quantities are much more complicated. Third, in string theory we do not have the freedom to introduce additional fields to measure local observables (analogous to the way electroweak processes are used to probe strongly interacting systems): we must use the strings themselves, or other objects that we will discuss (D-branes and solitons) that are intrinsic to the theory.

One can define off-shell amplitudes, at least in perturbation theory, if one fixes a gauge (light-cone gauge being simplest for this) and uses stringy sources. The S-matrix formula (3.5.5) can then be derived in a way more akin to the reduction formula from quantum field theory, where it is possible to define finite-time transition amplitudes and then take the infinite-time limit. Incidentally, the reader who is trying to relate the present discussion to field theory should note that the path integral expression (3.5.5) is not analogous to a Green's function but will in fact have the properties of an S-matrix element. The analogous object in field theory is the Green's function with its external propagators amputated and its external momenta restricted to the mass shell.

The discussion of the local probe (3.6.12) also shows the problem with a contact interaction between strings. The simplest form for such an interaction would be

$$\int_M d^2\sigma\, g(\sigma)^{1/2} \int_M d^2\sigma'\, g(\sigma')^{1/2}\, \delta^D(X(\sigma) - X(\sigma')) \,, \tag{3.6.13}$$

which is nonvanishing whenever the world-sheet self-intersects. However, the delta function involves all momenta, as in eq. (3.6.12), so this is not Weyl-invariant. The structure of string theory is quite rigid: strings can couple to strings only in the way we have described at the beginning of this chapter.

Massless closed string vertex operators

It is interesting, if a little messy, to carry the Polyakov treatment of vertex operators one level further. There is no way to make a world-sheet scalar with exactly one derivative, so the next case is two derivatives. The diff-invariant possibilities are

$$V_1 = \frac{g_c}{\alpha'} \int d^2\sigma\, g^{1/2} \Big\{ (g^{ab} s_{\mu\nu} + i\epsilon^{ab} a_{\mu\nu}) [\, \partial_a X^\mu \partial_b X^\nu e^{ik\cdot X} \,]_r + \alpha' \phi R\, [\, e^{ik\cdot X} \,]_r \Big\} \,, \tag{3.6.14}$$

where $s_{\mu\nu}$, $a_{\mu\nu}$, and ϕ are respectively a symmetric matrix, an antisymmetric matrix, and a constant. The antisymmetric *tensor* ϵ^{ab} is normalized $g^{1/2}\epsilon^{12} = 1$, or $g^{1/2}\epsilon^{z\bar{z}} = -i$. The i accompanying the antisymmetric tensor in the vertex operator can be understood as arising from the Euclidean continuation, because this term necessarily has an odd number of time derivatives (one) . This result applies generally to Euclidean actions and vertex operators.

To extend the analysis of Weyl invariance we need to work out the Weyl dependence (3.6.11) of the geodesic distance to higher order. One finds

$$\partial_a \delta_W \Delta(\sigma,\sigma')\Big|_{\sigma'=\sigma} = \frac{1}{2}\alpha' \partial_a \delta\omega(\sigma) \,, \tag{3.6.15a}$$

$$\partial_a \partial'_b \delta_W \Delta(\sigma,\sigma')\Big|_{\sigma'=\sigma} = \frac{1+\gamma}{2}\alpha' \nabla_a \partial_b \delta\omega(\sigma) \,, \tag{3.6.15b}$$

$$\partial_a \partial_b \delta_W \Delta(\sigma,\sigma')\Big|_{\sigma'=\sigma} = -\frac{\gamma}{2}\alpha' \nabla_a \partial_b \delta\omega(\sigma) \,. \tag{3.6.15c}$$

Here, $\gamma = -\frac{2}{3}$, but we have left it as a parameter for later reference; the third equation can be obtained from the second and the gradient of the first. Using eq. (3.3.5) for the variation of the curvature and eqs. (3.6.7) and (3.6.15) for the variation of the renormalization gives

$$\delta_W V_1 = \frac{g_c}{2} \int d^2\sigma\, g^{1/2} \delta\omega \Big\{ (g^{ab} S_{\mu\nu} + i\epsilon^{ab} A_{\mu\nu}) [\, \partial_a X^\mu \partial_b X^\nu e^{ik\cdot X} \,]_r + \alpha' F\, [\, e^{ik\cdot X} \,]_r \Big\} \,, \tag{3.6.16}$$

where

$$S_{\mu\nu} = -k^2 s_{\mu\nu} + k_\nu k^\omega s_{\mu\omega} + k_\mu k^\omega s_{\nu\omega} - (1+\gamma) k_\mu k_\nu s^\omega{}_\omega + 4 k_\mu k_\nu \phi \,, \tag{3.6.17a}$$

$$A_{\mu\nu} = -k^2 a_{\mu\nu} + k_\nu k^\omega a_{\mu\omega} - k_\mu k^\omega a_{\nu\omega} \,, \tag{3.6.17b}$$

$$F = (\gamma - 1) k^2 \phi + \frac{1}{2}\gamma k^\mu k^\nu s_{\mu\nu} - \frac{1}{4}\gamma(1+\gamma) k^2 s^\nu{}_\nu \,. \tag{3.6.17c}$$

In deriving the Weyl variation (3.6.16) we have integrated by parts and also used the relation

$$[\, \nabla^2 X^\mu e^{ik\cdot X} \,]_r = i\frac{\alpha'\gamma}{4} k^\mu R\, [\, e^{ik\cdot X} \,]_r \,. \tag{3.6.18}$$

The left-hand side would vanish by the naive equation of motion, but here the equation of motion multiplies another operator at the same point. The general principle that holds in this case is that the operator need not vanish but it is not independent — it can be expanded in terms of the other local operators in the theory. In general the coefficients of these operators depend on the renormalization scheme; the result (3.6.18) can be obtained by taking the Weyl variation of each side.

Since $S_{\mu\nu}$, $A_{\mu\nu}$, and F multiply independent operators, the condition for Weyl invariance is

$$S_{\mu\nu} = A_{\mu\nu} = F = 0 \,. \tag{3.6.19}$$

In order to get an accurate count of the number of operators, we must note that not all V_1 of the form (3.6.14) are independent. Rather, under

$$s_{\mu\nu} \to s_{\mu\nu} + \xi_\mu k_\nu + k_\mu \xi_\nu \,, \tag{3.6.20a}$$

$$a_{\mu\nu} \to a_{\mu\nu} + \zeta_\mu k_\nu - k_\mu \zeta_\nu \,, \tag{3.6.20b}$$

$$\phi \to \phi + \frac{\alpha' \gamma}{2} k \cdot \xi \,, \tag{3.6.20c}$$

the change in V_1 integrates to zero using the equation of motion (3.6.18). Choose now a vector n^μ satisfying $n \cdot k = 1$ and $n^2 = 0$. A complete set of independent operators is obtained by restricting to

$$n^\mu s_{\mu\nu} = n^\mu a_{\mu\nu} = 0 \,. \tag{3.6.21}$$

This is $2D$ equations for the $2D$ parameters ξ_μ and ζ_μ; one can show that one equation and one parameter are trivial, but that eqs. (3.6.21) are otherwise nondegenerate and so define an independent set of vertex operators.

By using condition (3.6.21) and solving first $S_{\mu\nu} n^\mu n^\nu = 0$, then $S_{\mu\nu} n^\mu = A_{\mu\nu} n^\mu = 0$, and finally the full $S_{\mu\nu} = A_{\mu\nu} = F = 0$, one finds that

$$k^2 = 0 \,, \tag{3.6.22a}$$

$$k^\nu s_{\mu\nu} = k^\mu a_{\mu\nu} = 0 \,, \tag{3.6.22b}$$

$$\phi = \frac{1+\gamma}{4} s^\mu_{\ \mu} \,. \tag{3.6.22c}$$

Again we find a mass-shell condition (3.6.22a), this time corresponding to the first excited states of the closed string. There is also the condition (3.6.22b) that the polarization be transverse to the momentum, as required for a physical polarization of a massless tensor field. In the frame in which

$$k^\mu = (1,1,0,0,\ldots,0) \,, \quad n^\mu = \frac{1}{2}(-1,1,0,0,\ldots,0) \,, \tag{3.6.23}$$

conditions (3.6.21) and (3.6.22b) imply that all 0- and 1-components vanish, while eq. (3.6.22c) fixes ϕ. This leaves precisely $(D-2)^2$ independent operators, the same number of massless closed string states as found in the light-cone gauge.

The condition $n^\mu s_{\mu\nu} = n^\mu a_{\mu\nu} = 0$ is needed to remove zero-norm light-like polarizations. However, the introduction of the vector n^μ spoils manifest Lorentz invariance, and the theory is Lorentz-invariant only because different choices of n^μ are equivalent due to (3.6.20). The equivalence relations (3.6.20) are therefore essential in order for the theory to be Lorentz-invariant with a positive Hilbert space norm. These relations are the signatures of local spacetime symmetries. In particular, ξ_μ is an infinitesimal *spacetime* coordinate transformation, and ζ_μ is the local spacetime symmetry of the antisymmetric tensor field. We now see in the interacting theory the existence of local spacetime symmetries, which we argued in section 1.4 had to be present because of the existence of massless spin-1 and spin-2 fields.

There is a technical point we should mention. Different renormalization schemes assign different names to the renormalized operators. The most common scheme is dimensional regularization; the operators in this scheme are related to those in the scheme above by

$$[e^{ik\cdot X}]_{\rm DR} = [e^{ik\cdot X}]_{\rm r} \,, \tag{3.6.24a}$$

$$[\partial_a X^\mu e^{ik\cdot X}]_{\rm DR} = [\partial_a X^\mu e^{ik\cdot X}]_{\rm r} \,, \tag{3.6.24b}$$

$$[\partial_a X^\mu \partial_b X^\nu e^{ik\cdot X}]_{\rm DR} = [\partial_a X^\mu \partial_b X^\nu e^{ik\cdot X}]_{\rm r} - \frac{\alpha'}{12} g_{ab} \eta^{\mu\nu} R[e^{ik\cdot X}]_{\rm r} \,, \tag{3.6.24c}$$

$$[\nabla_a \partial_b X^\mu e^{ik\cdot X}]_{\rm DR} = [\nabla_a \partial_b X^\mu e^{ik\cdot X}]_{\rm r} + i \frac{\alpha'}{12} g_{ab} k^\mu R[e^{ik\cdot X}]_{\rm r} \,. \tag{3.6.24d}$$

Comparing with eq. (3.6.18), we see that in dimensional regularization $\gamma = 0$, so the equations of motion hold inside the regulated operator. This is a convenience — many equations simplify. In particular, in the spacetime gauge transformation (3.6.20), ϕ is invariant.

Open string vertex operators

The extension to the open string is straightforward. We leave the details for the reader as exercise 3.9, giving only some results. The tachyon vertex operator is

$$g_{\rm o} \int_{\partial M} ds \, [e^{ik\cdot X}]_{\rm r} \tag{3.6.25}$$

and is Weyl-invariant for $k^2 = 1/\alpha'$. The photon vertex operator is

$$-i \frac{g_{\rm o}}{(2\alpha')^{1/2}} e_\mu \int_{\partial M} ds \, [\dot{X}^\mu e^{ik\cdot X}]_{\rm r} \,, \tag{3.6.26}$$

where the normalization relative to the tachyon is obtained either from the state–operator mapping or unitarity, the i is as discussed for the closed string vertex operator (3.6.14), and the sign is a matter of convention. This operator is Weyl-invariant if $k^2 = 0$ and $k \cdot e = 0$. There is an equivalence $e_\mu \cong e_\mu + \lambda k_\mu$, which is a spacetime gauge transformation. This leaves $D-2$ transverse polarizations.

3.7 Strings in curved spacetime

Now we are ready to consider an important new issue, strings moving in a curved spacetime. Let us recall first that the point-particle action has a natural extension to curved spacetime. Replacing the flat metric $\eta_{\mu\nu}$ with a general metric $G_{\mu\nu}(x)$ gives the action

$$S_{\rm pp} = \frac{1}{2}\int d\tau \left(\eta^{-1} G_{\mu\nu}(X)\dot{X}^\mu \dot{X}^\nu - \eta m^2\right) . \tag{3.7.1}$$

After eliminating η this becomes the invariant proper time along the particle world-line. Its variation is well known to give the geodesic equation, representing the motion of a particle in a gravitational field.

Making the same replacement in the Polyakov action gives

$$S_\sigma = \frac{1}{4\pi\alpha'}\int_M d^2\sigma\, g^{1/2} g^{ab} G_{\mu\nu}(X)\partial_a X^\mu \partial_b X^\nu . \tag{3.7.2}$$

This is a natural guess, but the reader might make the following objection. We have learned that the graviton itself is one state of the string. A curved spacetime is, roughly speaking, a coherent background of gravitons, and therefore in string theory it is a coherent state of strings. Merely putting a curved metric in the action (3.7.2) doesn't seem stringy enough. To see why it is justified, consider a spacetime that is close to flat,

$$G_{\mu\nu}(X) = \eta_{\mu\nu} + \chi_{\mu\nu}(X) \tag{3.7.3}$$

with $\chi_{\mu\nu}$ small. The integrand in the world-sheet path integral is

$$\exp(-S_\sigma) = \exp(-S_{\rm P})\left(1 - \frac{1}{4\pi\alpha'}\int_M d^2\sigma\, g^{1/2} g^{ab}\chi_{\mu\nu}(X)\partial_a X^\mu \partial_b X^\nu + \ldots\right) . \tag{3.7.4}$$

The term of order χ is precisely the vertex operator for the graviton state of the string, eq. (3.6.14) with

$$\chi_{\mu\nu}(X) = -4\pi g_{\rm c} e^{ik\cdot X} s_{\mu\nu} . \tag{3.7.5}$$

So indeed, inserting the curved metric does seem to make contact with what we already know. The action (3.7.2) can be thought of as describing a coherent state of gravitons by exponentiating the graviton vertex operator.

This suggests a natural generalization: include backgrounds of the other massless string states as well. From the form of the vertex operators, eq. (3.6.14), one has

$$S_\sigma = \frac{1}{4\pi\alpha'}\int_M d^2\sigma\, g^{1/2}\left[\left(g^{ab}G_{\mu\nu}(X) + i\epsilon^{ab}B_{\mu\nu}(X)\right)\partial_a X^\mu \partial_b X^\nu + \alpha' R\Phi(X)\right] . \tag{3.7.6}$$

The field $B_{\mu\nu}(X)$ is the antisymmetric tensor, and the dilaton involves both Φ and the diagonal part of $G_{\mu\nu}$, as implied by the equation of motion (3.6.22c).

As a check, let us see that the spacetime gauge invariances are respected. Under a change of variables in the path integral corresponding to a field redefinition $X'^\mu(X)$, the action (3.7.6) is invariant with $G_{\mu\nu}$ and $B_{\mu\nu}$ transforming as tensors and Φ as a scalar. From the spacetime point of view, this is a coordinate transformation. The action is also invariant under

$$\delta B_{\mu\nu}(X) = \partial_\mu \zeta_\nu(X) - \partial_\nu \zeta_\mu(X) , \tag{3.7.7}$$

which adds a total derivative to the Lagrangian density. This is a generalization of the electromagnetic gauge transformation to a potential with two antisymmetric indices. The three-index field strength

$$H_{\omega\mu\nu} = \partial_\omega B_{\mu\nu} + \partial_\mu B_{\nu\omega} + \partial_\nu B_{\omega\mu} \tag{3.7.8}$$

is invariant. There is a similar generalization to an antisymmetric n-tensor potential, which plays an important role in the superstring and will be developed in volume two.

The theory thus depends only on gauge-invariant objects built out of the metric and other fields. One can think of the fields X^μ as being coordinates on a manifold. This is called the *target space*, because the X^μ define an embedding,

$$\text{world-sheet} \rightarrow \text{target} . \tag{3.7.9}$$

In string theory, the target space is spacetime itself. A field theory such as (3.7.2), in which the kinetic term is field-dependent and so field space is effectively a curved manifold, is known for historic reasons as a *nonlinear sigma model*. Nonlinear sigma models have had many applications in particle physics and quantum field theory. For example, the neutral and charged pion fields can be regarded to good approximation as coordinates on the group manifold $SU(2)$.

The nonlinear sigma model action is no longer quadratic in X^μ, so the path integral is now an *interacting* two-dimensional quantum field theory. Expand the path integral around the classical solution $X^\mu(\sigma) = x_0^\mu$ for any chosen point x_0^μ. With $X^\mu(\sigma) = x_0^\mu + Y^\mu(\sigma)$,

$$G_{\mu\nu}(X)\partial_a X^\mu \partial_b X^\nu = \Big[G_{\mu\nu}(x_0) + G_{\mu\nu,\omega}(x_0)Y^\omega + \tfrac{1}{2}G_{\mu\nu,\omega\rho}(x_0)Y^\omega Y^\rho + \ldots\Big]\partial_a Y^\mu \partial_b Y^\nu . \tag{3.7.10}$$

The first term in this expansion is a quadratic kinetic term for the field Y^μ. The next term is a cubic interaction, and so forth. The coupling constants in the expansion, $G_{\mu\nu,\omega}(x_0)$ and so on, involve derivatives of

the metric at x_0. In a target space with characteristic radius of curvature R_c, derivatives of the metric are of order R_c^{-1}. The effective dimensionless coupling in the theory is therefore $\alpha'^{1/2}R_c^{-1}$. If the radius of curvature R_c is much greater than the characteristic length scale of the string, then this coupling is small and perturbation theory in the two-dimensional field theory is a useful tool. Notice that in this same regime we can employ another tool as well. Since the characteristic length scale is long compared to the string, we can ignore the internal structure of the string and use *low energy effective field theory*. String theory enters in determining the effective low energy action for the fields, as we will soon see. We have also implicitly used $\alpha'^{1/2}R_c^{-1} \ll 1$ when restricting our attention to *massless* backgrounds: when wavelengths are long compared to the string scale, massive string states are not created.

The nonlinear sigma model (3.7.2) is a renormalizable theory: the fields Y^μ have dimension zero and so the interactions all have dimension two. Nevertheless the couplings, which are the coefficients in the expansion of the metric, are infinite in number (unless some symmetry restricts the form of the metric). In fact it is more useful to think of the metric function itself, rather than its expansion coefficients, as defining the couplings of the theory.

Weyl invariance

We have emphasized that Weyl invariance is essential to the consistency of string theory. The action (3.7.6) will define a consistent string theory only if the two-dimensional quantum field theory is Weyl-invariant. This action happens to be the most general action that is classically invariant under a *rigid* Weyl transformation, with $\delta\omega$ independent of σ. This is easy to see: under a rigid Weyl transformation, the scaling of a term with n derivatives will be proportional to $2-n$. Of course, it is necessary also to consider local Weyl transformations, under which the $G_{\mu\nu}$ and $B_{\mu\nu}$ terms in the action (3.7.6) are classically invariant while the Φ term is not, and to include the quantum contributions to the Weyl transformation properties. In the limit that $B_{\mu\nu}$ and Φ are small, and $G_{\mu\nu}$ is close to $\eta_{\mu\nu}$, we can use the results of the previous section to find the Weyl transformation. Write $S_\sigma = S_P - V_1 + \ldots$, with S_P the flat-space action and V_1 the vertex operator (3.6.14). Then

$$G_{\mu\nu}(X) = \eta_{\mu\nu} - 4\pi g_c s_{\mu\nu} e^{ik\cdot X} , \tag{3.7.11a}$$

$$B_{\mu\nu}(X) = -4\pi g_c a_{\mu\nu} e^{ik\cdot X} , \tag{3.7.11b}$$

$$\Phi(X) = -4\pi g_c \phi e^{ik\cdot X} . \tag{3.7.11c}$$

We could of course take linear combinations of vertex operators with different momenta. To first order the Weyl variation is then given in eq. (3.6.16); we take for convenience a renormalization scheme in which $\gamma = 0$. Relating this Weyl transformation to $T^a{}_a$ as in eq. (3.4.6) gives

$$T^a{}_a = -\frac{1}{2\alpha'}\beta^G_{\mu\nu} g^{ab}\partial_a X^\mu \partial_b X^\nu - \frac{i}{2\alpha'}\beta^B_{\mu\nu}\epsilon^{ab}\partial_a X^\mu \partial_b X^\nu - \frac{1}{2}\beta^\Phi R , \tag{3.7.12}$$

where, to linear order in $\chi_{\mu\nu}$, $B_{\mu\nu}$ and Φ,

$$\beta^G_{\mu\nu} \approx -\frac{\alpha'}{2}\left(\partial^2\chi_{\mu\nu} - \partial_\nu\partial^\omega\chi_{\mu\omega} - \partial_\mu\partial^\omega\chi_{\omega\nu} + \partial_\mu\partial_\nu\chi^\omega{}_\omega\right) + 2\alpha'\partial_\mu\partial_\nu\Phi , \tag{3.7.13a}$$

$$\beta^B_{\mu\nu} \approx -\frac{\alpha'}{2}\partial^\omega H_{\omega\mu\nu} , \tag{3.7.13b}$$

$$\beta^\Phi \approx \frac{D-26}{6} - \frac{\alpha'}{2}\partial^2\Phi . \tag{3.7.13c}$$

We have included in β^Φ the flat spacetime anomaly found in section 3.4, including the contribution proportional to 26 from the ghost fields. The symbol β is used for the coefficients in $T^a{}_a$ because these are essentially the renormalization group beta functions, governing the dependence of the physics on world-sheet scale. We will discuss this connection further, though it does not come up in detail until chapter 15.

The Weyl anomalies (3.7.13) have further contributions from higher orders in the fields. For example, expanding the path integral to second order in the cubic interaction from the expansion (3.7.10) will give rise to divergences when the two interaction vertices approach one another. Their OPE includes a singularity of the form $|z|^{-2}\partial X^\mu \bar\partial X^\nu$ times the square of the derivative of $G_{\alpha\beta}$. Integrating d^2z produces a logarithmic divergence. Diff invariance requires that the integral be cut off in terms of the invariant distance, $|z|\exp(\omega) > a_0$. This introduces a dependence on the scale of the metric, $\ln z_{\min} = -\omega + \ln a_0$. The Weyl variation $\beta^G_{\mu\nu}$ thus gets a contribution proportional to $O(G_{\alpha\beta,\gamma})^2$ which combines with the linear second derivative term to form the spacetime Ricci tensor.

We will not work this out in detail, but quote the result keeping all terms with up to two spacetime derivatives

$$\beta^G_{\mu\nu} = \alpha' R_{\mu\nu} + 2\alpha'\nabla_\mu\nabla_\nu\Phi - \frac{\alpha'}{4}H_{\mu\lambda\omega}H_\nu{}^{\lambda\omega} + O(\alpha'^2) , \tag{3.7.14a}$$

$$\beta^B_{\mu\nu} = -\frac{\alpha'}{2}\nabla^\omega H_{\omega\mu\nu} + \alpha'\nabla^\omega\Phi H_{\omega\mu\nu} + O(\alpha'^2) , \tag{3.7.14b}$$

$$\beta^\Phi = \frac{D-26}{6} - \frac{\alpha'}{2}\nabla^2\Phi + \alpha'\nabla_\omega\Phi\nabla^\omega\Phi - \frac{\alpha'}{24}H_{\mu\nu\lambda}H^{\mu\nu\lambda} + O(\alpha'^2) . \tag{3.7.14c}$$

Several terms in (3.7.14) can be recognized from the linear approximation

(3.7.13), but now made covariant under a change of spacetime coordinates. The spacetime Ricci tensor is $R_{\mu\nu}$, distinguished from the world-sheet Ricci tensor R_{ab}. Terms with more derivatives are higher order in the world-sheet $\alpha'^{1/2}R_c^{-1}$ expansion. Incidentally, the most efficient method to carry the calculation of the Weyl anomaly to higher order is not the one we have described, but rather dimensional regularization.

The condition that the world-sheet theory be Weyl-invariant is thus

$$\beta^G_{\mu\nu} = \beta^B_{\mu\nu} = \beta^\Phi = 0 \,. \tag{3.7.15}$$

These are sensible-looking equations of motion. The equation $\beta^G_{\mu\nu} = 0$ resembles Einstein's equation with source terms from the antisymmetric tensor field and the dilaton. The equation $\beta^B_{\mu\nu} = 0$ is the antisymmetric tensor generalization of Maxwell's equation, determining the divergence of the field strength.

Backgrounds

One other qualitative feature of the field equations is readily understood: the field Φ always occurs differentiated, so there is an invariance under X^μ-independent shifts of Φ. This is because such a shift only changes the world-sheet action (3.7.6) by a term proportional to the Euler number, and so does not affect local properties like Weyl invariance. In particular, the background

$$G_{\mu\nu}(X) = \eta_{\mu\nu} \,, \quad B_{\mu\nu}(X) = 0 \,, \quad \Phi(X) = \Phi_0 \tag{3.7.16}$$

is exactly Weyl-invariant for any constant Φ_0. This is just the flat spacetime action (3.2.2), with

$$\lambda = \Phi_0 \,. \tag{3.7.17}$$

The value of λ determines the coupling strength between strings, but we now see that this does not mean that there are different string theories with different values of this parameter. Different values of λ correspond not to different *theories*, but to different *backgrounds* in a single theory. The other parameter appearing in the Nambu–Goto and Polyakov actions, the Regge slope α', is also not really a free parameter because it is dimensionful. It simply defines the unit of length, and can be absorbed in the definition of X^μ.

Changing the fields $G_{\mu\nu}$, $B_{\mu\nu}$, and Φ in the world-sheet action would seem, from the two-dimensional point of view, to give a new theory. From the point of view of the full string theory, one is merely looking at a different background — a different state — in the same theory. This is one of the features that makes string theory attractive. In the Standard Model, and in most attempts to unify the Standard Model within

quantum field theory, there are many constants that are not determined by the theory; any values for these constants give a consistent theory. In string theory there are no such free parameters — the coupling constants depend on the state and are determined in principle by the dynamics. Of course, for now this only moves the difficulty elsewhere, because we do not understand the dynamics well enough to know the values of the background fields.

It is striking that Einstein's equation turns up in what seems to be a rather out-of-the-way place, as the condition for Weyl invariance of a two-dimensional field theory. String theory provides the physical connection between these two ideas. It is also notable that, because of the need for Weyl invariance, strings can propagate consistently only in a background that satisfies appropriate field equations. This parallels our earlier discovery that only on-shell vertex operators make sense.

Another observation: the condition $D = 26$ came from the R term in T^a_a. In the nonlinear sigma model this is generalized to $\beta^\Phi = 0$. We see in eq. (3.7.14c) that the leading term in β^Φ is proportional to $D - 26$ but that there are corrections involving the gradients of the fields. Evidently we can have other values of D if the fields are not constant. This is true, although we cannot really conclude it from eq. (3.7.14c) because this has been derived in the approximation $\alpha'^{1/2}R_c^{-1} \ll 1$ where the correction terms in β^Φ are small. In fact, exact solutions with $D \neq 26$ are known. For now we give one simple example, saving others for later. This is

$$G_{\mu\nu}(X) = \eta_{\mu\nu} \,, \quad B_{\mu\nu}(X) = 0 \,, \quad \Phi(X) = V_\mu X^\mu \,. \tag{3.7.18}$$

The beta functions (3.7.14) vanish if

$$V_\mu V^\mu = \frac{26 - D}{6\alpha'} \,. \tag{3.7.19}$$

This result is actually *exact*, because the fields (3.7.18) are constant or linear in X^μ, so the world-sheet path integral remains Gaussian. Varying g_{ab} to determine T_{ab} for this theory, one finds that it is none other than the linear dilaton CFT (hence the name), whose central charge $c = D + 6\alpha' V_\mu V^\mu$ is indeed 26 under the condition (3.7.19). Because Φ must have a large gradient in order to change D, this CFT does not describe our roughly static and homogeneous four-dimensional spacetime, but it may have cosmological applications. This CFT is also important because the cases $D = 1$ and $D = 2$ are simple enough for the string theory to be exactly solvable, as we will discuss in section 9.9.

The spacetime action

The field equations (3.7.15) can be derived from the spacetime action

$$S = \frac{1}{2\kappa_0^2}\int d^D x\,(-G)^{1/2} e^{-2\Phi}\Big[-\frac{2(D-26)}{3\alpha'} + R - \frac{1}{12}H_{\mu\nu\lambda}H^{\mu\nu\lambda} + 4\partial_\mu\Phi\partial^\mu\Phi + O(\alpha')\Big] . \tag{3.7.20}$$

The normalization constant κ_0 is not determined by the field equations and has no physical significance since it can be changed by a redefinition of Φ. One can verify that

$$\delta S = -\frac{1}{2\kappa_0^2\alpha'}\int d^D x\,(-G)^{1/2} e^{-2\Phi}\Big[\delta G_{\mu\nu}\beta^{G\mu\nu} + \delta B_{\mu\nu}\beta^{B\mu\nu} + \Big(2\delta\Phi - \tfrac{1}{2}G^{\mu\nu}\delta G_{\mu\nu}\Big)(\beta^{G\omega}{}_{\omega} - 4\beta^{\Phi})\Big] . \tag{3.7.21}$$

The action is written in bold as a reminder that this is the effective action governing the low energy spacetime fields, unlike the world-sheet actions that have appeared everywhere previously in this book.

It is often useful to make a field redefinition of the form

$$\tilde{G}_{\mu\nu}(x) = \exp(2\omega(x))G_{\mu\nu}(x) , \tag{3.7.22}$$

which is a spacetime Weyl transformation. The Ricci scalar constructed from $\tilde{G}_{\mu\nu}$ is

$$\tilde{R} = \exp(-2\omega)\Big[R - 2(D-1)\nabla^2\omega - (D-2)(D-1)\partial_\mu\omega\partial^\mu\omega\Big] . \tag{3.7.23}$$

For the special case $D = 2$, this is the Weyl transformation (3.3.5). Let $\omega = 2(\Phi_0 - \Phi)/(D-2)$ and define

$$\tilde{\Phi} = \Phi - \Phi_0 , \tag{3.7.24}$$

which has vanishing expectation value. The action becomes

$$S = \frac{1}{2\kappa^2}\int d^D X\,(-\tilde{G})^{1/2}\Big[-\frac{2(D-26)}{3\alpha'}e^{4\tilde{\Phi}/(D-2)} + \tilde{R} - \frac{1}{12}e^{-8\tilde{\Phi}/(D-2)}H_{\mu\nu\lambda}\tilde{H}^{\mu\nu\lambda} - \frac{4}{D-2}\partial_\mu\tilde{\Phi}\tilde{\partial}^\mu\tilde{\Phi} + O(\alpha')\Big] , \tag{3.7.25}$$

where tildes have been inserted as a reminder that indices here are raised with $\tilde{G}^{\mu\nu}$. In terms of $\tilde{G}_{\mu\nu}$, the gravitational Lagrangian density takes the standard Hilbert form $(-\tilde{G})^{1/2}\tilde{R}/2\kappa^2$. The constant $\kappa = \kappa_0 e^{\Phi_0}$ is the observed gravitational coupling constant,

$$\kappa = (8\pi G_N)^{1/2} = \frac{(8\pi)^{1/2}}{M_P} = (2.43\times 10^{18}\ \text{GeV})^{-1} . \tag{3.7.26}$$

Commonly, $G_{\mu\nu}$ is called the *sigma model metric* or *string metric*, and $\tilde{G}_{\mu\nu}$ the *Einstein metric*. Because of the force from dilaton exchange, there is no equivalence principle and so no way to single out a preferred definition of the metric. Looking ahead, there are higher order effects in string theory that can give the dilaton a mass and so a finite range — a good thing, because the equivalence principle does of course hold to good accuracy.

Note that the dilaton appears in the action (3.7.20) in an overall factor $e^{-2\Phi}$ and otherwise is always differentiated. This is consistent with the fact that adding an overall constant to the dilaton has no effect on the Weyl anomaly. From the spacetime point of view this multiplies the action by a constant, which does not change the equations of motion. It also reflects the fact that in quantum field theory it is possible to rescale the fields in such a way that the coupling g appears only as an overall factor of g^{-2}, which in string theory is $e^{-2\Phi}$. For example, the Yang–Mills action is usually introduced in the form

$$S = -\frac{1}{4}\int d^D x\,\mathrm{Tr}\,(F_{\mu\nu}F^{\mu\nu}) , \tag{3.7.27a}$$

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu - ig[A_\mu, A_\nu] \tag{3.7.27b}$$

(matrix notation). By defining

$$gA_\mu = A'_\mu , \quad gF_{\mu\nu} = F'_{\mu\nu} , \tag{3.7.28}$$

g is removed from the field strength, covariant derivative, and gauge transformation and appears only in an overall g^{-2} in the action. In string theory this holds when the action is written in terms of the fields appearing on the string world-sheet. If we make a dilaton-dependent field redefinition as in going to the form (3.7.25) then it no longer holds.

When g appears only in the normalization of the action, every propagator contains a factor of g^2 and every interaction a factor of g^{-2}. It is not hard to show that an L-loop contribution to the effective action has exactly $L-1$ more propagators than vertices and so scales as $g^{2(L-1)}$. In string theory this is $e^{-\chi\Phi}$, as we would expect.

The discussion of backgrounds extends to open strings. We have seen that open string vertex operators are integrated along the boundary, so open string fields appear as boundary terms in the sigma model action.

Compactification and CFT

The four string theories studied thus far (with or without boundaries, oriented or unoriented) have a number of features in common. One good feature is the automatic inclusion of general relativity. There are also several bad features: the need for 26 dimensions, the existence of the tachyon, and the absence of fermions in the spectrum. The elimination

of the tachyon and the introduction of fermions will have to wait until volume two. However, the presence of general relativity provides a natural way to account for the extra dimensions. In general relativity the geometry of spacetime is dynamical, not fixed. Flat spacetime is only one of many solutions to the field equations. There may well be solutions in which some dimensions are large and flat and others are small and highly curved. The metric would be

$$g_{MN} = \begin{bmatrix} \eta_{\mu\nu} & 0 \\ 0 & g_{mn}(x^p) \end{bmatrix} . \tag{3.7.29}$$

We have divided the 26 coordinates $M, N = 0, \ldots, 25$ into four 'spacetime' coordinates $\mu, \nu = 0, \ldots, 3$ and 22 'internal' coordinates $m, n, p = 4, \ldots, 25$. The metric (3.7.29) is flat in the four spacetime dimensions and curved in the 22 internal dimensions. The internal dimensions are assumed to be compact. As one example, the background field equations (3.7.15) are satisfied to order α' if the dilaton field is constant, the antisymmetric tensor is zero, and the metric for the internal space is Ricci-flat, $R_{mn} = 0$.

In other words, spacetime is of the form $M^d \times K$, where M^d is d-dimensional Minkowski space and K is some $(26-d)$-dimensional compact Riemannian space.[4] In such a spacetime, the physics on length scales much longer than the size of K is the same as in a d-dimensional Minkowski space. One says that $26 - d$ dimensions have been *compactified*, though cosmologically it seems more appropriate to say that the large dimensions have decompactified. As we discussed briefly in the introduction, there are good reasons, independent of string theory, to consider the possibility that our spacetime might be of this form. The compact dimensions must be small enough not to have been observed. In fact, we will see in chapter 8, and in more detail in chapter 18, that the size of any extra dimensions must be within a few orders of magnitude of the Planck length, 1.6×10^{-32} cm.

We can also think about generalizing this idea further. The basic physical consistency conditions are few in number. We require Lorentz invariance, and we require that quantum mechanical probabilities be positive and conserved — to date, nothing in string theory seems to require that the standard quantum mechanical framework be modified, though it would be interesting to learn otherwise. Although these conditions are simple there is a tension between them, in gauge theory and gravitation as well as in string theory. In a manifestly covariant gauge there are negative norm timelike excitations that must decouple from physical amplitudes. In the light-cone gauge the inner product is positive but covariance is no longer manifest.

[4] We will generally use D to refer to the total dimension of a string theory and d when we may be talking about a subset of the world-sheet fields.

One necessary condition is therefore two-dimensional diff invariance, so that the unphysical oscillations are merely oscillations of the coordinate system. Weyl invariance is a more technical requirement: the extra degree of freedom that appears when Weyl invariance is lost is less intuitive than an oscillation of the coordinate system, but we have argued in section 3.4 that it is incompatible with Lorentz invariance.

We will also make an additional technical assumption: that the embedding time X^0 appears in the world-sheet action only in the term

$$-\frac{1}{4\pi\alpha'} \int_M d^2\sigma \, g^{1/2} g^{ab} \partial_a X^0 \partial_b X^0 . \tag{3.7.30}$$

For example, in the sigma model this means that the fields are static with $G_{0\mu} = \eta_{0\mu}$ and $B_{0\mu} = 0$. This assumption is not much of a restriction. The form (3.7.30) is sufficient if one is interested in static states. Time-dependent backgrounds are certainly of interest for cosmological and other reasons, but they can be analyzed using low energy effective field theory except in the extreme case that the time scale is of the order of the string scale. The reason for making the assumption (3.7.30) is to put the main problem, the wrong-sign nature of X^0, in a rather explicit form.

The requirement that the local invariances be maintained leads to the following proposal for a more general string theory: *Replace the 25 spatial X^μ fields with any unitary CFT having $c = \tilde{c} = 25$.* This ensures that the full two-dimensional theory, with X^0 and the ghosts, can be coupled to a curved metric in a (diff×Weyl)-invariant way. The unitary (positive inner product) condition is necessary because there is only enough local symmetry to remove the unphysical X^0 excitation; the world-sheet inner product is relevant here because it becomes the spacetime inner product in the one-particle sector. We will see in volume two that further generalization is possible, enlarging simultaneously the local symmetry group and the set of unphysical excitations.

If we are interested in a bosonic string theory whose low energy physics looks four-dimensional, we keep the free fields X^μ for $\mu = 0, 1, 2, 3$, and replace the remaining X^μ with a compact unitary CFT having $c = \tilde{c} = 22$. By compact, we mean that the spectrum is discrete, just as for a quantum field theory in a finite volume. There have been some efforts to construct all CFTs, or all CFTs having various additional properties. We will describe some of these in chapter 15. Incidentally, all of these theories have tachyons, the product of the unit state $|1\rangle$ from the internal CFT with the ground state $|0;k\rangle$ from the four-dimensional CFT.

How large a generalization have we just made? If the local symmetries are the only essential restriction, we are free to introduce fields on the world-sheet with various world-sheet and spacetime quantum numbers. This seems to take us far beyond the original theory of 26 X^μ fields. How-

ever, in two dimensions there are many equivalences between seemingly different quantum field theories, so that the situation is not clear.

It is generally assumed that all string theories based on different CFTs but the same world-sheet gauge symmetries and topology are different ground states of the same theory. This causes a semantic problem: we often refer to strings with different world-sheet Lagrangians as different *theories*, though they are most likely just different vacua in one theory. We will take this even further in chapter 14, seeing that *all* string theories appear to be vacua of a single theory.

As we develop the techniques of string theory in volume one, we will usually take the simplest case of flat 26-dimensional spacetime in explicit examples. However, many results will be developed in such a way as to apply to the very general string theory we have described above.

Exercises

3.1 (a) Evaluate the Euler number χ for a flat disk and for a disk with the metric of a hemisphere.
(b) Show that the Euler number of a sphere with b holes is $2-b$. You may take any convenient metric on the sphere.
(c) Extend this to the result $2-b-2g$ for the sphere with b holes and g handles.

3.2 (a) Show that an n-index traceless symmetric tensor in two dimensions has exactly two independent components. The traceless condition means that the tensor vanishes when any pair of indices is contracted with g^{ab}.
(b) Find a differential operator P_n that takes n-index traceless symmetric tensors into $(n+1)$-index traceless symmetric tensors.
(c) Find a differential operator P_n^T that does the reverse.
(d) For u an $(n+1)$-index traceless symmetric tensor and v an n-index traceless symmetric tensor, show that with appropriate normalization $(u, P_n v) = (P_n^T u, v)$. The inner product of traceless symmetric tensors of any rank is

$$(t,t') = \int d^2\sigma\, g^{1/2} t \cdot t' ,$$

the dot denoting contraction of all indices.

3.3 (a) Work out the form of the covariant derivative in conformal gauge. Show in particular that ∇_z reduces to ∂_z when acting on a tensor component with all $\bar{z}$ indices, and vice versa. Since raising or lowering reverses z and $\bar{z}$ indices, this means that it is always possible to work with simple derivatives by appropriately positioning the indices.
(b) Use this to give simple expressions for P_n and P_n^T.

3.4 To gain familiarity with the Faddeev–Popov procedure, it is useful to consider it in a more general and abstract context. This is described in the first page of section 4.2 (which is independent of section 4.1). Derive the general result (4.2.3).

3.5 (A continuation of the previous problem.) Consider two different gauge choices, $F^A(\phi) = 0$ and $G^A(\phi) = 0$. The Faddeev–Popov procedure shows that the gauge-fixed path integrals (4.2.3) are both equal to the original gauge-invariant path integral and so equal to each other. However, this passes through the somewhat heuristically defined gauge-invariant path integral. Adapt the Faddeev–Popov procedure to show directly the equivalence of the two gauge-fixed path integrals. This is what is relevant in practice, showing for example that manifestly covariant and manifestly unitary gauges define the same theory, which therefore has both properties. In chapter 4 we will show this equivalence by a more abstract method, BRST symmetry.

3.6 On dimensional grounds, a current that is conserved on a flat world-sheet has the possible anomaly

$$\nabla_a j^a = aR .$$

By comparing the Weyl transformations of the two sides, show that the sum of the coefficients of z^{-3} in Tj and of $\bar{z}^{-3}$ in $\tilde{T}\tilde{j}$ is $-4a$.

3.7 In our second calculation of the Weyl anomaly, we have in a roundabout way calculated a Feynman graph with a loop of scalars or ghosts and with two external gravitational fields. Do the calculation for scalars directly, in momentum space. You will need to use your favorite invariant regulator: dimensional, Pauli–Villars, or zeta function. Show that the result agrees with eq. (3.4.19). The linearized expression for the curvature can be conveniently read from eqs. (3.7.13a) and (3.7.14a). (Warning: the full calculation is a bit lengthy.)

3.8 Verify the Weyl transformation (3.6.16) of the vertex operator V_1.

3.9 Find the Weyl invariance conditions for the open string tachyon and photon vertex operators. It is simplest to take a coordinate system in which the boundary is a straight line. Recall from exercise 2.10 that the propagator includes an image charge term, so it is effectively doubled for fields directly on the boundary.

3.10 Find the general vertex operator and the conditions for Weyl invariance at the level $m^2 = 1/\alpha'$ of the open string. Find a complete set of independent Weyl-invariant operators and compare with the states found in the light-cone quantization. [Answer: there appears to one extra operator. In fact its amplitudes vanish, but we need the formalism of the next chapter to understand the reason, that it corresponds to a null state.]

3.11 Calculate the order H^2 term in $\beta^G_{\mu\nu}$ by the method suggested above

eq. (3.7.14). You need precisely the term $(z\bar{z})^{-1}$ in the OPE; more singular terms require a shift in the tachyon background, but this is an artifact of the bosonic string.

3.12 If $\beta^G_{\mu\nu} = \beta^B_{\mu\nu} = 0$, the nonlinear sigma model will be conformally invariant on a flat world-sheet. According to the discussion at the end of section 3.4, in this case β^Φ must be independent of X. Show this for the form (3.7.14); you will need the Bianchi identities.

3.13 Consider a spacetime with d flat dimensions and 3 dimensions in the shape of a 3-sphere. Let H be proportional to the completely antisymmetric tensor on the 3-sphere and let the dilaton be constant. Using the form (3.7.14) for the equations of motion, show that there are solutions with $d + 3 \neq 26$. These solutions are outside the range of validity of eq. (3.7.14), but we will see in chapter 15 that there are exact solutions of this form, though with H limited to certain quantized values.

3.14 Verify that the background (3.7.18) gives the linear dilaton energy-momentum tensor (2.5.1).

4

The string spectrum

4.1 Old covariant quantization

In conformal gauge, the world-sheet fields are X^μ and the Faddeev–Popov ghosts b_{ab} and c^a. The Hilbert space is bigger than the actual physical spectrum of the string: the D sets of α^μ oscillators include unphysical oscillations of the coordinate system, and there are ghost oscillators. As is generally the case in covariant gauges, there are negative norm states from the timelike oscillators (because the commutator is proportional to the spacetime metric $\eta_{\mu\nu}$), and also from the ghosts.

The actual physical space is smaller. To see how to identify this smaller space, consider the amplitude on the infinite cylinder for some initial state $|i\rangle$ to propagate to some final state $|f\rangle$. Suppose that we have initially used the local symmetries to fix the metric to a form $g_{ab}(\sigma)$. Consider now a different gauge, with metric $g_{ab}(\sigma) + \delta g_{ab}(\sigma)$. A physical amplitude should not depend on this choice. Of course, for a change in the metric we know how the path integral changes. From the definition (3.4.4) of T^{ab},

$$\delta\langle f|i\rangle = -\frac{1}{4\pi}\int d^2\sigma\, g(\sigma)^{1/2}\delta g_{ab}(\sigma)\langle f|T^{ab}(\sigma)|i\rangle \ . \tag{4.1.1}$$

In order that the variation vanish for arbitrary changes in the metric, we need

$$\langle\psi|T^{ab}(\sigma)|\psi'\rangle = 0 \tag{4.1.2}$$

for arbitrary physical states $|\psi\rangle$, $|\psi'\rangle$.

There is another way to think about this. The original equation of motion from variation of g_{ab} was $T^{ab} = 0$. After gauge-fixing, this does not hold as an operator equation: we have a missing equation of motion because we do not vary g_{ab} in the gauge-fixed theory. The condition (4.1.2) says that the missing equation of motion must hold for matrix elements between physical states. When we vary the gauge we must take into

account the change in the Faddeev–Popov determinant, so the energy-momentum tensor in the matrix element is the sum of the X^μ and ghost contributions,

$$T_{ab} = T^X_{ab} + T^{\rm g}_{ab} \,. \tag{4.1.3}$$

The X^μ may be replaced by a more general CFT (which we will refer to as the *matter* CFT), in which case

$$T_{ab} = T^{\rm m}_{ab} + T^{\rm g}_{ab} \,. \tag{4.1.4}$$

In the remainder of this section we will impose the condition (4.1.2) in a simple but somewhat *ad hoc* way. This is known as the *old covariant quantization (OCQ)*, and it is sufficient for many purposes. In the next section we will take a more systematic approach, BRST quantization. These are in fact equivalent, as will be shown in section 4.4.

In the *ad hoc* approach, we will simply ignore the ghosts and try to restrict the matter Hilbert space so that the missing equation of motion $T^{\rm m}_{ab} = 0$ holds for matrix elements. In terms of the Laurent coefficients this is $L^{\rm m}_n = 0$, and in the closed string also $\tilde{L}^{\rm m}_n = 0$. One might first try to require physical states to satisfy $L^{\rm m}_n|\psi\rangle = 0$ for all n, but this is too strong; acting on this equation with $L^{\rm m}_m$ and forming the commutator, one encounters an inconsistency due to the central charge in the Virasoro algebra. However, it is sufficient that only the Virasoro lowering and zero operator annihilate physical states,

$$(L^{\rm m}_n + A\delta_{n,0})|\psi\rangle = 0 \quad \text{for } n \geq 0 \,. \tag{4.1.5}$$

For $n < 0$ we then have

$$\langle\psi|L^{\rm m}_n|\psi'\rangle = \langle L^{\rm m}_{-n}\psi|\psi'\rangle = 0 \,. \tag{4.1.6}$$

We have used

$$L^{\rm m\dagger}_n = L^{\rm m}_{-n} \tag{4.1.7}$$

as follows from the Hermiticity of the energy-momentum tensor. The conditions (4.1.5) are consistent with the Virasoro algebra. At $n = 0$ we have as usual included the possibility of an ordering constant. A state satisfying (4.1.5) is called *physical*. In the terminology of eq. (2.9.8), a physical state is a highest weight state of weight $-A$. The condition (4.1.5) is similar to the Gupta–Bleuler quantization of electrodynamics.

Using the adjoint (4.1.7), one sees that a state of the form

$$|\chi\rangle = \sum_{n=1}^{\infty} L^{\rm m}_{-n}|\chi_n\rangle \tag{4.1.8}$$

is orthogonal to all physical states for any $|\chi_n\rangle$. Such a state is called *spurious*. A state that is both physical and spurious is called *null*. If $|\psi\rangle$ is physical and $|\chi\rangle$ is null, then $|\psi\rangle + |\chi\rangle$ is also physical and its inner product with any physical state is the same as that of $|\psi\rangle$. Therefore these two states are physically indistinguishable, and we identify

$$|\psi\rangle \cong |\psi\rangle + |\chi\rangle \,. \tag{4.1.9}$$

The real physical Hilbert space is then the set of equivalence classes,

$$\mathscr{H}_{\rm OCQ} = \frac{\mathscr{H}_{\rm phys}}{\mathscr{H}_{\rm null}} \,. \tag{4.1.10}$$

Let us see how this works for the first two levels of the open string in flat spacetime, not necessarily assuming $D = 26$. The only relevant terms are

$$L^{\rm m}_0 = \alpha' p^2 + \alpha_{-1}\cdot\alpha_1 + \ldots \,, \tag{4.1.11a}$$

$$L^{\rm m}_{\pm 1} = (2\alpha')^{1/2} p\cdot\alpha_{\pm 1} + \ldots \,. \tag{4.1.11b}$$

At the lowest mass level, the only state is $|0;k\rangle$. The only nontrivial condition at this level is $(L^{\rm m}_0 + A)|\psi\rangle = 0$, giving $m^2 = A/\alpha'$. There are no null states at this level — the Virasoro generators in the spurious state (4.1.8) all contain raising operators. Thus there is one equivalence class, corresponding to a scalar particle.

At the next level there are D states

$$|e;k\rangle = e\cdot\alpha_{-1}|0;k\rangle \,. \tag{4.1.12}$$

The norm is

$$\begin{aligned}\langle e;k|e;k'\rangle &= \langle 0;k|e^*\cdot\alpha_1\, e\cdot\alpha_{-1}|0;k'\rangle \\ &= \langle 0;k|(e^*\cdot e + e^*\cdot\alpha_{-1} e\cdot\alpha_1)|0;k'\rangle \\ &= e^{\mu *} e_\mu (2\pi)^D \delta^D(k-k') \,.\end{aligned} \tag{4.1.13}$$

We have used

$$\alpha^{\mu\dagger}_n = \alpha^\mu_{-n} \,, \tag{4.1.14}$$

as follows from the Hermiticity of X^μ, and also

$$\langle 0;k|0;k'\rangle = (2\pi)^D \delta^D(k-k') \,, \tag{4.1.15}$$

as follows from momentum conservation. The timelike excitation has a negative norm.

The $L^{\rm m}_0$ condition gives

$$m^2 = \frac{1+A}{\alpha'} \,. \tag{4.1.16}$$

The other nontrivial physical state condition is

$$L^{\rm m}_1|e;k\rangle \propto p\cdot\alpha_1 e\cdot\alpha_{-1}|0;k\rangle = e\cdot k|0;k\rangle = 0 \tag{4.1.17}$$

and so $k \cdot e = 0$. There is a spurious state at this level:

$$L_{-1}^{\mathrm{m}}|0;k\rangle = 2k \cdot \alpha_{-1}|0;k\rangle \ . \tag{4.1.18}$$

That is, $e^\mu \propto k^\mu$ is spurious. There are now three cases:

(i) If $A > -1$, the mass-squared is positive. Going to the rest frame, the physical state condition removes the negative norm timelike polarization. The spurious state (4.1.18) is not physical, $k \cdot k \neq 0$, so there are no null states and the spectrum consists of the $D-1$ positive-norm states of a massive vector particle.

(ii) If $A = -1$, the mass-squared is zero. Now $k \cdot k = 0$ so the spurious state is physical and null. Thus,

$$k \cdot e = 0 \ , \quad e_\mu \cong e_\mu + \gamma k_\mu \ . \tag{4.1.19}$$

This describes the $D-2$ positive-norm states of a massless vector particle.

(iii) If $A < -1$, the mass-squared is negative. The momentum is spacelike, so the physical state condition removes a positive-norm *spacelike* polarization. The spurious state is not physical, so we are left with a tachyonic vector particle with $D-2$ positive-norm states and one negative-norm state.

Case *(iii)* is unacceptable. Case *(ii)* is the same as the light-cone quantization. Case *(i)* is not the same as the light-cone spectrum, having different masses and an extra state at the first level, but does not have any obvious inconsistency. The difficulty is that there is no known way to give such a theory consistent interactions. Case *(ii)* is the one of interest in string theory, and the one we will recover from the more general approach of the next section.

The result at the next level is quite interesting: it depends on the constant A and also on the spacetime dimension D. Restricting to the value $A = -1$ found at the first excited level, the spectrum only agrees with the light-cone spectrum if in addition $D = 26$. If $D > 26$, there are negative-norm states; if $D < 26$ the OCQ spectrum has positive norm but more states than in light-cone quantization. The derivation is left to the reader. The OCQ at $A = -1$ and $D = 26$ is in fact the same as light-cone quantization at all mass levels,

$$\mathscr{H}_{\mathrm{OCQ}} = \mathscr{H}_{\text{light-cone}} \ , \tag{4.1.20}$$

as will be shown in section 4.4. It is only in this case that consistent interactions are known.

The extension to the closed string is straightforward. There are two sets of oscillators and two sets of Virasoro generators, so at each level the spectrum is the product of two copies of the open string spectrum (except for the normalization of p^2 in L_0^{m}). The first two levels are then

$$|0;k\rangle \ , \quad m^2 = -\frac{4}{\alpha'} \ ; \tag{4.1.21a}$$

$$e_{\mu\nu}\alpha_{-1}^\mu \tilde{\alpha}_{-1}^\nu |0;k\rangle \ , \quad m^2 = 0 \ , \ k^\mu e_{\mu\nu} = k^\nu e_{\mu\nu} = 0 \ , \tag{4.1.21b}$$

$$e_{\mu\nu} \cong e_{\mu\nu} + a_\mu k_\nu + k_\mu b_\nu \ , \ a \cdot k = b \cdot k = 0 \ . \tag{4.1.21c}$$

The relevant values are again $A = -1$, $D = 26$. As in light-cone quantization, there are $(D-2)^2$ massless states forming a traceless symmetric tensor, an antisymmetric tensor, and a scalar.

Mnemonic

For more general string theories, it is useful to have a quick mnemonic for obtaining the zero-point constant. As will be derived in the following sections, the L_0^{m} condition can be understood as

$$(L_0^{\mathrm{m}} + L_0^{\mathrm{g}})|\psi, \downarrow\rangle = 0 \ . \tag{4.1.22}$$

That is, one includes the ghost contribution with the ghosts in their ground state $|\downarrow\rangle$, which has $L_0^{\mathrm{g}} = -1$. The L_0 generators differ from the Hamiltonian by a shift (2.6.10) proportional to the central charge, but the total central charge is zero in string theory so we can just as well write this condition as

$$(H^{\mathrm{m}} + H^{\mathrm{g}})|\psi, \downarrow\rangle = 0 \ . \tag{4.1.23}$$

Now apply the mnemonic for zero-point energies given at the end of chapter 2. In particular, the ghosts always cancel the $\mu = 0, 1$ oscillators because they have the same periodicities but opposite statistics. So the rule is that A is given as the sum of the zero-point energies of the transverse oscillators. This is the same rule as in the light-cone, giving here $A = 24(-\frac{1}{24}) = -1$.

Incidentally, for the purposes of counting the physical states (but not for their precise form) one can simply ignore the ghost and $\mu = 0, 1$ oscillators and count transverse excitations as in the light-cone.

The condition (4.1.22) requires the weight of the matter state to be 1. Since the physical state condition requires the matter state to be a highest weight state, the vertex operator must then be a weight 1 or (1,1) tensor. This agrees with the condition from section 3.6 that the integrated vertex operator be conformally invariant, which gives another understanding of the condition $A = -1$.

4.2 BRST quantization

We now return to a more systematic study of the spectrum. The condition (4.1.2) is not sufficient to guarantee gauge invariance. It implies invariance for arbitrary fixed choices of g_{ab}, but this is not the most general gauge. In light-cone gauge, for example, we placed some conditions on X^μ and some conditions on g_{ab}. To consider the most general possible variation of the gauge condition, we must allow δg_{ab} to be an *operator*, that is, to depend on the fields in the path integral.

In order to derive the full invariance condition, it is useful to take a more general and abstract point of view. Consider a path integral with a local symmetry. The path integral fields are denoted ϕ_i, which in the present case would be $X^\mu(\sigma)$ and $g_{ab}(\sigma)$. Here we use a very condensed notation, where i labels the field and *also* represents the coordinate σ. The gauge invariance is $\epsilon^\alpha \delta_\alpha$, where again α includes the coordinate. By assumption the gauge parameters ϵ^α are real, since we can always separate a complex parameter into its real and imaginary parts. The gauge transformations satisfy an algebra

$$[\delta_\alpha, \delta_\beta] = f^\gamma{}_{\alpha\beta}\delta_\gamma \ . \tag{4.2.1}$$

Now fix the gauge by conditions

$$F^A(\phi) = 0 \ , \tag{4.2.2}$$

where once again A includes the coordinate. Following the same Faddeev–Popov procedure as in section 3.3, the path integral becomes

$$\int \frac{[d\phi_i]}{V_{\text{gauge}}} \exp(-S_1) \to \int [d\phi_i \, dB_A \, db_A \, dc^\alpha] \exp(-S_1 - S_2 - S_3) \ , \tag{4.2.3}$$

where S_1 is the original gauge-invariant action, S_2 is the gauge-fixing action

$$S_2 = -iB_A F^A(\phi) \ , \tag{4.2.4}$$

and S_3 is the Faddeev–Popov action

$$S_3 = b_A c^\alpha \delta_\alpha F^A(\phi) \ . \tag{4.2.5}$$

We have introduced the field B_A to produce an integral representation of the gauge-fixing $\delta(F^A)$.

There are two things to notice about this action. The first is that it is invariant under the *Becchi–Rouet–Stora–Tyutin (BRST) transformation*

$$\delta_B \phi_i = -i\epsilon c^\alpha \delta_\alpha \phi_i \ , \tag{4.2.6a}$$

$$\delta_B B_A = 0 \ , \tag{4.2.6b}$$

$$\delta_B b_A = \epsilon B_A \ , \tag{4.2.6c}$$

$$\delta_B c^\alpha = \frac{i}{2}\epsilon f^\alpha{}_{\beta\gamma} c^\beta c^\gamma \ . \tag{4.2.6d}$$

This transformation mixes commuting and anticommuting objects, so that ϵ must be taken to be anticommuting. There is a conserved ghost number, which is $+1$ for c^α, -1 for b_A and ϵ, and 0 for all other fields. The original action S_1 is invariant by itself, because the action of δ_B on ϕ_i is just a gauge transformation with parameter $i\epsilon c^\alpha$. The variation of S_2 cancels the variation of b_A in S_3, while the variations of $\delta_\alpha F^A$ and c^α in S_3 cancel.

The second key property is that

$$\delta_B(b_A F^A) = i\epsilon(S_2 + S_3) \ . \tag{4.2.7}$$

Now consider a small change δF in the gauge-fixing condition. The change in the gauge-fixing and ghost actions gives

$$\begin{aligned} \epsilon\delta\langle f|i\rangle &= i\langle f|\delta_B(b_A\delta F^A)|i\rangle \\ &= -\epsilon\langle f|\{Q_B, b_A\delta F^A\}|i\rangle \ , \end{aligned} \tag{4.2.8}$$

where we have written the BRST variation as an anticommutator with the corresponding conserved charge Q_B.

Therefore, physical states must satisfy

$$\langle\psi|\{Q_B, b_A\delta F^A\}|\psi'\rangle = 0 \ . \tag{4.2.9}$$

In order for this to hold for arbitrary δF, it must be that

$$Q_B|\psi\rangle = Q_B|\psi'\rangle = 0 \ . \tag{4.2.10}$$

This is the essential condition: *physical states must be BRST-invariant*. We have assumed that $Q_B^\dagger = Q_B$. There are several ways to see that this must indeed be the case. One is that if $Q_B^\dagger$ were different, it would have to be some other symmetry, and there is no candidate. A better argument is that the fields c^α and b_A are like anticommuting versions of the gauge parameter ϵ^α and Lagrange multiplier B_A, and so inherit their reality properties.

As an aside, we may also add to the action a term proportional to

$$\epsilon^{-1}\delta_B(b_A B_B M^{AB}) = -B_A B_B M^{AB} \tag{4.2.11}$$

for any constant matrix M^{AB}. By the above argument, amplitudes between physical states are unaffected. The integral over B_A now produces a Gaussian rather than a delta function: these are the *Gaussian-averaged gauges*, which include for example the covariant α-gauges of gauge theory.

There is one more key idea. In order to move around in the space of gauge choices, the BRST charge must remain conserved. Thus it must commute with the change in the Hamiltonian,

$$\begin{aligned} 0 &= [Q_B, \{Q_B, b_A\delta F^A\}] \\ &= Q_B^2 b_A\delta F^A - Q_B b_A\delta F^A Q_B + Q_B b_A\delta F^A Q_B - b_A\delta F^A Q_B^2 \\ &= [Q_B^2, b_A\delta F^A] \ . \end{aligned} \tag{4.2.12}$$

In order for this to vanish for general changes of gauge, we need

$$Q_B^2 = 0 . \tag{4.2.13}$$

That is, the BRST charge is *nilpotent*; the possibility $Q_B^2 =$ constant is excluded because Q_B^2 has ghost number 2. The reader can check that acting twice with the BRST transformation (4.2.6) leaves all fields invariant. In particular,

$$\delta_B(\delta'_B c^\alpha) = -\frac{1}{2}\epsilon\epsilon' f^\alpha{}_{\beta\gamma} f^\gamma{}_{\delta\epsilon} c^\beta c^\delta c^\epsilon = 0 . \tag{4.2.14}$$

The product of ghosts is antisymmetric on the indices β, δ, ϵ, and the product of structure functions then vanishes by the Jacobi identity.

We should mention that we have made two simplifying assumptions about the gauge algebra (4.2.1). The first is that the structure constants $f^\alpha{}_{\beta\gamma}$ *are* constants, independent of the fields, and the second is that the algebra does not have, on the right-hand side, additional terms proportional to the equations of motion. More generally, both of these assumptions break down. In these cases, the BRST formalism as we have described it does not give a nilpotent transformation, and a generalization, the *Batalin–Vilkovisky (BV) formalism*, is needed. The BV formalism has had various applications in string theory, but we will not have a need for it.

The nilpotence of Q_B has an important consequence. A state of the form

$$Q_B|\chi\rangle \tag{4.2.15}$$

will be annihilated by Q_B for any χ and so is physical. However, it is orthogonal to all physical states including itself:

$$\langle\psi|(Q_B|\chi\rangle) = (\langle\psi|Q_B)|\chi\rangle = 0 \tag{4.2.16}$$

if $Q_B|\psi\rangle = 0$. All physical amplitudes involving such a *null state* thus vanish. Two physical states that differ by a null state,

$$|\psi'\rangle = |\psi\rangle + Q_B|\chi\rangle \tag{4.2.17}$$

will have the same inner products with all physical states and are therefore physically equivalent. So, just as for the OCQ, we identify the true physical space with a set of *equivalence classes*, states differing by a null state being equivalent. This is a natural construction for a nilpotent operator, and is known as the *cohomology* of Q_B. Other examples of nilpotent operators are the exterior derivative in differential geometry and the boundary operator in topology. In cohomology, the term *closed* is often used for states annihilated by Q_B and the term *exact* for states of the form (4.2.15). Thus, our prescription is

$$\mathscr{H}_{\text{BRST}} = \frac{\mathscr{H}_{\text{closed}}}{\mathscr{H}_{\text{exact}}} . \tag{4.2.18}$$

We will see explicitly in the remainder of the chapter that this space has the expected form in string theory. Essentially, the invariance condition removes one set of unphysical X^μ oscillators and one set of ghost oscillators, and the equivalence relation removes the other set of unphysical X^μ oscillators and the other set of ghost oscillators.

Point-particle example

Let us consider an example, the point particle. Expanding out the condensed notation above, the local symmetry is coordinate reparameterization $\delta\tau(\tau)$, so the index α just becomes τ and a basis of infinitesimal transformations is $\delta_{\tau_1}\tau(\tau) = \delta(\tau - \tau_1)$. These act on the fields as

$$\delta_{\tau_1} X^\mu(\tau) = -\delta(\tau - \tau_1)\partial_\tau X^\mu(\tau) , \quad \delta_{\tau_1} e(\tau) = -\partial_\tau[\delta(\tau - \tau_1)X^\mu(\tau)] . \tag{4.2.19}$$

Acting with a second transformation and forming the commutator, we have

$$\begin{aligned}[\delta_{\tau_1}, \delta_{\tau_2}]X^\mu(\tau) &= -\Big[\delta(\tau - \tau_1)\partial_\tau\delta(\tau - \tau_2) - \delta(\tau - \tau_2)\partial_\tau\delta(\tau - \tau_1)\Big]\partial_\tau X^\mu(\tau) \\ &\equiv \int d\tau_3 f^{\tau_3}{}_{\tau_1\tau_2}\delta_{\tau_3} X^\mu(\tau) .\end{aligned} \tag{4.2.20}$$

From the commutator we have determined the structure function

$$f^{\tau_3}{}_{\tau_1\tau_2} = \delta(\tau_3 - \tau_1)\partial_{\tau_3}\delta(\tau_3 - \tau_2) - \delta(\tau_3 - \tau_2)\partial_{\tau_3}\delta(\tau_3 - \tau_1) . \tag{4.2.21}$$

The BRST transformation is then

$$\delta_B X^\mu = i\epsilon c\dot{X}^\mu , \tag{4.2.22a}$$

$$\delta_B e = i\epsilon(\dot{c}e) , \tag{4.2.22b}$$

$$\delta_B B = 0 , \tag{4.2.22c}$$

$$\delta_B b = \epsilon B , \tag{4.2.22d}$$

$$\delta_B c = i\epsilon c\dot{c} . \tag{4.2.22e}$$

The gauge $e(\tau) = 1$ is analogous to the unit gauge for the string, using the single coordinate freedom to fix the single component of the tetrad, so $F(\tau) = 1 - e(\tau)$. The gauge-fixed action is then

$$S = \int d\tau \left(\frac{1}{2}e^{-1}\dot{X}^\mu\dot{X}_\mu + \frac{1}{2}em^2 + iB(e-1) + eb\dot{c}\right) . \tag{4.2.23}$$

We will find it convenient to integrate B out, thus fixing $e = 1$. This leaves only the fields X^μ, b, and c, with action

$$S = \int d\tau \left(\frac{1}{2}\dot{X}^\mu\dot{X}_\mu + \frac{1}{2}m^2 + \dot{b}c\right) \tag{4.2.24}$$

and BRST transformation

$$\delta_B X^\mu = i\epsilon c\dot{X}^\mu , \tag{4.2.25a}$$

$$\delta_B b = i\epsilon\left(-\frac{1}{2}\dot{X}^\mu\dot{X}_\mu + \frac{1}{2}m^2 + b\dot{c}\right) , \tag{4.2.25b}$$

$$\delta_B c = i\epsilon c\dot{c} . \tag{4.2.25c}$$

Since B is no longer present, we have used the equation of motion from e to replace B in the transformation law for b. The reader can check that the new transformation (4.2.25) is a symmetry of the action and is nilpotent. This always works when the fields B^A are integrated out, though $\delta_B\delta'_B b$ is no longer identically zero but is proportional to the equations of motion. This is perfectly satisfactory: $Q_B^2 = 0$ holds as an operator equation, which is what we need.

The canonical commutators are

$$[p^\mu, X^\nu] = -i\eta^{\mu\nu} , \quad \{b, c\} = 1 , \tag{4.2.26}$$

with $p^\mu = i\dot{X}^\mu$, the i being from the Euclidean signature. The Hamiltonian is $H = \frac{1}{2}(p^2 + m^2)$, and the Noether procedure gives the BRST operator

$$Q_B = cH . \tag{4.2.27}$$

The structure here is analogous to what we will find for the string. The constraint (the missing equation of motion) is $H = 0$, and the BRST operator is c times this.

The ghosts generate a two-state system, so a complete set of states is $|k,\downarrow\rangle$, $|k,\uparrow\rangle$, where

$$p^\mu|k,\downarrow\rangle = k^\mu|k,\downarrow\rangle , \quad p^\mu|k,\uparrow\rangle = k^\mu|k,\uparrow\rangle , \tag{4.2.28a}$$

$$b|k,\downarrow\rangle = 0 , \quad b|k,\uparrow\rangle = |k,\downarrow\rangle , \tag{4.2.28b}$$

$$c|k,\downarrow\rangle = |k,\uparrow\rangle , \quad c|k,\uparrow\rangle = 0 . \tag{4.2.28c}$$

The action of the BRST operator on these is

$$Q_B|k,\downarrow\rangle = (k^2 + m^2)|k,\uparrow\rangle , \quad Q_B|k,\uparrow\rangle = 0 . \tag{4.2.29}$$

From this it follows that the closed states are

$$|k,\downarrow\rangle , \quad k^2 + m^2 = 0 , \tag{4.2.30a}$$

$$|k,\uparrow\rangle , \quad \text{all } k^\mu , \tag{4.2.30b}$$

and that the exact states are

$$|k,\uparrow\rangle , \quad k^2 + m^2 \neq 0 . \tag{4.2.31}$$

The closed states that are not exact are

$$|k,\downarrow\rangle , \; k^2 + m^2 = 0 ; \quad |k,\uparrow\rangle , \; k^2 + m^2 = 0 . \tag{4.2.32}$$

That is, physical states must satisfy the mass-shell condition, but we have *two copies* of the expected spectrum. In fact, only states $|k,\downarrow\rangle$ satisfying the additional condition

$$b|\psi\rangle = 0 \tag{4.2.33}$$

appear in physical amplitudes. The origin of this additional condition is kinematic. For $k^2 + m^2 \neq 0$, the states $|k,\uparrow\rangle$ are exact — they are are orthogonal to all physical states and their amplitudes must vanish identically. So the amplitudes can only be proportional to $\delta(k^2 + m^2)$. But amplitudes in field theory and in string theory, while they may have poles and cuts, never have delta functions (except in $D = 2$ where the kinematics is special). So they must vanish identically.

4.3 BRST quantization of the string

In string theory, the BRST transformation is

$$\delta_B X^\mu = i\epsilon(c\partial + \tilde{c}\bar{\partial})X^\mu , \tag{4.3.1a}$$

$$\delta_B b = i\epsilon(T^X + T^g) , \quad \delta_B\tilde{b} = i\epsilon(\tilde{T}^X + \tilde{T}^g) , \tag{4.3.1b}$$

$$\delta_B c = i\epsilon(c\partial + \tilde{c}\bar{\partial})c , \quad \delta_B\tilde{c} = i\epsilon(c\partial + \tilde{c}\bar{\partial})\tilde{c} . \tag{4.3.1c}$$

The reader can derive this following the point-particle example. To the sum of the Polyakov and ghost actions, add the gauge-fixing term

$$\frac{i}{4\pi}\int d^2\sigma\, g^{1/2}B^{ab}(\delta_{ab} - g_{ab}) . \tag{4.3.2}$$

The transformation $\delta_B b_{ab} = \epsilon B_{ab}$ becomes (4.3.1) after integrating over B_{ab} and using the g_{ab} equation of motion to replace B_{ab} in the transformation. The Weyl ghost is just a Lagrange multiplier, making b_{ab} traceless. Again, the reader can check nilpotence up to the equations of motion. The similarity between the string BRST transformation (4.3.1) and the particle case (4.2.25) is evident. Replacing the X^μ with a general matter CFT, the BRST transformation of the matter fields is a conformal transformation with $v(z) = c(z)$, while T^m replaces T^X in the transformation of b.

Noether's theorem gives the BRST current

$$j_B = cT^m + \frac{1}{2}:cT^g: + \frac{3}{2}\partial^2 c ,$$

$$= cT^m + :bc\partial c: + \frac{3}{2}\partial^2 c , \tag{4.3.3}$$

and correspondingly for $\tilde{j}_B$. The final term in the current is a total derivative and does not contribute to the BRST charge; it has been added by hand to make the BRST current a tensor. The OPEs of the BRST

current with the ghost fields and with a general matter tensor field are

$$j_B(z)b(0) \sim \frac{3}{z^3} + \frac{1}{z^2}j^g(0) + \frac{1}{z}T^{m+g}(0) , \tag{4.3.4a}$$

$$j_B(z)c(0) \sim \frac{1}{z}c\partial c(0) , \tag{4.3.4b}$$

$$j_B(z)\mathcal{O}^m(0) \sim \frac{h}{z^2}c\mathcal{O}^m(0) + \frac{1}{z}\left[h(\partial c)\mathcal{O}^m(0) + c\partial\mathcal{O}^m(0)\right] . \tag{4.3.4c}$$

The simple poles reflect the BRST transformations (4.3.1) of these fields.

The BRST operator is

$$Q_B = \frac{1}{2\pi i}\oint (dz\, j_B - d\bar{z}\,\tilde{j}_B) . \tag{4.3.5}$$

By the usual contour argument, the OPE implies

$$\{Q_B, b_m\} = L_m^m + L_m^g . \tag{4.3.6}$$

In terms of the ghost modes,

$$Q_B = \sum_{n=-\infty}^{\infty}(c_n L_{-n}^m + \tilde{c}_n \tilde{L}_{-n}^m) + \sum_{m,n=-\infty}^{\infty}\frac{(m-n)}{2}{:}(c_m c_n b_{-m-n} + \tilde{c}_m\tilde{c}_n\tilde{b}_{-m-n}){:} + a^B(c_0 + \tilde{c}_0) . \tag{4.3.7}$$

The ordering constant is $a^B = a^g = -1$, as follows from the anticommutator

$$\{Q_B, b_0\} = L_0^m + L_0^g . \tag{4.3.8}$$

There is an anomaly in the gauge symmetry when $c^m \neq 26$, so we expect some breakdown in the BRST formalism. The BRST current is still conserved: all of the terms in the current (4.3.3) are analytic for any value of the central charge. However, it is no longer *nilpotent*,

$$\{Q_B, Q_B\} = 0 \quad \text{only if } c^m = 26 . \tag{4.3.9}$$

The shortest derivation of this uses the Jacobi identity, exercise 4.3. More directly it follows from the OPE

$$j_B(z)j_B(0) \sim -\frac{c^m - 18}{2z^3}c\partial c(0) - \frac{c^m - 18}{4z^2}c\partial^2 c(0) - \frac{c^m - 26}{12z}c\partial^3 c(0) . \tag{4.3.10}$$

This requires a bit of calculation, being careful about minus signs from anticommutation. The single pole implies that $\{Q_B, Q_B\} = 0$ when $c^m = 26$.

Note also the OPE

$$T(z)j_B(0) \sim \frac{c^m - 26}{2z^4}c(0) + \frac{1}{z^2}j_B(0) + \frac{1}{z}\partial j_B(0) , \tag{4.3.11}$$

which implies that j_B is a tensor only when $c^m = 26$.

Let us point out some important features. The missing equation of motion is just the vanishing of the generator T_{ab} of the conformal transformations, the part of the local symmetry group not fixed by the gauge choice. This is a general result. When the gauge conditions fix the symmetry completely, one can show that the missing equations of motion are trivial due to the gauge invariance of the action. When they do not, so that there is a residual symmetry group, the missing equations of motion require its generators to vanish. They must vanish in the sense (4.1.2) of matrix elements, or more generally in the BRST sense. Denote the generators of the residual symmetry group by G_I. The G_I are called *constraints;* for the string these are L_m and $\tilde{L}_m$. They form an algebra

$$[G_I, G_J] = ig^K{}_{IJ}G_K . \tag{4.3.12}$$

Associated with each generator is a pair of ghosts, b_I and c^J, where

$$\{c^I, b_J\} = \delta^I{}_J , \quad \{c^I, c^J\} = \{b_I, b_J\} = 0 . \tag{4.3.13}$$

The general form of the BRST operator, as illustrated by the string case (4.3.7), is

$$Q_B = c^I G_I^m - \frac{i}{2}g^K{}_{IJ}c^I c^J b_K = c^I\left(G_I^m + \frac{1}{2}G_I^g\right) , \tag{4.3.14}$$

where G_I^m is the matter part of G_I and

$$G_I^g = -ig^K{}_{IJ}c^J b_K \tag{4.3.15}$$

is the ghost part. The G_I^m and G_I^g satisfy the same algebra (4.3.12) as the G_I. Using the commutators (4.3.12) and (4.3.13), one finds

$$Q_B^2 = \frac{1}{2}\{Q_B, Q_B\} = -\frac{1}{2}g^K{}_{IJ}g^M{}_{KL}c^I c^J c^L b_M = 0 . \tag{4.3.16}$$

The last equality follows from the *GGG* Jacobi identity, which requires $g^K{}_{IJ}g^M{}_{KL}$ to vanish when antisymmetrized on IJL. We have neglected central charge terms; these need to be checked by hand.

To reiterate, a *constraint* algebra is a world-sheet symmetry algebra that is required to vanish in physical matrix elements. When we go on to generalize the bosonic string theory in volume two, it will be easiest to do so directly in terms of the constraint algebra, and we will write the BRST charge directly in the form (4.3.14).

BRST cohomology of the string

Let us now look at the BRST cohomology at the lowest levels of the string. The inner product is defined by specifying

$$(\alpha_m^\mu)^\dagger = \alpha_{-m}^\mu\,, \quad (\tilde\alpha_m^\mu)^\dagger = \tilde\alpha_{-m}^\mu\,, \tag{4.3.17a}$$

$$(b_m)^\dagger = b_{-m}\,, \quad (\tilde b_m)^\dagger = \tilde b_{-m}\,, \tag{4.3.17b}$$

$$(c_m)^\dagger = c_{-m}\,, \quad (\tilde c_m)^\dagger = \tilde c_{-m}\,. \tag{4.3.17c}$$

In particular, the Hermiticity of the BRST charge requires that the ghost fields be Hermitean as well. The Hermiticity of the ghost zero modes forces the inner products of the ground state to take the form

$$\text{open string:} \quad \langle 0;k|c_0|0;k'\rangle = (2\pi)^{26}\delta^{26}(k-k')\,, \tag{4.3.18a}$$

$$\text{closed string:} \quad \langle 0;k|\tilde c_0 c_0|0;k'\rangle = i(2\pi)^{26}\delta^{26}(k-k')\,. \tag{4.3.18b}$$

Here $|0;k\rangle$ denotes the matter ground state times the ghost ground state $|{\downarrow}\downarrow\rangle$, with momentum k. The c_0 and $\tilde c_0$ insertions are necessary for a nonzero result. For example, $\langle 0;k|0;k'\rangle = \langle 0;k|(c_0b_0 + b_0c_0)|0;k'\rangle = 0$; the last equality follows because b_0 annihilates both the bra and the ket. The factor of i is needed in the ghost zero-mode inner product for Hermiticity. Inner products of general states are then obtained, as in the earlier calculation (4.1.13), by use of the commutation relations and the adjoints (4.3.17).

We will focus on the open string; the closed string discussion is entirely parallel but requires twice as much writing. As in the point-particle case (4.2.33), we will assert (and later show) that physical states must satisfy the additional condition

$$b_0|\psi\rangle = 0\,. \tag{4.3.19}$$

This also implies

$$L_0|\psi\rangle = \{Q_B, b_0\}|\psi\rangle = 0\,, \tag{4.3.20}$$

since Q_B and b_0 both annihilate $|\psi\rangle$. The operator L_0 is

$$L_0 = \alpha'(p^\mu p_\mu + m^2)\,, \tag{4.3.21}$$

where

$$\alpha' m^2 = \sum_{n=1}^{\infty} n\left(N_{bn} + N_{cn} + \sum_{\mu=0}^{25} N_{\mu n}\right) - 1\,. \tag{4.3.22}$$

Thus the L_0 condition (4.3.20) determines the mass of the string. BRST invariance, with the extra condition (4.3.19), implies that every string state is on the mass shell. We will denote by $\hat{\mathscr{H}}$ the space of states satisfying the conditions (4.3.19) and (4.3.20). From the commutators $\{Q_B, b_0\} = L_0$ and $[Q_B, L_0] = 0$, it follows that Q_B takes $\hat{\mathscr{H}}$ into itself.

The inner product (4.3.18a) is not quite the correct one to use in $\hat{\mathscr{H}}$. It is not well defined: the ghost zero modes give zero, while the $\delta^{26}(k-k')$ contains a factor $\delta(0)$ because the momentum support is restricted to the mass shell. Therefore we use in $\hat{\mathscr{H}}$ a reduced inner product $\langle\ \|\ \rangle$ in which we simply ignore the X^0 and ghost zero modes. It is this inner product that will be relevant for the probabilistic interpretation. One can check that Q_B is still Hermitean with the reduced inner product, in the space $\hat{\mathscr{H}}$. Note that the mass-shell condition determines k^0 in terms of the spatial momentum $\mathbf{k}$ (we focus on the incoming case, $k^0 > 0$), and that we have used covariant normalization for states.

Let us now work out the first levels of the $D = 26$ flat spacetime string. At the lowest level, $N = 0$, we have

$$|0;\mathbf{k}\rangle\,, \quad -k^2 = -\frac{1}{\alpha'}\,. \tag{4.3.23}$$

This state is invariant,

$$Q_B|0;\mathbf{k}\rangle = 0\,, \tag{4.3.24}$$

because every term in Q_B contains either lowering operators or L_0. This also shows that there are no exact states at this level, so each invariant state corresponds to a cohomology class. These are just the states of the tachyon. The mass-shell condition is the same as found in the light-cone quantization of section 1.3 and from the open string vertex operators in section 3.6. The tachyon mass-squared, determined in a heuristic way in chapter 1, is here fixed by the normal ordering constant in Q_B, which is determined by the requirement of nilpotence.

At the next level, $N = 1$, there are $26 + 2$ states,

$$|\psi_1\rangle = (e\cdot\alpha_{-1} + \beta b_{-1} + \gamma c_{-1})|0;\mathbf{k}\rangle\,, \quad -k^2 = 0\,, \tag{4.3.25}$$

depending on a 26-vector e_μ and two constants, β and γ. The norm of this state is

$$\begin{aligned}\langle\psi_1\|\psi_1\rangle &= \langle 0;\mathbf{k}\|(e^*\cdot\alpha_1 + \beta^* b_1 + \gamma^* c_1)(e\cdot\alpha_{-1} + \beta b_{-1} + \gamma c_{-1})|0;\mathbf{k}'\rangle \\ &= (e^*\cdot e + \beta^*\gamma + \gamma^*\beta)\langle 0;\mathbf{k}\|0;\mathbf{k}'\rangle\,.\end{aligned} \tag{4.3.26}$$

Going to an orthogonal basis, there are 26 positive-norm states and 2 negative-norm states. The BRST condition is

$$\begin{aligned}0 = Q_B|\psi_1\rangle &= (2\alpha')^{1/2}(c_{-1}k\cdot\alpha_1 + c_1 k\cdot\alpha_{-1})|\psi_1\rangle \\ &= (2\alpha')^{1/2}(k\cdot e c_{-1} + \beta k\cdot\alpha_{-1})|0;\mathbf{k}\rangle\,.\end{aligned} \tag{4.3.27}$$

The terms proportional to c_0 sum to zero by the mass-shell condition, and are omitted. An invariant state therefore satisfies $k\cdot e = \beta = 0$. There are

26 linearly independent states remaining, of which 24 have positive norm and 2 have zero norm, being orthogonal to all physical states including themselves. The zero-norm invariant states are created by c_{-1} and $k\cdot\alpha_{-1}$. A general $|\chi\rangle$ is of the same form (4.3.25) with constants e'_μ, β', γ', so the general BRST-exact state at this level is

$$Q_B|\chi\rangle = (2\alpha')^{1/2}(k\cdot e' c_{-1} + \beta' k\cdot\alpha_{-1})|0;\mathbf{k}\rangle\ . \tag{4.3.28}$$

Thus the ghost state $c_{-1}|0;\mathbf{k}\rangle$ is BRST-exact, while the polarization is transverse with the equivalence relation $e_\mu \cong e_\mu + 2\beta' k_\mu$. This leaves the 24 positive-norm states expected for a massless vector particle, in agreement with the light-cone quantization and with the OCQ at $A=-1$.

This pattern is general: there are two extra positive-norm and two extra negative-norm families of oscillators, as compared to the light-cone quantization. The physical state condition eliminates two of these and leaves two combinations with vanishing inner products. These null oscillators are BRST-exact and are removed by the equivalence relation.

An aside: the states $b_{-1}|0;\mathbf{k}\rangle$ and $c_{-1}|0;\mathbf{k}\rangle$ can be identified with the two Faddeev–Popov ghosts that appear in the BRST-invariant quantization of a massless vector field in field theory. The world-sheet BRST operator acts on these states in the same way as the corresponding *spacetime* BRST operator in gauge field theory. Acting on the full string Hilbert space, the open string BRST operator is then some infinite-dimensional generalization of the spacetime gauge theory BRST invariance, while the closed string BRST operator is a generalization of the spacetime general coordinate BRST invariance, in the free limit.

The generalization to the closed string is straightforward. We restrict attention to the space $\hat{\mathscr{H}}$ of states satisfying[1]

$$b_0|\psi\rangle = \tilde{b}_0|\psi\rangle = 0\ , \tag{4.3.29}$$

implying also

$$L_0|\psi\rangle = \tilde{L}_0|\psi\rangle = 0\ . \tag{4.3.30}$$

In the closed string,

$$L_0 = \frac{\alpha'}{4}(p^2+m^2)\ ,\quad \tilde{L}_0 = \frac{\alpha'}{4}(p^2+\tilde{m}^2)\ , \tag{4.3.31}$$

[1] S. Weinberg has pointed out that the simple argument given at the end of section 4.2 would only lead to the weaker condition $(b_0+\tilde{b}_0)|\psi\rangle = 0$. Nevertheless, we shall see from the detailed form of string amplitudes in chapter 9 that there is a projection onto states satisfying conditions (4.3.29).

where

$$\frac{\alpha'}{4}m^2 = \sum_{n=1}^{\infty} n\left(N_{bn}+N_{cn}+\sum_{\mu=0}^{25}N_{\mu n}\right) - 1\ , \tag{4.3.32a}$$

$$\frac{\alpha'}{4}\tilde{m}^2 = \sum_{n=1}^{\infty} n\left(\tilde{N}_{bn}+\tilde{N}_{cn}+\sum_{\mu=0}^{25}\tilde{N}_{\mu n}\right) - 1\ . \tag{4.3.32b}$$

Repeating the earlier exercise, we find at $m^2 = -4/\alpha'$ the tachyon, and at $m^2 = 0$ the 24×24 states of the graviton, dilaton, and antisymmetric tensor.

4.4 The no-ghost theorem

In this section we prove that the BRST cohomology of the string has a positive inner product and is isomorphic to the light-cone and OCQ spectra. We are identifying the BRST cohomology as the physical Hilbert space. We will also need to verify, in our study of string amplitudes, that the amplitudes are well defined on the cohomology (that is, that equivalent states have equal amplitudes) and that the S-matrix is unitary within the space of physical states.

We will work in the general framework described at the end of chapter 3. That is, the world-sheet theory consists of d free X^μ fields including $\mu=0$, plus some compact unitary CFT K of central charge $26-d$, plus the ghosts. The Virasoro generators are a sum

$$L_m = L_m^X + L_m^K + L_m^{\mathrm{g}}\ . \tag{4.4.1}$$

By compact we mean that L_0^K has a discrete spectrum. For example, in the case that K corresponds to strings on a compact manifold, the term $\alpha' p^2$ in L_0 is replaced by the Laplacian $-\alpha'\nabla^2$, which has a discrete spectrum on a compact space.

The general state is labeled

$$|N,I;k\rangle\ ,\quad |N,\tilde{N},I;k\rangle\ , \tag{4.4.2}$$

in the open and closed strings respectively, where N (and $\tilde{N}$) refer to both the d-dimensional and ghost oscillators, k is the d-momentum, and I labels the states of the compact CFT with given boundary conditions. The b_0 condition (4.3.19) or (4.3.29) is imposed as above, implying the mass-shell

condition

$$-\sum_{\mu=0}^{d-1} k_\mu k^\mu = m^2 \,, \tag{4.4.3a}$$

$$\alpha' m^2 = \sum_{n=1}^{\infty} n\left(N_{bn} + N_{cn} + \sum_{\mu=0}^{d-1} N_{\mu n}\right) + L_0^K - 1 \,, \tag{4.4.3b}$$

for the open string and

$$-\sum_{\mu=0}^{d-1} k_\mu k^\mu = m^2 = \tilde{m}^2 \,, \tag{4.4.4a}$$

$$\frac{\alpha'}{4} m^2 = \sum_{n=1}^{\infty} n\left(N_{bn} + N_{cn} + \sum_{\mu=0}^{d-1} N_{\mu n}\right) + L_0^K - 1 \,, \tag{4.4.4b}$$

$$\frac{\alpha'}{4} \tilde{m}^2 = \sum_{n=1}^{\infty} n\left(\tilde{N}_{bn} + \tilde{N}_{cn} + \sum_{\mu=0}^{d-1} \tilde{N}_{\mu n}\right) + \tilde{L}_0^K - 1 \,, \tag{4.4.4c}$$

for the closed string. That is, the contributions of the $d \le \mu \le 25$ oscillators are replaced by the eigenvalue of L_0^K or $\tilde{L}_0^K$ from the compact CFT. The only information we use about the compact CFT is that it is conformally invariant with the appropriate central charge, so there is a nilpotent BRST operator, and that it has a positive inner product. The basis I can be taken to be orthonormal, so the reduced inner products are

$$\langle 0, I; \mathbf{k} \| 0, I'; \mathbf{k}' \rangle = \langle 0, 0, I; \mathbf{k} \| 0, 0, I'; \mathbf{k}' \rangle = 2k^0 (2\pi)^{d-1} \delta^{d-1}(\mathbf{k} - \mathbf{k}') \delta_{II'} \,. \tag{4.4.5}$$

Now let us see what we expect for the physical Hilbert space. Define the transverse Hilbert space $\mathscr{H}^\perp$ to consist of those states in $\hat{\mathscr{H}}$ that have no longitudinal (X^0, X^1, b, or c) excitations. Since these oscillators are the source of the indefinite inner product, $\mathscr{H}^\perp$ has a positive inner product. Light-cone gauge-fixing eliminates the longitudinal oscillators directly — the light-cone Hilbert space is isomorphic to $\mathscr{H}^\perp$, as one sees explicitly in the flat-spacetime case from chapter 1. We will show, in the general case, that the BRST cohomology is isomorphic to $\mathscr{H}^\perp$. That is, it has the same number of states at each mass level, and has a positive inner product. This is the *no-ghost theorem.*

Proof

The proof has two parts. The first is to find the cohomology of a simplified BRST operator Q_1, which is quadratic in the oscillators, and the second is to show that the cohomology of the full Q_B is identical to that of Q_1.

Define the light-cone oscillators

$$\alpha_m^\pm = 2^{-1/2}(\alpha_m^0 \pm \alpha_m^1) \,, \tag{4.4.6}$$

which satisfy

$$[\alpha_m^+, \alpha_n^-] = -m\delta_{m,-n} \,, \quad [\alpha_m^+, \alpha_n^+] = [\alpha_m^-, \alpha_n^-] = 0 \,. \tag{4.4.7}$$

We will use extensively the quantum number

$$N^{lc} = \sum_{\substack{m=-\infty \\ m\neq 0}}^{\infty} \frac{1}{m} \alpha_{-m}^+ \alpha_m^- \,. \tag{4.4.8}$$

The number N^{lc} counts the number of $-$ excitations minus the number of $+$ excitations; it is not a Lorentz generator because the center-of-mass piece has been omitted. We choose a Lorentz frame in which the momentum component k^+ is nonzero.

Now decompose the BRST generator using the quantum number N^{lc}:

$$Q_B = Q_1 + Q_0 + Q_{-1} \,, \tag{4.4.9}$$

where Q_j changes N^{lc} by j units,

$$[N^{lc}, Q_j] = jQ_j \,. \tag{4.4.10}$$

Also, each of the Q_j increases the ghost number N^g by one unit:

$$[N^g, Q_j] = Q_j \,. \tag{4.4.11}$$

Expanding out $Q_B^2 = 0$ gives

$$\left(Q_1^2\right) + \left(\{Q_1, Q_0\}\right) + \left(\{Q_1, Q_{-1}\} + Q_0^2\right) + \left(\{Q_0, Q_{-1}\}\right) + \left(Q_{-1}^2\right) = 0 \,. \tag{4.4.12}$$

Each group in parentheses has a different N^{lc} and so must vanish separately. In particular, Q_1 itself is nilpotent, and so has a cohomology.

Explicitly,

$$Q_1 = -(2\alpha')^{1/2} k^+ \sum_{\substack{m=-\infty \\ m\neq 0}}^{\infty} \alpha_{-m}^- c_m \,. \tag{4.4.13}$$

For $m < 0$ this destroys a $+$ mode and creates a c, and for $m > 0$ it creates a $-$ mode and destroys a b. One can find the cohomology directly by considering the action of Q_1 in the occupation basis. We will leave this as an exercise, and instead use a standard strategy which will also be useful in generalizing to Q_B. Define

$$R = \frac{1}{(2\alpha')^{1/2} k^+} \sum_{\substack{m=-\infty \\ m\neq 0}}^{\infty} \alpha_{-m}^+ b_m \tag{4.4.14}$$

and

$$S \equiv \{Q_1, R\} = \sum_{m=1}^{\infty} (mb_{-m}c_m + mc_{-m}b_m - \alpha^+_{-m}\alpha^-_m - \alpha^-_{-m}\alpha^+_m)$$
$$= \sum_{m=1}^{\infty} m(N_{bm} + N_{cm} + N^+_m + N^-_m) \,. \tag{4.4.15}$$

The normal ordering constant is determined by noting that Q_1 and R both annihilate the ground state. Note that Q_1 commutes with S. We can then calculate the cohomology within each eigenspace of S, and the full cohomology is the union of the results. If $|\psi\rangle$ is Q_1-invariant with $S|\psi\rangle = s|\psi\rangle$, then for nonzero s

$$|\psi\rangle = \frac{1}{s}\{Q_1, R\}|\psi\rangle = \frac{1}{s}Q_1 R|\psi\rangle \,, \tag{4.4.16}$$

and so $|\psi\rangle$ is actually Q_1-exact. Therefore the Q_1 cohomology can be nonzero only at $s = 0$. By the definition (4.4.15) of S, the $s = 0$ states have no longitudinal excitations — the $s = 0$ space is just $\mathscr{H}^\perp$. The operator Q_1 annihilates all states in $\mathscr{H}^\perp$, so they are all Q_1-closed and there are no Q_1-exact states in this space. Therefore the cohomology is $\mathscr{H}^\perp$ itself. We have proven the no-ghost theorem, but for the operator Q_1, not Q_B.

The proof had two steps, the first being to show that the cohomology could only come from $s = 0$ states (the kernel of S), and the second to show that $s = 0$ states were Q_1-invariant. It is useful to prove the second step in a more abstract way, using the property that all $s = 0$ states have the same ghost number, in this case $-\frac{1}{2}$. Suppose $S|\psi\rangle = 0$. Since S and Q_1 commute we have

$$0 = Q_1 S|\psi\rangle = SQ_1|\psi\rangle \,. \tag{4.4.17}$$

The state $|\psi\rangle$ has ghost number $-\frac{1}{2}$, so $Q_1|\psi\rangle$ has ghost number $+\frac{1}{2}$. Since S is invertible at this ghost number, it must be that $Q_1|\psi\rangle = 0$ as we wished to show.

It remains to show that the cohomology of Q_B is the same as the cohomology of Q_1. The idea here is to use in place of S the operator

$$S + U \equiv \{Q_B, R\} \,. \tag{4.4.18}$$

Now, $U = \{Q_0 + Q_{-1}, R\}$ lowers N^{lc} by one or two units. In terms of N^{lc}, S is diagonal and U is lower triangular. By general properties of lower triangular matrices, the kernel of $S + U$ can be no larger than the kernel of its diagonal part S. In fact they are isomorphic: if $|\psi_0\rangle$ is annihilated by S, then

$$|\psi\rangle = (1 - S^{-1}U + S^{-1}US^{-1}U - \ldots)|\psi_0\rangle \tag{4.4.19}$$

is annihilated by $S + U$. The factors of S^{-1} make sense because they always act on states of $N^{lc} < 0$, where S is invertible. For the same reason, $S + U$ is invertible except at ghost number $-\frac{1}{2}$. Eqs. (4.4.16) and (4.4.17) can now be applied to Q_B, with $S + U$ replacing S. They imply that the Q_B cohomology is isomorphic to the kernel of $S + U$, which is isomorphic to the kernel of S, which is isomorphic to the cohomology of Q_1, as we wished to show.

We must still check that the inner product is positive. All terms after the first on the right-hand side of eq. (4.4.19) have strictly negative N^{lc}. By the commutation relations, the inner product is nonzero only between states whose N^{lc} adds to zero. Then for two states (4.4.19) in the kernel of $S + U$,

$$\langle\psi\|\psi'\rangle = \langle\psi_0\|\psi'_0\rangle \,. \tag{4.4.20}$$

The positivity of the inner product on the kernel of $S + U$ then follows from that for the kernel of S, and we are done.

After adding the tilded operators to Q_B, N^{lc}, R, etc., the closed string proof is identical.

BRST–OCQ equivalence

We now prove the equivalence

$$\mathscr{H}_{OCQ} = \mathscr{H}_{BRST} = \mathscr{H}_{\text{light-cone}} \,. \tag{4.4.21}$$

To a state $|\psi\rangle$ in the matter Hilbert space, associate the state

$$|\psi, \downarrow\rangle \tag{4.4.22}$$

from the full matter plus ghost theory. Again the ghost vacuum $|\downarrow\rangle$ is annihilated by all of the ghost lowering operators, b_n for $n \geq 0$ and c_n for $n > 0$. Acting with Q_B,

$$Q_B|\psi, \downarrow\rangle = \sum_{n=0}^{\infty} c_{-n}(L^m_n - \delta_{n,0})|\psi, \downarrow\rangle = 0 \,. \tag{4.4.23}$$

All terms in Q_B that contain ghost lowering operators drop out, and the constant in the $n = 0$ term is known from the mode expansion of Q_B, eq. (4.3.7). Each OCQ physical state thus maps to a BRST-closed state. The ordering constant $A = -1$ arises here from the L_0 eigenvalue of the ghost vacuum.

To establish the equivalence (4.4.21), we need to show more. First, we need to show that we have a well-defined map of equivalence classes: that if ψ and ψ' are equivalent OCQ physical states, they map into the same BRST class:

$$|\psi, \downarrow\rangle - |\psi', \downarrow\rangle \tag{4.4.24}$$

must be BRST-exact. By (4.4.23), this state is BRST-closed. Further, since $|\psi\rangle' - |\psi\rangle$ is OCQ null, the state (4.4.24) has zero norm. From the BRST no-ghost theorem, the inner product on the cohomology is positive, so a zero-norm closed state is BRST-exact, as we needed to show.

To conclude that we have an isomorphism, we need to show that the map is one-to-one and onto. One-to-one means that OCQ physical states ψ and ψ' that map into the same BRST class must be in the same OCQ class — if

$$|\psi, \downarrow\rangle - |\psi', \downarrow\rangle = Q_B|\chi\rangle \,, \tag{4.4.25}$$

then $|\psi\rangle - |\psi\rangle'$ must be OCQ null. To see this, expand

$$|\chi\rangle = \sum_{n=1}^{\infty} b_{-n}|\chi_n, \downarrow\rangle + \dots \,; \tag{4.4.26}$$

$|\chi\rangle$ has ghost number $-\frac{3}{2}$, so the ellipsis stands for terms with at least one c and two b excitations. Insert this into the form (4.4.25) and keep on both sides only terms with the ghost ground state. This gives

$$\begin{aligned} |\psi, \downarrow\rangle - |\psi', \downarrow\rangle &= \sum_{m,n=1}^{\infty} c_m L^{\mathrm{m}}_{-m} b_{-n}|\chi_n, \downarrow\rangle \\ &= \sum_{n=1}^{\infty} L^{\mathrm{m}}_{-n}|\chi_n, \downarrow\rangle \,. \end{aligned} \tag{4.4.27}$$

Terms with ghost excitations must vanish separately, and so have been omitted. Thus, $|\psi\rangle - |\psi'\rangle = \sum_{n=1}^{\infty} L^{\mathrm{m}}_{-n}|\chi_n\rangle$ is OCQ null, and the map is one-to-one.

Finally, we must show that the map is onto, that every Q_B class contains at least one state of the form (4.4.22). In fact, the specific representatives (4.4.19), the states annihilated by $S + U$, are of this form. To see this, consider the quantum number $N' = 2N^- + N_b + N_c$, involving the total numbers of $-$, b, and c excitations. The operator R has $N' = -1$: terms in R with $m > 0$ reduce N_c by one unit and terms with $m < 0$ reduce N^- by one unit and increase N_b by one unit. Examining $Q_0 + Q_{-1}$, one finds various terms with $N' = 1$, but no greater. So $U = \{R, Q_0 + Q_{-1}\}$ cannot increase N'. Examining the state (4.4.19), and noting that S and $|\psi_0\rangle$ have $N' = 0$, we see that all terms on the right-hand side have $N' \leq 0$. By definition N' is nonnegative, so it must be that $N'|\psi\rangle = 0$. This implies no $-$, b, or c excitations, and so this state is of the form (4.4.22). Thus the equivalence (4.4.21) is shown.

The full power of the BRST method is needed for understanding the general structure of string amplitudes. However, for many practical purposes it is a great simplification to work with the special states of the form $|\psi, \downarrow\rangle$, and so it is useful to know that every BRST class contains at least one state in which the ghost modes are unexcited. We will call these OCQ-type states.

The OCQ physical state condition (4.1.5) requires that the matter state be a highest weight state with $L_0 = 1$. The corresponding vertex operator is a weight-1 tensor field $\mathscr{V}^{\mathrm{m}}$ constructed from the matter fields. Including the ghost state $|\downarrow\rangle$, the full vertex operator is $c\mathscr{V}^{\mathrm{m}}$. For the closed string the full vertex operator is $\tilde{c}c\mathscr{V}^{\mathrm{m}}$ with $\mathscr{V}^{\mathrm{m}}$ a $(1,1)$ tensor. The matter part of the vertex operator is the same as found in the Polyakov formalism in section 3.6, while the ghost part has a simple interpretation that we will encounter in the next chapter.

Eq. (4.4.19) defines an OCQ-physical state in each cohomology class. For the special case of flat spacetime this state can be constructed more explicitly by using the *Del Giudice–Di Vecchia–Fubini (DDF) operators*. To explain these we will need to develop some more vertex operator technology, so this is deferred to chapter 8.

Exercises

4.1 Extend the OCQ, for $A = -1$ and general D, to the second excited level of the open string. Verify the assertions made in the text about extra positive- and negative-norm states.

4.2 (a) In the OCQ, show that the state $L^{\mathrm{m}}_{-1}|\chi\rangle$ is physical (and therefore null) if $|\chi\rangle$ is a highest weight state with $L^{\mathrm{m}}_0 = 0$.
(b) Show that the state $(L^{\mathrm{m}}_{-2} + \frac{3}{2}L^{\mathrm{m}}_{-1}L^{\mathrm{m}}_{-1})|\chi\rangle$ is physical and null if $|\chi\rangle$ is a highest weight state with $L^{\mathrm{m}}_0 = -1$.

4.3 By writing out terms, demonstrate the Jacobi identity

$$\{[Q_B, L_m], b_n\} - \{[L_m, b_n], Q_B\} - [\{b_n, Q_B\}, L_m] = 0 \,.$$

This is the sum of cyclic permutations, with commutators or anticommutators as appropriate, and minus signs from anticommutation. Use the Jacobi identity and the (anti)commutators (2.6.24) and (4.3.6) of b_n with L_n and Q_B to show that $\{[Q_B, L_m], b_n\}$ vanishes when the total central charge is zero. This implies that $[Q_B, L_m]$ contains no c modes. However, this commutator has $N^g = 1$, and so it must vanish: the charge Q_B is conformally invariant. Now use the $Q_B\, Q_B\, b_n$ Jacobi identity in the same way to show that Q_B is nilpotent when the total central charge is zero.

4.4 In more general situations, one encounters *graded Lie algebras*, with bosonic generators (fermion number F_I even) and fermionic generators (F_I odd). The algebra of constraints is

$$G_I G_J - (-1)^{F_I F_J} G_J G_I = i g^K{}_{IJ} G_K \,.$$

Construct a nilpotent BRST operator. You will need bosonic ghosts for the fermionic constraints.

4.5 (a) Carry out the BRST quantization for the first two levels of the closed string explicitly.
(b) Carry out the BRST quantization for the third level of the open string, $m^2 = 1/\alpha'$.

4.6 (a) Consider the operator $Q' = \alpha_{-1}^{-} c_1 + \alpha_1^{-} c_{-1}$, obtained by truncating Q_1 to the $m = \pm 1$ oscillators only. Calculate the cohomology of Q' directly, by considering its action on a general linear combination of the states

$$(b_{-1})^{N_b}(c_{-1})^{N_c}(\alpha_{-1}^{+})^{N^+}(\alpha_{-1}^{-})^{N^-}|0\rangle \ .$$

(b) Generalize this to the full Q_1.

5
The string S-matrix

In chapter 3 we wrote the string S-matrix as a path integral over compact two-dimensional surfaces with vertex operators. In this chapter we will reduce the path integral to gauge-fixed form.

5.1 The circle and the torus

We would like to identify a gauge slice, a choice of one configuration from each (diff×Weyl)-equivalent set as in figure 3.7. Locally we did this by fixing the metric, but globally there is a small mismatch between the space of metrics and the world-sheet gauge group.

Once again the point particle is a good illustration. We want to evaluate the Euclidean path integral

$$\int [de\, dX] \exp\left[-\frac{1}{2}\int d\tau\, (e^{-1}\dot{X}^\mu \dot{X}_\mu + em^2)\right] \ . \tag{5.1.1}$$

Consider a path forming a closed loop in spacetime, so the topology is a circle. The parameter τ can be taken to run from 0 to 1 with the endpoints identified. That is, $X^\mu(\tau)$ and $e(\tau)$ are periodic on $0 \le \tau \le 1$. The tetrad $e(\tau)$ has one component and there is one local symmetry, the choice of parameter, so as for the string there is just enough local symmetry to fix the tetrad completely. The tetrad transforms as $e'd\tau' = ed\tau$. The gauge choice $e' = 1$ thus gives a differential equation for $\tau'(\tau)$,

$$\frac{\partial \tau'}{\partial \tau} = e(\tau) \ . \tag{5.1.2}$$

Integrating this with the boundary condition $\tau'(0) = 0$ determines

$$\tau'(\tau) = \int_0^\tau d\tau''\, e(\tau'') \ . \tag{5.1.3}$$

The complication is that in general $\tau'(1) \neq 1$, so the periodicity is not preserved. In fact,

$$\tau'(1) = \int_0^1 d\tau\, e(\tau) = l \tag{5.1.4}$$

is the invariant length of the circle. So we cannot simultaneously set $e' = 1$ and keep the coordinate region fixed. We can hold the coordinate region fixed and set e' to the constant value $e' = l$, or set $e' = 1$ and let the coordinate region vary:

$$e' = l\,, \quad 0 \leq \tau \leq 1\,, \tag{5.1.5a}$$

or

$$e' = 1\,, \quad 0 \leq \tau \leq l\,. \tag{5.1.5b}$$

In either case, after fixing the gauge invariance we are left with an ordinary integral over l. In other words, not all tetrads on the circle are diff-equivalent. There is a one-parameter family of inequivalent tetrads, parameterized by l.

Both descriptions (5.1.5) will have analogs in the string. In practice, we will define the path integral using a description like (5.1.5a), where the fields are functions on a fixed coordinate region, and then convert to a description like (5.1.5b) where the metric is fixed and the moduli are encoded in the coordinate range.

There is a second complication with the gauge-fixing. The condition $e =$ constant is preserved by the rigid translation

$$\tau \to \tau + v \mod 1\,. \tag{5.1.6}$$

That is, we have not said where on the circle to put the origin of the coordinate system. Fixing the metric thus leaves a small part of the local symmetry unfixed. So there is a small mismatch in *both* directions, metrics that are not gauge-equivalent to one another and gauge transformations that are not fixed by the choice of metric.

Moving on to the string we take as an example the torus, where the same complications arise. Start with the coordinate region

$$0 \leq \sigma^1 \leq 2\pi\,, \quad 0 \leq \sigma^2 \leq 2\pi\,, \tag{5.1.7}$$

with $X^\mu(\sigma^1, \sigma^2)$ and $g_{ab}(\sigma_1, \sigma_2)$ periodic in both directions. Equivalently we can think of this as the σ plane with the identification of points,

$$(\sigma^1, \sigma^2) \cong (\sigma^1, \sigma^2) + 2\pi(m, n) \tag{5.1.8}$$

for integer m and n.

To what extent is the field space diff×Weyl redundant? The theorem is that it is *not* possible to bring a general metric to unit form by a diff×Weyl transformation that leaves invariant the periodicity (5.1.7), but it *is* possible to bring it to the form

$$ds^2 = |d\sigma^1 + \tau d\sigma^2|^2 \tag{5.1.9}$$

where τ is a complex constant. For $\tau = i$ this would be the unit metric δ_{ab}.

To see this, repeat the steps used in the local discussion in section 3.3. We can first make the metric flat by a Weyl transformation satisfying $2\nabla^2\omega = R$. By expanding in eigenmodes of ∇^2, one sees that with the periodic boundary conditions this has a solution unique up to addition of a constant to ω. It is important here that $\int d^2\sigma\, g^{1/2}R$, which is 4π times the Euler number, vanishes for the torus. A transformation to new coordinates $\tilde{\sigma}^a$ then brings the metric to unit form. However, as in the point-particle case, there is no guarantee that this respects the original periodicity. Rather, we may now have

$$\tilde{\sigma}^a \cong \tilde{\sigma}^a + 2\pi(mu^a + nv^a) \tag{5.1.10}$$

with general translations u^a and v^a. By rotating and rescaling the coordinate system, accompanied by a shift in ω to keep the metric normalized, we can always set $u = (1, 0)$. This leaves two parameters, the components of v. Defining $w = \tilde{\sigma}^1 + i\tilde{\sigma}^2$, the metric is $dwd\bar{w}$ and the periodicity is

$$w \cong w + 2\pi(m + n\tau)\,, \tag{5.1.11}$$

where $\tau = v^1 + iv^2$. The torus is a parallelogram in the w-plane with periodic boundary conditions, as shown in figure 5.1.

Alternatively, define σ^a by $w = \sigma^1 + \tau\sigma^2$. The original periodicity (5.1.8) is preserved but the metric now takes the more general form (5.1.9). The integration over metrics reduces to two ordinary integrals, over the real and imaginary parts of τ. The metric (5.1.9) is invariant under complex

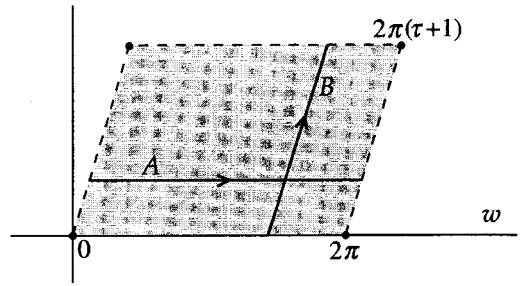


Fig. 5.1. Torus with modulus τ, in gauge $ds^2 = dwd\bar{w}$. The upper and lower edges are identified, as are the right- and left-hand edges. Closed curves A and B are marked for later reference.

conjugation of τ and degenerate for τ real, so we can restrict attention to $\mathrm{Im}\,\tau > 0$. As in the case (5.1.5) of the circle, we can put these parameters either in the metric (5.1.9) or the periodicity (5.1.11). The parameter τ is known as a *Teichmüller parameter* or more commonly a *modulus*. A difference from the point-particle case is that there is no simple invariant expression for τ analogous to the invariant length of the circle.

There is some additional redundancy, which does not have an analog in the point-particle case. The value $\tau + 1$ generates the same set of identifications (5.1.11) as τ, replacing $(m, n) \to (m-n, n)$. So also does $-1/\tau$, defining $w' = \tau w$ and replacing $(m, n) \to (n, -m)$. Repeated application of these two transformations,

$$T: \ \tau' = \tau + 1 \ , \quad S: \ \tau' = -1/\tau \ , \tag{5.1.12}$$

generates

$$\tau' = \frac{a\tau + b}{c\tau + d} \tag{5.1.13}$$

for all integer a, b, c, d such that $ad - bc = 1$.

We can also think about this as follows. The transformation

$$\begin{bmatrix} \sigma^1 \\ \sigma^2 \end{bmatrix} = \begin{bmatrix} d & b \\ c & a \end{bmatrix} \begin{bmatrix} \sigma'^1 \\ \sigma'^2 \end{bmatrix} \tag{5.1.14}$$

takes the metric (5.1.9) for σ into a metric of the same form in σ' but with the modulus τ'. This is a diffeomorphism of the torus. It is one-to-one as a consequence of $ad - bc = 1$ and it preserves the periodicity (5.1.8). However, it cannot be obtained from the identity by successive infinitesimal transformations — it is a so-called *large coordinate transformation*. The curve A in the coordinate σ maps to a curve in the σ' coordinate that runs a times in the A' direction and $-c$ times in the B' direction. These large coordinate transformations form the group $SL(2, \mathbf{Z})$, integer-valued 2×2 matrices with unit determinant. The group on the τ-plane is $SL(2, \mathbf{Z})/\mathbf{Z}_2 = PSL(2, \mathbf{Z})$, because τ' is unchanged if all the signs of a, b, c, d are reversed. The group of transformations (5.1.14) is known as the *modular group*.

Using the modular transformations (5.1.13), it can be shown that every τ is equivalent to exactly one point in the region F_0 shown in figure 5.2,

$$-\frac{1}{2} \le \mathrm{Re}\,\tau \le \frac{1}{2} \ , \quad |\tau| \ge 1 \ , \tag{5.1.15}$$

except on the boundaries which are identified as shown in the figure. This is called a *fundamental region* for the upper half-plane mod $PSL(2, \mathbf{Z})$. Because of the identifications one should think of F_0 as being rolled up, and open only at $\mathrm{Im}\,\tau \to \infty$. The fundamental region F_0 is one representation of the moduli space of (diff×Weyl)-inequivalent metrics.

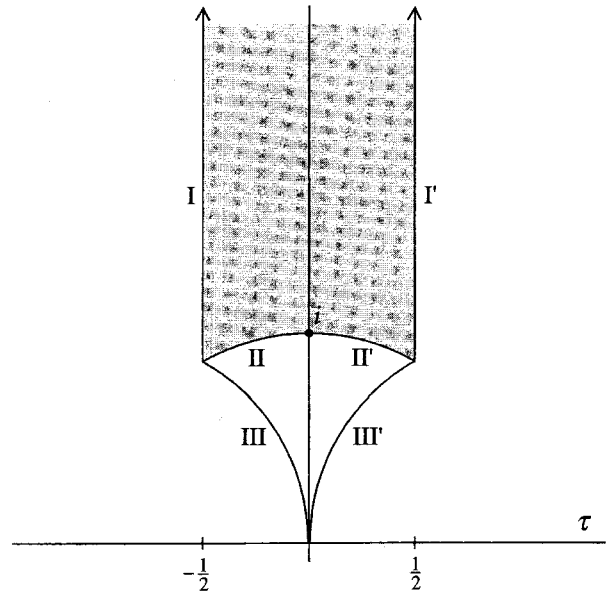


Fig. 5.2. The standard fundamental region F_0 for the moduli space of the torus, shaded. The lines I and I′ are identified, as are the arcs II and II′. A different fundamental region, mapped into F_0 by the modular transformation S, is bounded by II, II′, III, and III′.

There is a further complication as there was for the point particle. Requiring the metric to take the form (5.1.9) with τ in a given fundamental region does not fix all of the diff×Weyl invariance. The metric and periodicity are left invariant by *rigid* translations,

$$\sigma^a \to \sigma^a + v^a \ , \tag{5.1.16}$$

so this two-parameter subgroup of diff×Weyl is not fixed. In addition, the discrete transformation $\sigma^a \to -\sigma^a$ leaves the metric invariant, and in the unoriented case also $(\sigma^1, \sigma^2) \to (-\sigma^1, \sigma^2)$ with $\tau \to -\bar{\tau}$. Thus there are again two kinds of mismatch between the metric degrees of freedom and the local symmetries: parameters in the metric that cannot be removed by the symmetries, and symmetries that are not fixed by the choice of metric. The unfixed symmetries are known as the *conformal Killing group (CKG)*.

When there are enough vertex operators in the amplitude, we can fully fix the gauge invariance by fixing some of the vertex operator positions. On the torus with n vertex operators, the invariance (5.1.16) can be used

to fix one vertex operator position, leaving an integral over τ and the $n-1$ other positions. The $\mathbf{Z}_2$ from $\sigma^a \to -\sigma^a$ can be fixed by restricting a second vertex operator to half the torus. Actually, for finite overcounting it is often more convenient to leave some symmetry unfixed and divide by an appropriate factor. Also, if there are too few vertex operators, some conformal Killing invariance remains unfixed and we must allow explicitly for the overcounting.

A note to the reader

In the remainder of this chapter we discuss moduli in more generality, and obtain the measure for the gauge-fixed integral. The principal goal is the gauge-fixed expression (5.3.9) for the string S-matrix. In fact, most of the interesting string physics can be seen in the tree-level and one-loop amplitudes. For these the measure can be obtained by shortcuts. For the tree-level amplitudes there are no metric moduli but some vertex operator positions must be fixed. The measure can be deduced as explained below eq. (6.4.4). For one-loop amplitudes the measure can be understood in a fairly intuitive way, by analogy to the discussion of the point particle measure below eq. (7.3.9). Thus it is possible for the reader to skip the remainder of this chapter (or to skim as far as eq. (5.3.9)), and over all sections dealing with the Faddeev–Popov ghosts.

5.2 Moduli and Riemann surfaces

Now let us repeat the preceding discussion in a more general and abstract way. We start with the integral over all metrics on some given topology r. Call the space of metrics $\mathscr{G}_r$. For closed oriented surfaces we can label r simply by the number g of handles, also known as the *genus*. After taking into account the diff×Weyl redundancy, we are left with the moduli space

$$\mathscr{M}_r = \frac{\mathscr{G}_r}{(\text{diff} \times \text{Weyl})_r} . \tag{5.2.1}$$

As in the case of the torus, this space is parameterized by a finite number of *moduli*. For the torus, $\mathscr{M}_1$ is the upper half-plane mod $PSL(2,\mathbf{Z})$, or equivalently any fundamental region, say F_0. There may also be a subgroup of diff×Weyl, the CKG, that leaves the metric invariant.

When there are vertex operators in the path integral it is useful to treat their positions on the same footing as the moduli from the metric, referring to all as moduli. We specify the *metric moduli* when we need to make the distinction. One way to deal with the CKG, which is applicable if there are vertex operators in the path integral, is to specify the gauge further by fixing the coordinates of some vertex operators. The Polyakov

path integral includes an integral over $\mathscr{G}_r$ and integrals of the n vertex operator positions over the world-sheet $\mathscr{M}$. The moduli space at topology r with n vertex operators is then

$$\mathscr{M}_{r,n} = \frac{\mathscr{G}_r \times \mathscr{M}^n}{(\text{diff} \times \text{Weyl})_r} . \tag{5.2.2}$$

As we have seen for the torus, diff_r is not in general connected. Picking out the connected component diff_{r0} that contains the identity, the quotient

$$\frac{\text{diff}_r}{\text{diff}_{r0}} \tag{5.2.3}$$

is the modular group.

It is interesting to study the diff×Weyl redundancy of the metric at the infinitesimal level. That is, we are looking for small variations of the metric that are not equivalent to diff×Weyl transformations and so correspond to changes in the moduli. We are also looking for small diff×Weyl transformations that do not change the metric; these would be infinitesimal elements of the CKG, called *conformal Killing vectors (CKVs)*. This will give further insight into the origin of the moduli and the CKG.

An infinitesimal diff×Weyl transformation changes the metric by

$$\delta g_{ab} = -2(P_1 \delta\sigma)_{ab} + (2\delta\omega - \nabla \cdot \delta\sigma) g_{ab} , \tag{5.2.4}$$

where P_1 is the traceless symmetric linear combination of derivatives (3.3.17). Moduli correspond to variations $\delta' g_{ab}$ of the metric that are orthogonal to all variations (5.2.4):

$$\begin{aligned} 0 &= \int d^2\sigma\, g^{1/2}\, \delta' g_{ab} \left[-2(P_1 \delta\sigma)^{ab} + (2\delta\omega - \nabla \cdot \delta\sigma) g^{ab} \right] \\ &= \int d^2\sigma\, g^{1/2} \left[-2(P_1^T \delta' g)_a \delta\sigma^a + \delta' g_{ab} g^{ab} (2\delta\omega - \nabla \cdot \delta\sigma) \right] . \end{aligned} \tag{5.2.5}$$

The transpose $(P_1^T u)_a = -\nabla^b u_{ab}$ is defined as in exercise 3.2. In order for the overlap (5.2.5) to vanish for general $\delta\omega$ and $\delta\sigma$ we need

$$g^{ab} \delta' g_{ab} = 0 , \tag{5.2.6a}$$

$$(P_1^T \delta' g)_a = 0 . \tag{5.2.6b}$$

The first condition requires that $\delta' g_{ab}$ be traceless, and it is on traceless symmetric tensors that P_1^T acts. For each solution of these equations there will be a modulus.

CKVs are transformations (5.2.4) such that $\delta g_{ab} = 0$. The trace of this equation fixes $\delta\omega$ uniquely, leaving the *conformal Killing equation*

$$(P_1 \delta\sigma)_{ab} = 0 . \tag{5.2.7}$$

Eqs. (5.2.6) and (5.2.7) become simple for variations around conformal gauge,

$$\partial_{\bar{z}}\delta' g_{zz} = \partial_z \delta' g_{\bar{z}\bar{z}} = 0, \tag{5.2.8a}$$

$$\partial_{\bar{z}}\delta z = \partial_z \delta \bar{z} = 0\,, \tag{5.2.8b}$$

so variations of the moduli correspond to *holomorphic quadratic differentials* and CKVs to *holomorphic vector fields*. On the torus, the only holomorphic doubly periodic functions are the constants, so there are two real moduli and two real CKVs, in agreement with the discussion in the previous section.

The metric moduli correspond to the kernel (5.2.6) of P_1^T and the CKVs to the kernel (5.2.7) of P_1. The Riemann–Roch theorem relates the number of metric moduli $\mu = \dim\ker P_1^T$ and the number of CKVs $\kappa = \dim\ker P_1$ to the Euler number χ,

$$\mu - \kappa = -3\chi\,. \tag{5.2.9}$$

We will be able to derive this later in the chapter. For a closed oriented surface $-3\chi = 6g - 6$. This counting refers to real moduli, separating complex moduli such as τ into their real and imaginary parts.

Further, κ vanishes for $\chi < 0$ and μ vanishes for $\chi > 0$. To show this, we first cite the fact that it is always possible by a Weyl transformation to find a metric for which the scalar curvature R is constant. The sign of R is then the same as that of χ. Now, $P_1^T P_1 = -\frac{1}{2}\nabla^2 - \frac{1}{4}R$, so

$$\begin{aligned}\int d^2\sigma\, g^{1/2}(P_1\delta\sigma)_{ab}(P_1\delta\sigma)^{ab} &= \int d^2\sigma\, g^{1/2}\delta\sigma_a(P_1^T P_1\delta\sigma)^a \\ &= \int d^2\sigma\, g^{1/2}\left(\frac{1}{2}\nabla_a\delta\sigma_b\nabla^a\delta\sigma^b - \frac{R}{4}\delta\sigma_a\delta\sigma^a\right).\end{aligned} \tag{5.2.10}$$

For negative χ the right-hand side is strictly positive, so $P_1\delta\sigma$ cannot vanish. A similar argument shows that $P_1^T\delta' g$ cannot vanish for positive χ. Combining these results we have

$$\chi > 0: \quad \kappa = 3\chi\,, \quad \mu = 0\,, \tag{5.2.11a}$$

$$\chi < 0: \quad \kappa = 0\,, \quad \mu = -3\chi\,. \tag{5.2.11b}$$

Riemann surfaces

The familiar way to describe moduli space would be to choose one metric from each equivalence class. This would be a family of metrics $\hat{g}_{ab}(t;\sigma)$, depending on the moduli t^k. For the torus, eq. (5.1.9) gives such a slice. It is often convenient to use an alternative description of the

form (5.1.11), where the metric $dwd\bar{w}$ is fixed and the moduli are encoded in the coordinate region. We can formalize this idea as follows.

Let us first recall how *differentiable* manifolds are defined. One covers the manifold with a set of overlapping patches, with coordinates σ_m^a in the m^{th} patch, with a running from 1 to the dimension of the manifold. When patches m and n overlap, the coordinates in the two patches are related by

$$\sigma_m^a = f_{mn}^a(\sigma_n)\,, \tag{5.2.12}$$

where the transition functions f_{mn} are required to be differentiable a given number of times. For a *Riemannian* manifold a metric $g_{m,ab}(\sigma_m)$ is also given in each patch, with the usual tensor transformation law relating the values in the overlaps.

For a *complex* manifold, there are complex coordinates z_m^a in each patch, where now a runs from 1 to half the dimension of the manifold. The transition functions are required to be holomorphic,

$$z_m^a = f_{mn}^a(z_n)\,. \tag{5.2.13}$$

One can now define holomorphic functions on the manifold, since the holomorphicity will not depend on which coordinates z_m^a are used. Just as two differentiable manifolds are equivalent if there is a one-to-one differentiable map between them, two complex manifolds are equivalent if there is a one-to-one holomorphic map between them. In particular, making a holomorphic change of coordinates within each patch gives an equivalent surface.

In the two-dimensional case (one complex coordinate), a complex manifold is known as a *Riemann surface*. In this case there is a one-to-one correspondence

$$\text{Riemann surfaces} \leftrightarrow \text{Riemannian manifolds mod Weyl}\,. \tag{5.2.14}$$

On the right we put only 'mod Weyl' because 'mod diff' is already implicit in the definition of a Riemannian manifold. To see this isomorphism, start with a Riemannian manifold. We know from our discussion of conformal gauge that we can find in each coordinate patch a coordinate z_m such that

$$ds^2 \propto dz_m d\bar{z}_m\,. \tag{5.2.15}$$

In neighboring patches the coordinate need not be the same, but because $ds^2 \propto dz_m d\bar{z}_m \propto dz_n d\bar{z}_n$, the transition functions will be holomorphic.[1] This is the map from Riemannian manifolds to complex manifolds in two real dimensions. For the inverse map, one can take the metric $dz_m d\bar{z}_m$ in

[1] To be precise, an antiholomorphic transition function is also possible. For the oriented string we forbid this by choosing a standard orientation; for the unoriented string it is allowed. The term Riemann surface in the strict sense refers to an oriented manifold without boundary.

the m^{th} patch and smooth it in the overlaps (this is a standard construction, the partition of unity) to produce a Riemannian manifold. The two characterizations of a Riemann surface are thus equivalent.

The description of the torus in terms of the complex coordinate w of the parallelogram (figure 5.1) illustrates the idea of a complex manifold. Imagine taking a single coordinate patch which is a little larger than the fundamental parallelogram in figure 5.1. The periodicity conditions

$$w \cong w + 2\pi \cong w + 2\pi\tau \tag{5.2.16}$$

are then the transition functions on the overlap between the opposite edges of the patch. The surface is defined by these transition functions. To define the original path integral over metrics, it is simplest to start by regarding the metric as a function of fixed coordinates such as the square (5.1.7), or at least on a fixed set of coordinate patches with fixed transition functions. To study the quantum field theory on a given surface, it then is easiest to work with the unit metric with moduli-dependent transition functions, as in the parallelogram (5.1.11).

One can also think of a Riemann surface as follows. Defining the world-sheet as a union of patches, we can use the diff×Weyl freedom to reach the metric $dzd\bar{z}$ in each. The gauge choices on overlapping patches can then differ only by diff×Weyl invariances of this metric, which as we have discussed are the conformal transformations. A Riemann surface is therefore the natural place for a CFT to live, because the fields in a CFT have definite transformation properties under conformal transformations.

5.3 The measure for moduli

We now revisit the gauge-fixing of the Polyakov path integral. We have learned that the gauge redundancy does not completely eliminate the path integral over metrics but leaves behind a finite-dimensional integral over moduli space. It is necessary to refine the discussion from section 3.3 in order to take this into account.

The Polyakov path integral for the S-matrix is

$$S_{j_1\dots j_n}(k_1,\dots,k_n) = \sum_{\substack{\text{compact}\\ \text{topologies}}} \int \frac{[d\phi\, dg]}{V_{\text{diff}\times\text{Weyl}}} \exp(-S_{\text{m}} - \lambda\chi) \prod_{i=1}^{n} \int d^2\sigma_i\, g(\sigma_i)^{1/2} \mathscr{V}_{j_i}(k_i,\sigma_i)\,, \tag{5.3.1}$$

which is the earlier expression (3.5.5) but now written with a general $c = \tilde{c} = 26$ matter theory, the matter fields being denoted ϕ. In gauge-fixing, the integral over metrics and positions is converted to an integral over the gauge group, the moduli, and the unfixed positions,

$$[dg]\, d^{2n}\sigma \to [d\zeta]\, d^{\mu}t\, d^{2n-\kappa}\sigma\,. \tag{5.3.2}$$

After factoring out the gauge volume, the Jacobian for this transformation becomes the measure on moduli space.

The steps are the same as in section 3.3, now taking account of the moduli and the CKVs. In particular, the gauge choice will now fix κ of the vertex operator coordinates, $\sigma_i^a \to \hat{\sigma}_i^a$. Label the set of fixed coordinates (a,i) by f. Define the Faddeev–Popov measure on moduli space by

$$1 = \Delta_{\text{FP}}(g,\sigma) \int_F d^{\mu}t \int_{\text{diff}\times\text{Weyl}} [d\zeta]\, \delta(g - \hat{g}(t)^{\zeta}) \prod_{(a,i)\in f} \delta(\sigma_i^a - \hat{\sigma}_i^{\zeta a})\,. \tag{5.3.3}$$

By definition, every metric is (diff×Weyl)-equivalent to $\hat{g}(t)$ for at least one value of t and ζ. The delta function for the fixed coordinates then picks out a unique value of ζ. Actually, as discussed at the end of section 5.1, there may be a residual discrete group of symmetries of finite order n_R, so the delta functions are nonzero at n_R points.

Inserting the expression (5.3.3) into the path integral (5.3.1) and following the same steps as before leads to

$$S_{j_1\dots j_n}(k_1,\dots,k_n) = \sum_{\substack{\text{compact}\\ \text{topologies}}} \int_F d^{\mu}t\, \Delta_{\text{FP}}(\hat{g}(t),\hat{\sigma}) \int [d\phi] \int \prod_{(a,i)\notin f} d\sigma_i^a \times \exp\Big(-S_{\text{m}}[\phi,\hat{g}(t)] - \lambda\chi\Big) \prod_{i=1}^{n} \Big[\hat{g}(\sigma_i)^{1/2} \mathscr{V}_{j_i}(k_i;\sigma_i)\Big]\,. \tag{5.3.4}$$

The integral over metrics and vertex operator coordinates is now reduced to an integral over the moduli space F for the metric and over the unfixed coordinates, with the measure given by Δ_{FP}. In the vertex operators, κ of the positions are fixed.

Now we evaluate the Faddeev–Popov measure. The delta functions are nonzero at n_R points that are related by symmetry, so we consider one such point and divide by n_R. Expand the definition (5.3.3) of Δ_{FP} near the point. The general metric variation is equal to a local symmetry variation plus a change in the moduli t^k,

$$\delta g_{ab} = \sum_{k=1}^{\mu} \delta t^k \partial_{t^k} \hat{g}_{ab} - 2(\hat{P}_1 \delta\sigma)_{ab} + (2\delta\omega - \partial\cdot\hat{\nabla}\sigma)\hat{g}_{ab}\,. \tag{5.3.5}$$

The inverse Faddeev–Popov determinant is then

$$\Delta_{\text{FP}}(\hat{g},\hat{\sigma})^{-1} = n_R \int d^{\mu}\delta t\, [d\delta\omega\, d\delta\sigma]\, \delta(\delta g_{ab}) \prod_{(a,i)\in f} \delta(\delta\sigma^a(\hat{\sigma}_i))$$

$$= n_R \int d^\mu \delta t\, d^\kappa x\, [d\beta'\, d\delta\sigma] \times \exp\left[2\pi i(\beta', 2\hat{P}_1 \delta\sigma - \delta t^k \partial_k \hat{g}) + 2\pi i \sum_{(a,i)\in f} x_{ai} \delta\sigma^a(\hat{\sigma}_i) \right] , \quad (5.3.6)$$

the inner product in the second line having been defined in exercise 3.2. We have followed the same steps as in the local discussion in section 3.3, writing the delta functions and functionals as integrals over x and β_{ab} and then integrating out $\delta\omega$ to obtain the constraint that β'_{ab} be traceless.

As before, invert the integral (5.3.6) by replacing all bosonic variables with Grassmann variables:

$$\delta\sigma^a \to c^a , \quad (5.3.7a)$$
$$\beta'_{ab} \to b_{ab} , \quad (5.3.7b)$$
$$x_{ai} \to \eta_{ai} , \quad (5.3.7c)$$
$$\delta t^k \to \xi^k . \quad (5.3.7d)$$

Then with a convenient normalization of the fields,

$$\Delta_{FP}(\hat{g}, \hat{\sigma}) = \frac{1}{n_R} \int [db\, dc]\, d^\mu \xi\, d^\kappa \eta \times \exp\left[-\frac{1}{4\pi}(b, 2\hat{P}_1 c - \xi^k \partial_k \hat{g}) + \sum_{(a,i)\in f} \eta_{ai} c^a(\hat{\sigma}_i) \right]$$
$$= \frac{1}{n_R} \int [db\, dc]\, \exp(-S_g) \prod_{k=1}^{\mu} \frac{1}{4\pi}(b, \partial_k \hat{g}) \prod_{(a,i)\in f} c^a(\hat{\sigma}_i) . \quad (5.3.8)$$

In the final line we have integrated over the Grassmann parameters η_{ai} and ξ^k.

The appropriate measure for integration on moduli space is generated by this ghost path integral with insertions. We will not try to keep track of the overall sign at intermediate steps; since this is supposed to be a Jacobian, we implicitly choose the overall sign to give a positive result. The full expression for the S-matrix is then

$$S_{j_1 \dots j_n}(k_1, \dots, k_n) = \sum_{\substack{\text{compact} \\ \text{topologies}}} \int_F \frac{d^\mu t}{n_R} \int [d\phi\, db\, dc]\, \exp(-S_m - S_g - \lambda\chi) \times \prod_{(a,i)\notin f} \int d\sigma_i^a \prod_{k=1}^{\mu} \frac{1}{4\pi}(b, \partial_k \hat{g}) \prod_{(a,i)\in f} c^a(\hat{\sigma}_i) \prod_{i=1}^{n} \hat{g}(\sigma_i)^{1/2} \mathscr{V}_{j_i}(k_i, \sigma_i) . \quad (5.3.9)$$

This result extends readily to all bosonic string theories — closed and open, oriented and unoriented — the only difference being which topologies are included and which vertex operators allowed. Eq. (5.3.9) is a useful and

elegant result. The complications caused by the moduli and the CKVs are taken into account by the c and b insertions in the path integral. For each fixed coordinate, $d\sigma_i^a$ is replaced by c_i^a, while each metric modulus gives rise to a b insertion.

Expression in terms of determinants

In eq. (5.3.8) we have represented the Faddeev–Popov determinant as an integral over Grassmann parameters and fields. We now reduce it to a product of finite-dimensional and functional determinants. When we go on to study the string S-matrix in coming chapters, this direct path integral approach is just one of the methods we will use.

Expand the ghost fields in suitable complete sets,

$$c^a(\sigma) = \sum_J c_J \mathsf{C}_J^a(\sigma) , \quad b_{ab}(\sigma) = \sum_K b_K \mathsf{B}_{Kab}(\sigma) . \quad (5.3.10)$$

The complete sets C_J^a and B_{Kab} are defined as follows. The derivative P_1 appearing in the ghost action takes vectors into traceless symmetric tensors. Its transpose does the reverse. The ghost action can be written with either,

$$S_g = \frac{1}{2\pi}(b, P_1 c) = \frac{1}{2\pi}(P_1^T b, c) . \quad (5.3.11)$$

We cannot diagonalize P_1 in a diff-invariant way because it turns one kind of field into another, but we can diagonalize $P_1^T P_1$ and $P_1 P_1^T$:

$$P_1^T P_1 \mathsf{C}_J^a = \nu_J'^2 \mathsf{C}_J^a , \quad P_1 P_1^T \mathsf{B}_{Kab} = \nu_K^2 \mathsf{B}_{Kab} . \quad (5.3.12)$$

The eigenfunctions can be chosen real, and are normalized in the respective inner products

$$(\mathsf{C}_J, \mathsf{C}_{J'}) = \int d^2\sigma\, g^{1/2} \mathsf{C}_J^a \mathsf{C}_{J'a} = \delta_{JJ'} , \quad (5.3.13a)$$
$$(\mathsf{B}_K, \mathsf{B}_{K'}) = \int d^2\sigma\, g^{1/2} \mathsf{B}_{Kab} \mathsf{B}_{K'}^{ab} = \delta_{KK'} . \quad (5.3.13b)$$

Now note that

$$(P_1 P_1^T) P_1 \mathsf{C}_J = P_1 (P_1^T P_1) \mathsf{C}_J = \nu_J'^2 P_1 \mathsf{C}_J , \quad (5.3.14)$$

so $P_1 \mathsf{C}_J$ is an eigenfunction of $P_1 P_1^T$. In the same way, $P_1^T \mathsf{B}_K$ is an eigenfunction of $P_1^T P_1$. Thus there is a one-to-one correspondence between the eigenfunctions, *except* when $P_1 \mathsf{C}_J = 0$ or $P_1^T \mathsf{B}_K = 0$. The latter correspond to zero eigenvalues of $P_1^T P_1$ or $P_1 P_1^T$. These are just the CKVs and holomorphic quadratic differentials, and so their numbers are respectively κ and μ. Let us label the eigenfunctions with zero eigenvalue as C_{0j} or B_{0k}, while the nonzero eigenvalues are labeled with $J, K = 1, \dots$.

For the latter, the normalized eigenfunctions are related

$$\mathsf{B}_{Jab} = \frac{1}{v_J}(P_1 \mathsf{C}_J)_{ab} , \quad v_J = v_J' \neq 0 . \tag{5.3.15}$$

In terms of modes, the ghost path integral $\Delta_{\rm FP}$ becomes

$$\int \prod_{k=1}^{\mu} db_{0k} \prod_{j=1}^{\kappa} dc_{0j} \prod_J db_J\, dc_J \exp\left(-\frac{v_J b_J c_J}{2\pi}\right) \prod_{k=1}^{\mu} \frac{1}{4\pi}(b, \partial_k \hat{g}) \prod_{(a,i)\in f} c^a(\sigma_i) . \tag{5.3.16}$$

From section A.2, a Grassmann integral vanishes unless the variable appears in the integrand. The c_{0j} and b_{0k} do not appear in the action but only in the insertions. In fact, the number of insertions of each type in the integral (5.3.16), κ and μ respectively, just matches the number of ghost zero modes. We have just enough insertions to give a nonzero answer, and only the zero-mode part of the ghost fields contributes in the insertions. Thus,

$$\begin{aligned}\Delta_{\rm FP} = \int \prod_{k=1}^{\mu} db_{0k} \prod_{k'=1}^{\mu} \left[\sum_{k''=1}^{\mu} \frac{b_{0k''}}{4\pi}(\mathsf{B}_{0k''}, \partial_{k'}\hat{g})\right] \\ \times \int \prod_{j=1}^{\kappa} dc_{0j} \prod_{(a,i)\in f} \left[\sum_{j'=1}^{\kappa} c_{0j'} \mathsf{C}^a_{0j'}(\sigma_i)\right] \\ \times \int \prod_J db_J\, dc_J \exp\left(-\frac{v_J b_J c_J}{2\pi}\right) .\end{aligned} \tag{5.3.17}$$

Summing over all ways of saturating the zero-mode integrals with the Grassmann variables generates a finite-dimensional determinant in each case, while the nonzero modes produce an infinite product, a functional determinant. In all,

$$\Delta_{\rm FP} = \det \frac{(\mathsf{B}_{0k}, \partial_{k'}\hat{g})}{4\pi} \det \mathsf{C}^a_{0j}(\sigma_i) \det{}' \left(\frac{P_1^T P_1}{2\pi}\right)^{1/2} . \tag{5.3.18}$$

Note that $\mathsf{C}^a_{0j}(\sigma_i)$ is a square matrix, in that $(a,i) \in f$ runs over κ values as does j. The prime on the functional determinant denotes omission of zero eigenvalues.

The Riemann–Roch theorem

We can now give a path integral derivation of the Riemann–Roch theorem. For the ghost current of a holomorphic bc system, the current conservation anomaly derived in exercise 3.6 is

$$\nabla_a j^a = \frac{2\lambda - 1}{4} R . \tag{5.3.19}$$

Noether's theorem relates conserved currents to invariances. For a current that is not conserved, one finds by the same argument that

$$\frac{\delta([d\phi]\exp(-S))}{[d\phi]\exp(-S)} = -\frac{i\epsilon}{2\pi}\int d^2\sigma\, g^{1/2}\nabla_a j^a \to -i\frac{2\lambda-1}{2}\chi . \tag{5.3.20}$$

The ghost number symmetry acts as $\delta b = -i\epsilon b$, $\delta c = i\epsilon c$. When the path integral is nonvanishing, the transformations of the measure, action, and insertions must cancel. Thus we learn that from the anomaly the number of c minus the number of b insertions is $3\chi/2$. The path integral calculation relates the number of insertions to the number of zero modes and so gives for the same difference

$$\frac{1}{2}(\kappa - \mu) = \frac{1}{2}(\dim\ker P_1 - \dim\ker P_1^T) . \tag{5.3.21}$$

The factor of $\frac{1}{2}$ appears because we have taken the holomorphic theory only; the antiholomorphic theory gives an equal contribution. Equating the result from the anomaly with the result from the zero modes gives the Riemann–Roch theorem $\kappa - \mu = 3\chi$. For the general bc system one finds in the same way that

$$\dim\ker P_n - \dim\ker P_n^T = (2n+1)\chi , \tag{5.3.22}$$

the operators being defined in exercise 3.2.

5.4 More about the measure

Here we collect some general properties of the gauge-fixed string amplitude. These are primarily useful for the more formal considerations in chapter 9.

Gauge invariance

The Faddeev–Popov procedure guarantees that the gauge-fixed amplitude is BRST-invariant and independent of gauge choice, but it is useful to check explicitly that it has all the expected invariance properties. First, it is independent of the choice of coordinates t on moduli space. For new coordinates $t'^k(t)$,

$$\begin{aligned} d^\mu t' &= \left|\det \frac{\partial t'}{\partial t}\right| d^\mu t , \\ \prod_{k=1}^{\mu} \frac{1}{4\pi}(b, \partial'_k \hat{g}) &= \det\left(\frac{\partial t}{\partial t'}\right) \prod_{k=1}^{\mu} \frac{1}{4\pi}(b, \partial_k \hat{g}) , \end{aligned} \tag{5.4.1}$$

and the two Jacobians cancel (up to a sign, which as we have noted must be fixed by hand to give a positive measure). In other words, due to the b-ghost insertions the integrand transforms as a density on moduli space.

Second, let us see that the measure is invariant under a Weyl transformation of the gauge slice. Under a Weyl transformation,

$$\delta\hat{g}'_{ab}(t;\sigma) = 2\delta\omega(t;\sigma)\hat{g}_{ab}(t;\sigma)\,, \tag{5.4.2}$$

the variation of the action and measure gives an insertion of the local operator $T^a{}_a$, which vanishes in $D = 26$ by the equations of motion. We need therefore worry only about the effect on the various insertions. The vertex operator insertions are Weyl-invariant by construction. The insertions $c^a(\sigma_i)$ are Weyl-invariant as discussed below eq. (3.3.24). For the b insertions,

$$\begin{aligned}(b,\partial_k\hat{g}') &= \int d^2\sigma\,\hat{g}'^{1/2}\,b_{ab}\hat{g}'^{ac}\hat{g}'^{bd}\partial_k\hat{g}'_{cd}\\ &= \int d^2\sigma\,\hat{g}^{1/2}\,b_{ab}(\hat{g}^{ac}\hat{g}^{bd}\partial_k\hat{g}_{cd} + 2\hat{g}^{ab}\partial_k\omega)\\ &= (b,\partial_k\hat{g})\,,\end{aligned} \tag{5.4.3}$$

the last equality following from tracelessness of b.

Third, let us check invariance under an infinitesimal diff transformation, $\delta\sigma = \xi(t;\sigma)$. The extension to a general diff transformation is straightforward. The only terms in the amplitude (5.3.9) that are not immediately invariant are the b insertions and the vertex operators whose coordinates have been fixed. The former transform as

$$\begin{aligned}\delta(b,\partial_k\hat{g}) &= -2(b,P_1\partial_k\xi)\\ &= -2(P_1^T b,\partial_k\xi)\\ &= 0\end{aligned} \tag{5.4.4}$$

using the b equation of motion. The b equation of motion comes from $\delta S/\delta c = 0$, so there will be source terms at the c insertions; these are precisely what is needed to account for the effect of the coordinate transformation on the fixed vertex operators.

BRST invariance

We now verify the full BRST invariance of the amplitude (5.3.9). The path integral measure and action are known to be invariant from the local analysis, so we have to consider the effect of the BRST transformation on the insertions. In the gauge-fixed path integral (5.3.9), the unintegrated vertex operators are accompanied by a factor of c or $\tilde{c}c$. This is precisely what we described at the end of section 4.4 for the BRST-invariant vertex operator corresponding to an OCQ state. The integrated vertex operators,

on the other hand, are not BRST-invariant, but we leave it as an exercise to show that they have the BRST variation

$$\delta_B\mathscr{V}_m = i\epsilon\partial_a(c^a\mathscr{V}_m)\,, \tag{5.4.5}$$

which vanishes upon integration.

Finally there is the variation of the b insertion,

$$\delta_B(b,\partial_k\hat{g}) = i\epsilon(T,\partial_k\hat{g})\,. \tag{5.4.6}$$

This insertion of the energy-momentum tensor just produces a derivative with respect to the modulus t^k, so the BRST variation vanishes up to a possible term from the boundary of moduli space. We will examine the boundaries of moduli space in the coming chapters. In most cases there is no surface contribution, due to what is known as the *canceled propagator argument*. In certain circumstances this argument does not apply, and we will see how to deal with this.

It is important to note that the gauge-fixed result (5.3.9) could be written down directly from the requirement of BRST invariance, without reference to the gauge-invariant form. As we saw when evaluating the ghost path integral, there must be at least μ b insertions and κ c insertions, or else the path integral vanishes. In the amplitude (5.3.9) there are precisely the right number of ghost insertions to give a nonzero result. Once we include the necessary b factors, the BRST variation brings in the energy-momentum tensor convoluted with the t^k-derivative of the metric. This is proportional to a derivative with respect to the moduli, so we get an invariant result only by integrating over moduli space as we have done. The result (5.3.9) can be generalized in various ways, for example to the case that part of the conformal Killing invariance is left unfixed. We will illustrate this explicitly in chapter 7 for the example of the torus with no vertex operators.

BRST invariance implies that amplitudes for BRST-equivalent states are equal. If we add to any state a null piece $Q_B|\chi\rangle$, the effect is to insert in the path integral the variation $\delta_B\mathscr{V}_\chi$; this integrates to zero. This is important because the physical Hilbert space is identified with the cohomology, so equivalent states should have equal amplitudes.

Measure for Riemann surfaces

We derived the Faddeev–Popov measure in the framework where the gauge-fixed world-sheet is described by a moduli-dependent metric. We now recast the result in the Riemann surface framework, where the structure is encoded in moduli-dependent transition functions.

To express the measure at a given point t_0 in moduli space, let us take a set of coordinate patches, with complex coordinate z_m in the m^{th} patch

and holomorphic transition functions, where the metric $\hat{g}(t_0)$ is Weyl-equivalent to $dz_m d\bar{z}_m$ in each patch. Consider now a change in the moduli. We will first describe this as a change in the metric with fixed transition functions, and then convert this to a change in the transition functions with fixed metric. In the first description, define the *Beltrami differential*

$$\mu_{ka}{}^{b} = \frac{1}{2}\hat{g}^{bc}\partial_k \hat{g}_{ac} \,. \tag{5.4.7}$$

The b insertion for δt^k becomes

$$\frac{1}{2\pi}(b, \mu_k) = \frac{1}{2\pi}\int d^2z \left(b_{zz}\mu_{k\bar{z}}{}^{z} + b_{\bar{z}\bar{z}}\mu_{kz}{}^{\bar{z}}\right) \,. \tag{5.4.8}$$

In the second description, after a change δt^k in the moduli there will be new coordinates in each patch

$$z'_m = z_m + \delta t^k v_{km}^{z_m}(z_m, \bar{z}_m) \,. \tag{5.4.9}$$

The superscript on v_{km}^{a} is a vector index; note that v_{km}^{a} is defined only in the m^{th} patch. By the definition of a Riemann surface, $dz'_m d\bar{z}'_m$ is Weyl-equivalent to the metric at $t_0 + \delta t$,

$$dz'_m d\bar{z}'_m \propto dz_m d\bar{z}_m + \delta t^k \left(\mu_{k\bar{z}_m}{}^{z_m} dz_m dz_m + \mu_{kz_m}{}^{\bar{z}_m} d\bar{z}_m d\bar{z}_m\right) \,. \tag{5.4.10}$$

The coordinate change v_{km}^{a} is thus related to the Beltrami differential by

$$\mu_{kz_m}{}^{\bar{z}_m} = \partial_{z_m} v_{km}^{\bar{z}_m} \,, \quad \mu_{k\bar{z}_m}{}^{z_m} = \partial_{\bar{z}_m} v_{km}^{z_m} \,. \tag{5.4.11}$$

This is the infinitesimal version of *Beltrami's equation*. It implies that $v_{km}^{z_m}$ is not holomorphic — otherwise, it would not correspond to a change in the moduli. Also, $v_{km}^{z_m}$ has a holomorphic part not determined by Beltrami's equation, and $v_{km}^{\bar{z}_m}$ an antiholomorphic part; these just correspond to the freedom to make holomorphic reparameterizations within each patch.

Integrating by parts, the b insertion (5.4.8) becomes

$$\frac{1}{2\pi}(b, \mu_k) = \frac{1}{2\pi i}\sum_m \oint_{C_m} \left(dz_m v_{km}^{z_m} b_{z_m z_m} - d\bar{z}_m v_{km}^{\bar{z}_m} b_{\bar{z}_m \bar{z}_m}\right) \,, \tag{5.4.12}$$

the contour C_m circling the m^{th} patch counterclockwise. By eq. (5.4.9), the derivative of the coordinate with respect to the moduli at a given point is

$$\frac{dz_m}{dt^k} = v_{km}^{z_m} \,. \tag{5.4.13}$$

The change in the transition functions under a change in moduli is therefore

$$\left.\frac{\partial z_m}{\partial t^k}\right|_{z_n} = v_{km}^{z_m} - \left.\frac{\partial z_m}{\partial z_n}\right|_{t} v_{kn}^{z_n} = v_{km}^{z_m} - v_{kn}^{z_m} \,. \tag{5.4.14}$$

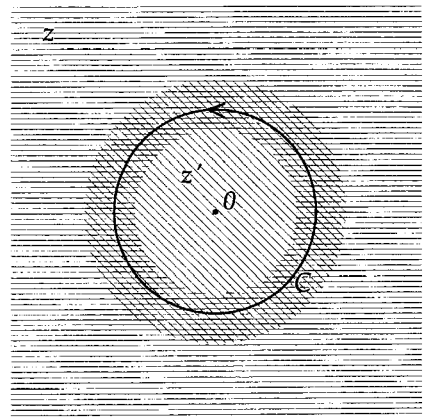


Fig. 5.3. Coordinate patches z' (diagonal hatching) and z (horizontal hatching) with vertex operator at $z' = 0$. In the annular region, $z = z' + z_v$.

The contour integrals (5.4.12) around adjacent patches then combine to give

$$\frac{1}{2\pi}(b, \mu_k) = \frac{1}{2\pi i}\sum_{(mn)}\int_{C_{mn}} \left(dz_m \left.\frac{\partial z_m}{\partial t^k}\right|_{z_n} b_{z_m z_m} - d\bar{z}_m \left.\frac{\partial \bar{z}_m}{\partial t^k}\right|_{z_n} b_{\bar{z}_m \bar{z}_m}\right) \,, \tag{5.4.15}$$

which is proportional to the derivatives of the transition functions. The sum runs over all pairs of overlapping patches. The contour C_{mn} runs between patches m and n, counterclockwise from the point of view of m. The (mn) term is symmetric in m and n, though this is not manifest. Each contour either closes or ends in a triple overlap of patches m, n, and p, where the contours C_{mn}, C_{np}, and C_{pm} meet at a point. The b insertion is now expressed entirely in terms of the data defining the Riemann surface, the transition functions modulo holomorphic equivalence.

As an illustration, we can use this to put the moduli for the metric and for the vertex operator positions on a more equal footing. Consider a vertex operator at position z_v in a coordinate system z. Let us make a small new coordinate patch z' centered on the vertex operator, as shown in figure 5.3. We keep the vertex operator at coordinate point $z' = 0$, and encode its position in the transition functions

$$z = z' + z_v \tag{5.4.16}$$

for the annular overlap region. The measure for the moduli z_v, $\bar{z}_v$ is given

by the form (5.4.15), where

$$\left.\frac{\partial z'}{\partial z_v}\right|_z = -1 \,. \tag{5.4.17}$$

Thus the ghost insertion is

$$\int_C \frac{dz'}{2\pi i} b_{z'z'} \int_C \frac{d\bar{z}'_m}{-2\pi i} b_{\bar{z}'\bar{z}'} = b_{-1}\tilde{b}_{-1} \cdot \,, \tag{5.4.18}$$

where C is any contour encircling the vertex operator as shown in the figure. The full expression for the S-matrix is then compactly written as

$$S(1;\ldots;n) = \sum_{\substack{\text{compact}\\ \text{topologies}}} e^{-\lambda\chi} \int_F \frac{d^m t}{n_R} \left\langle \prod_{k=1}^{m} B_k \prod_{i=1}^{n} \hat{\mathscr{V}}_i \right\rangle . \tag{5.4.19}$$

Here B_k is the abbreviation for the b ghost insertion (5.4.15), and $\hat{\mathscr{V}}$ denotes $\tilde{c}c\mathscr{V}_m$ for a closed string or $t_a c^a \mathscr{V}_m$ for an open string. That is, we are now treating *all* vertex operators as fixed, having traded their coordinates for extra parameters in the transition functions. The number m of moduli is

$$m = \mu + 2n_c + n_o - \kappa = -3\chi + 2n_c + n_o \,, \tag{5.4.20}$$

where n_c and n_o are the number of closed and open string vertex operators respectively.

The expression (5.4.19) shows that the hatted vertex operator, with $\tilde{c}c$ or $t_a c^a$, is the basic one. This is what comes out of the state–operator correspondence, and it is BRST-invariant. If the vertex operator is integrated, the ghost insertions (5.4.18) simply remove the $\tilde{c}c$ or $t_a c^a$. For example,

$$b_{-1}\tilde{b}_{-1} \cdot \tilde{c}c\mathscr{V}_m = \mathscr{V}_m \,, \tag{5.4.21}$$

leaving the integrated form of the vertex operator. We derived the result (5.4.19) from the Polyakov path integral, so the vertex operators come out in OCQ form, but we may now use arbitrary BRST-invariant vertex operators.

Exercises

5.1 Consider the sum (5.1.1), but now over all particle paths beginning and ending at given points in spacetime. In this case fixing the tetrad leaves no coordinate freedom.
(a) Derive the analog of the gauge-fixed path integral (5.3.9).
(b) Reduce the ghost path integral to determinants as in eq. (5.3.18).
(c) Reduce the X^μ path integral to determinants (as will be done in section 6.2 for the string).

(d) Evaluate the finite and functional determinants and show that the result is the scalar propagator for a particle of mass m.

5.2 Repeat the steps in the previous problem for paths forming a closed loop. In order to fix the residual gauge symmetry, introduce a fake vertex operator by differentiating with respect to m^2. The answer is given as eq. (7.3.9). If you do both exercises 5.1 and 5.2, discuss the relation between the results for the moduli measure.

5.3 (a) Show from the OPE with the BRST current that the vertex operator $\tilde{c}c\mathscr{V}_m$ is BRST-invariant for $\mathscr{V}_m$ a (1,1) matter tensor.
(b) Show that the vertex operator satisfies eq. (5.4.5).

6
Tree-level amplitudes

We are now ready to study string interactions. In this chapter we consider the lowest order amplitudes, coming from surfaces with positive Euler number. We first describe the relevant Riemann surfaces and calculate the CFT expectation values that will be needed. We next study scattering amplitudes, first for open strings and then for closed. Along the way we introduce an important generalization in the open string theory, the Chan–Paton factors. At the end of the section we return to CFT, and discuss some general properties of expectation values.

6.1 Riemann surfaces

There are three Riemann surfaces of positive Euler number: the sphere, the disk, and the projective plane.

The sphere

The sphere S_2 can be covered by two coordinate patches as shown in figure 6.1. Take the disks $|z| < \rho$ and $|u| < \rho$ for $\rho > 1$ and join them by identifying points such that

$$u = 1/z \ . \tag{6.1.1}$$

In fact, we may as well take $\rho \to \infty$. The coordinate z is then good everywhere except at the 'north pole,' $u = 0$. We can work mainly in the z patch except to check that things are well behaved at the north pole.

We can think of the sphere as a Riemann surface, taking the flat metric in both patches and connecting them by a conformal (coordinate plus Weyl) transformation. Or, we can think of it as a Riemannian manifold, with a globally defined metric. A general conformal gauge metric is

$$ds^2 = \exp(2\omega(z,\bar{z}))dzd\bar{z} \ . \tag{6.1.2}$$

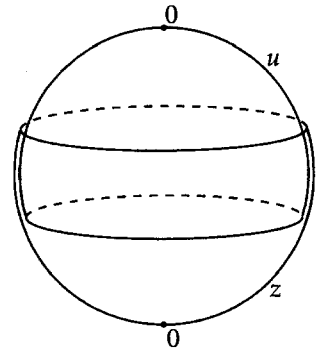


Fig. 6.1. The sphere built from z and u coordinate patches.

Since $dzd\bar{z} = |z|^4 dud\bar{u}$, the condition for the metric to be nonsingular at $u = 0$ is that $\exp(2\omega(z,\bar{z}))$ fall as $|z|^{-4}$ for $z \to \infty$. For example,

$$ds^2 = \frac{4r^2 dzd\bar{z}}{(1+z\bar{z})^2} = \frac{4r^2 dud\bar{u}}{(1+u\bar{u})^2} \tag{6.1.3}$$

describes a sphere of radius r and curvature $R = 2/r^2$.

According to the general discussion in section 5.2, the sphere has no moduli and six CKVs, so that in particular every metric is (diff×Weyl)-equivalent to the round metric (6.1.3). Let us see this at the infinitesimal level. As in eq. (5.2.8) one is looking for holomorphic tensor fields $\delta g_{zz}(z)$ and holomorphic vector fields $\delta z(z)$. These must be defined on the whole sphere, so we need to consider the transformation to the u-patch,

$$\delta u = \frac{\partial u}{\partial z}\delta z = -z^{-2}\delta z \ , \tag{6.1.4a}$$

$$\delta g_{uu} = \left(\frac{\partial u}{\partial z}\right)^{-2} \delta g_{zz} = z^4 \delta g_{zz} \ . \tag{6.1.4b}$$

Any holomorphic quadratic differential δg_{zz} would have to be holomorphic in z but vanish as z^{-4} at infinity, and so must vanish identically. A CKV δz, on the other hand, is holomorphic at $u = 0$ provided it grows no more rapidly than z^2 as $z \to \infty$. The general CKV is then

$$\delta z = a_0 + a_1 z + a_2 z^2 \ , \tag{6.1.5a}$$

$$\delta \bar{z} = a_0^* + a_1^* \bar{z} + a_2^* \bar{z}^2 \ , \tag{6.1.5b}$$

with three complex or six real parameters as expected from the Riemann–Roch theorem.

These infinitesimal transformations exponentiate to give the *Möbius*

group

$$z' = \frac{\alpha z + \beta}{\gamma z + \delta} \tag{6.1.6}$$

for complex $\alpha, \beta, \gamma, \delta$. Rescaling $\alpha, \beta, \gamma, \delta$ leaves the transformation unchanged, so we can fix $\alpha\delta - \beta\gamma = 1$ and identify under an overall sign reversal of $\alpha, \beta, \gamma, \delta$. This defines the group $PSL(2, \mathbf{C})$. This is the most general coordinate transformation that is holomorphic on all of S_2. It is one-to-one with the point at infinity included. Three of the six parameters correspond to ordinary rotations, forming an $SO(3)$ subgroup of $PSL(2, \mathbf{C})$.

The disk

It is useful to construct the disk D_2 from the sphere by identifying points under a reflection. For example, identify points z and z' such that

$$z' = 1/\bar{z} \ . \tag{6.1.7}$$

In polar coordinates $z = re^{i\phi}$, this inverts the radius and leaves the angle fixed, so the unit disk $|z| \leq 1$ is a fundamental region for the identification. The points on the unit circle are fixed by the reflection, so this becomes a boundary. It is often more convenient to use the conformally equivalent reflection

$$z' = \bar{z} \ . \tag{6.1.8}$$

The upper half-plane is now a fundamental region, and the real axis is the boundary.

The CKG of the disk is the subgroup of $PSL(2, \mathbf{C})$ that leaves the boundary of the disk fixed. For the reflection (6.1.8) this is just the subgroup of (6.1.6) with $\alpha, \beta, \gamma, \delta$ all real, which is $PSL(2, \mathbf{R})$, the Möbius group with real parameters. One CKV is the ordinary rotational symmetry of the disk. Again, all metrics are equivalent — there are no moduli.

The projective plane

The projective plane RP_2 can also be obtained as a $\mathbf{Z}_2$ identification of the sphere. Identify points z and z' with

$$z' = -1/\bar{z} \ . \tag{6.1.9}$$

These points are diametrically opposite in the round metric (6.1.3). There are no fixed points and so no boundary in the resulting space, but the space is not oriented. One fundamental region for the identification is the unit disk $|z| \leq 1$, with points $e^{i\phi}$ and $-e^{i\phi}$ identified. Another choice is the upper half-z-plane. There are no moduli. The CKG is the subgroup of $PSL(2, \mathbf{C})$ that respects the identification (6.1.9); this is just the ordinary rotation group $SO(3)$.

Both the disk and projective plane have been represented as the sphere with points identified under a $\mathbf{Z}_2$ transformation, or *involution*. In fact, every world-sheet can be obtained from a closed oriented world-sheet by identifying under one or two $\mathbf{Z}_2$s. The method of images can then be used to obtain the Green's functions.

6.2 Scalar expectation values

The basic quantities that we need are the expectation values of products of vertex operators. In order to develop a number of useful techniques and points of view we will calculate these in three different ways: by direct path integral evaluation, by using holomorphicity properties, and later in the chapter by operator methods.

The path integral method has already been used in section 5.3 for the Faddeev–Popov determinant and in the appendix for the harmonic oscillator. Start with a generating functional

$$Z[J] = \left\langle \exp\left(i \int d^2\sigma \, J(\sigma) \cdot X(\sigma) \right) \right\rangle \tag{6.2.1}$$

for arbitrary $J_\mu(\sigma)$. For now we work on an arbitrary compact two-dimensional surface M, and in an arbitrary spacetime dimension d. Expand $X^\mu(\sigma)$ in terms of a complete set $\mathsf{X}_I(\sigma)$,

$$X^\mu(\sigma) = \sum_I x_I^\mu \mathsf{X}_I(\sigma) \ , \tag{6.2.2a}$$

$$\nabla^2 \mathsf{X}_I = -\omega_I^2 \mathsf{X}_I \ , \tag{6.2.2b}$$

$$\int_M d^2\sigma g^{1/2} \mathsf{X}_I \mathsf{X}_{I'} = \delta_{II'} \ . \tag{6.2.2c}$$

Then

$$Z[J] = \prod_{I,\mu} \int dx_I^\mu \exp\left(-\frac{\omega_I^2 x_I^\mu x_{I\mu}}{4\pi\alpha'} + i x_I^\mu J_{I\mu} \right) , \tag{6.2.3}$$

where

$$J_I^\mu = \int d^2\sigma J^\mu(\sigma) \mathsf{X}_I(\sigma) \ . \tag{6.2.4}$$

The integrals are Gaussian except for the constant mode

$$\mathsf{X}_0 = \left(\int d^2\sigma g^{1/2} \right)^{-1/2} , \tag{6.2.5}$$

which has vanishing action and so gives a delta function. Carrying out the integrations leaves

$$Z[J] = i(2\pi)^d\delta^d(J_0)\prod_{I\neq 0}\left(\frac{4\pi^2\alpha'}{\omega_I^2}\right)^{d/2}\exp\left(-\frac{\pi\alpha' J_I\cdot J_I}{\omega_I^2}\right)$$
$$= i(2\pi)^d\delta^d(J_0)\left(\det{}'\frac{-\nabla^2}{4\pi^2\alpha'}\right)^{-d/2}$$
$$\times\exp\left(-\frac{1}{2}\int d^2\sigma\, d^2\sigma'\, J(\sigma)\cdot J(\sigma')G'(\sigma,\sigma')\right) . \quad (6.2.6)$$

As discussed in section 2.1 and further in section 3.2, the timelike modes x_I^0 give rise to wrong-sign Gaussians and are defined by the contour rotation[1] $x_I^0 \to -ix_I^d$, $I\neq 0$. The primed Green's function excludes the zero mode contribution,

$$G'(\sigma_1,\sigma_2) = \sum_{I\neq 0}\frac{2\pi\alpha'}{\omega_I^2}\mathsf{X}_I(\sigma_1)\mathsf{X}_I(\sigma_2) . \quad (6.2.7)$$

It satisfies the differential equation

$$-\frac{1}{2\pi\alpha'}\nabla^2 G'(\sigma_1,\sigma_2) = \sum_{I\neq 0}\mathsf{X}_I(\sigma_1)\mathsf{X}_I(\sigma_2)$$
$$= g^{-1/2}\delta^2(\sigma_1-\sigma_2) - \mathsf{X}_0^2 , \quad (6.2.8)$$

where the completeness of the X_I has been used. The ordinary Green's function with a delta function source does not exist. It would correspond to the electrostatic potential of a single charge, but on a compact surface the field lines from the source have no place to go. The X_0^2 term can be thought of as a neutralizing background charge distribution.

The sphere

Specializing to the sphere, the solution to the differential equation (6.2.8) is

$$G'(\sigma_1,\sigma_2) = -\frac{\alpha'}{2}\ln|z_{12}|^2 + f(z_1,\bar z_1) + f(z_2,\bar z_2) , \quad (6.2.9)$$

[1] A curious factor of i has been inserted into eq. (6.2.6) by hand because it is needed in the S-matrix, but it can be understood formally as arising from the same rotation. If we rotated the entire field $X^0 \to -iX^d$, the Jacobian should be 1, by the usual argument about rescaling fields (footnote 1 of the appendix). However, we do not want to rotate x_0^0, because this mode produces the energy delta function. So we have to rotate it back, giving a factor of i.

where

$$f(z,\bar z) = \frac{\alpha'\mathsf{X}_0^2}{4}\int d^2z'\,\exp[2\omega(z',\bar z')]\ln|z-z'|^2 + k . \quad (6.2.10)$$

The constant k is determined by the property that G' is orthogonal to X_0, but in any case we will see that the function f drops out of all expectation values. It comes from the background charge, but the delta function from the zero-mode integration forces overall neutrality, $J_0^\mu = 0$, and the background makes no net contribution.

Now consider the path integral with a product of tachyon vertex operators,

$$A_{S_2}^n(k,\sigma) = \left\langle \left[e^{ik_1\cdot X(\sigma_1)}\right]_{\mathrm{r}}\left[e^{ik_2\cdot X(\sigma_2)}\right]_{\mathrm{r}}\cdots\left[e^{ik_n\cdot X(\sigma_n)}\right]_{\mathrm{r}}\right\rangle_{S_2} . \quad (6.2.11)$$

This corresponds to

$$J(\sigma) = \sum_{i=1}^n k_i\delta^2(\sigma-\sigma_i) . \quad (6.2.12)$$

The amplitude (6.2.6) then becomes

$$A_{S_2}^n(k,\sigma) = iC_{S_2}^X(2\pi)^d\delta^d(\textstyle\sum_i k_i)$$
$$\times\exp\left(-\sum_{\substack{i,j=1\\ i<j}}^n k_i\cdot k_j G'(\sigma_i,\sigma_j) - \frac{1}{2}\sum_{i=1}^n k_i^2 G'_{\mathrm{r}}(\sigma_i,\sigma_i)\right) . \quad (6.2.13)$$

The constant here is

$$C_{S_2}^X = \mathsf{X}_0^{-d}\left(\det{}'\frac{-\nabla^2}{4\pi^2\alpha'}\right)_{S_2}^{-d/2} . \quad (6.2.14)$$

The determinant can be regularized and computed, but we will not need to do this explicitly. We are employing here the simple renormalization used in section 3.6, so that the self-contractions involve

$$G'_{\mathrm{r}}(\sigma,\sigma') = G'(\sigma,\sigma') + \frac{\alpha'}{2}\ln d^2(\sigma,\sigma') . \quad (6.2.15)$$

Note that

$$G'_{\mathrm{r}}(\sigma,\sigma) = 2f(z,\bar z) + \alpha'\omega(z,\bar z) \quad (6.2.16)$$

is finite by design. The path integral on the sphere is then

$$A_{S_2}^n(k,\sigma) = iC_{S_2}^X(2\pi)^d\delta^d(\textstyle\sum_i k_i)\exp\left(-\frac{\alpha'}{2}\sum_i k_i^2\omega(\sigma_i)\right)\prod_{\substack{i,j=1\\ i<j}}^n |z_{ij}|^{\alpha' k_i\cdot k_j} . \quad (6.2.17)$$

The function f has dropped out as promised. The dependence on the conformal factor $\omega(\sigma)$ is precisely that found in section 3.6 from the Weyl

anomaly in the vertex operator. It will cancel the variation of $g^{1/2}$ for an on-shell operator.[2] We can take a metric that is flat in a large region containing all the vertex operators ('pushing the curvature to infinity') and the term in $\omega(\sigma_i)$ drops out.

Higher vertex operators are exponentials times derivatives of X^μ, so we also need

$$\left\langle \prod_{i=1}^{n} \left[e^{ik_i \cdot X(z_i,\bar{z}_i)} \right]_{\mathrm{r}} \prod_{j=1}^{p} \partial X^{\mu_j}(z'_j) \prod_{k=1}^{q} \bar{\partial} X^{\nu_k}(\bar{z}''_k) \right\rangle_{S_2} . \tag{6.2.18}$$

This is given by summing over all contractions, where every ∂X or $\bar{\partial} X$ must be contracted either with an exponential or with another ∂X or $\bar{\partial} X$. The XX contraction is simply $-\frac{1}{2}\alpha' \ln|z|^2$, the fs again dropping out in the final expression. The result can be summarized as

$$iC^X_{S_2}(2\pi)^d \delta^d(\textstyle\sum_i k_i) \exp\left(-\frac{\alpha'}{2} \sum_i k_i^2 \omega(\sigma_i) \right) \prod_{\substack{i,j=1 \\ i<j}}^{n} |z_{ij}|^{\alpha' k_i \cdot k_j}$$

$$\times \left\langle \prod_{j=1}^{p} \left[v^{\mu_j}(z'_j) + q^{\mu_j}(z'_j) \right] \prod_{k=1}^{q} \left[\tilde{v}^{\nu_k}(\bar{z}''_k) + \tilde{q}^{\nu_k}(\bar{z}''_k) \right] \right\rangle \tag{6.2.19}$$

Here

$$v^\mu(z) = -i\frac{\alpha'}{2} \sum_{i=1}^{n} \frac{k_i^\mu}{z - z_i} , \quad \tilde{v}^\mu(\bar{z}) = -i\frac{\alpha'}{2} \sum_{i=1}^{n} \frac{k_i^\mu}{\bar{z} - \bar{z}_i} \tag{6.2.20}$$

come from the contractions with the exponentials. The expectation values of $q^\mu = \partial X^\mu - v^\mu$ are given by the sum over all contractions using $-\eta^{\mu\nu}(z - z')^{-2}\alpha'/2$, and those of $\tilde{q}^\nu$ by the conjugate.

Now we repeat the calculation in a different way, using holomorphicity. As an example consider

$$\langle \partial X^\mu(z_1) \partial X^\nu(z_2) \rangle_{S_2} . \tag{6.2.21}$$

The OPE determines this to be

$$-\frac{\alpha' \eta^{\mu\nu}}{2z_{12}^2} \langle 1 \rangle_{S_2} + g(z_1, z_2) , \tag{6.2.22}$$

where $g(z_1, z_2)$ is holomorphic in both variables. In the u-patch,

$$\partial_u X^\mu = -z^2 \partial_z X^\mu , \tag{6.2.23}$$

so the condition of holomorphicity at $u = 0$ is that expectation values of $\partial_z X^\mu$ fall as z^{-2} at infinity. More generally, a tensor of weight $(h, 0)$ must

[2] The constant $C^X_{S_2}$ also depends on the conformal factor, because the X^μ CFT by itself has nonzero central charge, but in the full string theory this will cancel.

behave as z^{-2h} at infinity. Focusing on the z_1-dependence in eq. (6.2.22) at fixed z_2, this implies that $g(z_1, z_2)$ falls as z_1^{-2} at infinity and so vanishes by holomorphicity. This determines the expectation value up to normalization. Comparing with the path integral result (6.2.19), there is agreement, though the normalization $\langle 1 \rangle_{S_2}$ is seen to diverge as $\delta(0)$. This is just the zero-mode divergence from the infinite volume of spacetime.

To obtain the expectation value of the product of vertex operators

$$A^n_{S_2}(k, \sigma) = \left\langle \prod_{i=1}^{n} :e^{ik_i \cdot X(z_i,\bar{z}_i)}: \right\rangle_{S_2} \tag{6.2.24}$$

by this method is less direct, because the exponentials are not holomorphic. In fact they can be factored into holomorphic and antiholomorphic parts, but this is subtle and is best introduced in operator language as we will do in chapter 8. We use the holomorphicity of the translation current. Consider the expectation value with one current $\partial X^\mu(z)$ added. The OPE of the current with the vertex operators determines the singularities in z,

$$\left\langle \partial X^\mu(z) \prod_{i=1}^{n} :e^{ik_i \cdot X(z_i,\bar{z}_i)}: \right\rangle_{S_2} = -\frac{i\alpha'}{2} A^n_{S_2}(k, \sigma) \sum_{i=1}^{n} \frac{k_i^\mu}{z - z_i}$$
$$+ \text{terms holomorphic in } z. \tag{6.2.25}$$

Now look at $z \to \infty$. The condition that $\partial_u X^\mu$ be holomorphic at $u = 0$ again requires that the expectation value vanish as z^{-2} when $z \to \infty$. The holomorphic term in eq. (6.2.25) must then vanish, and from the vanishing of the order z^{-1} term we recover momentum conservation,

$$A^n_{S_2}(k, \sigma) \sum_{i=1}^{n} k_i^\mu = 0. \tag{6.2.26}$$

Let us also say this in a slightly different way. Consider the contour integral of the spacetime translation current

$$p^\mu = \frac{1}{2\pi i} \oint_C (dz\, j_z^\mu - d\bar{z}\, j_{\bar{z}}^\mu) , \tag{6.2.27}$$

where the contour C encircles all the exponential operators. There are two ways to evaluate this. The first is to contract C until it is a small circle around each vertex operator, which picks out the $(z - z_i)^{-1}$ term from each OPE and gives $\sum_{i=1}^n k_i^\mu$. The second is to expand it until it is a small circle in the u-patch: by holomorphicity at $u = 0$ it must vanish. This sort of 'contour-pulling' argument is commonly used in CFT.

We have used the OPE to determine the singularities and then used holomorphicity to get the full z-dependence. Now we look at the second term of the OPE as $z \to z_1$. Expanding the result (6.2.25) gives

$$-\frac{i\alpha'}{2}A^n_{S_2}(k,\sigma)\left(\frac{k_1^\mu}{z-z_1}+\sum_{i=2}^n\frac{k_i^\mu}{z_1-z_i}+O(z-z_1)\right). \tag{6.2.28}$$

Contracting this with ik_1^μ, the term of order $(z-z_1)^0$ must agree with the OPE

$$ik_1\cdot\partial X(z)\ :e^{ik_1\cdot X(z_1,\bar z_1)}: = \frac{\alpha' k_1^2}{2(z-z_1)}\ :e^{ik_1\cdot X(z_1,\bar z_1)}: +\partial_{z_1}\ :e^{ik_1\cdot X(z_1,\bar z_1)}: +O(z-z_1)\,. \tag{6.2.29}$$

This implies

$$\partial_{z_1}A^n_{S_2}(k,\sigma)=\frac{\alpha'}{2}A^n_{S_2}(k,\sigma)\sum_{i=2}^n\frac{k_1\cdot k_i}{z_{1i}}\,. \tag{6.2.30}$$

Integrating, and using the conjugate equation and momentum conservation, determines the path integral up to normalization,

$$A^n_{S_2}(k,\sigma)\propto\delta^d(\textstyle\sum_i k_i)\displaystyle\prod_{\substack{i,j=1\\ i<j}}^n|z_{ij}|^{\alpha' k_i\cdot k_j}, \tag{6.2.31}$$

in agreement with the first method. Note that to compare we must push the curvature to infinity ($\omega(\sigma_i)=0$) so that $[\ ]_r = :\ :$.

It is worth noting that the intermediate steps in the path integral method depend in a detailed way on the particular choice of Riemannian metric. The metric is needed in order to preserve coordinate invariance in these steps. A different, Weyl-equivalent, metric gives a different Laplacian and different eigenfunctions, though at the very end this dependence must drop out in string theory. The second method uses only the basic data of a Riemann surface, its holomorphic structure.

The disk

The generalization to the disk is straightforward. We will represent the disk by taking the above representation of the sphere and restricting z to the upper half of the complex plane. The Neumann boundary term is accounted for by an image charge,

$$G^r(\sigma_1,\sigma_2)=-\frac{\alpha'}{2}\ln|z_1-z_2|^2-\frac{\alpha'}{2}\ln|z_1-\bar z_2|^2\,, \tag{6.2.32}$$

up to terms that drop out due to momentum conservation. Then

$$\left\langle\prod_{i=1}^n:e^{ik_i\cdot X(z_i,\bar z_i)}:\right\rangle_{D_2} = iC^X_{D_2}(2\pi)^d\delta^d(\textstyle\sum_i k_i)\displaystyle\prod_{i=1}^n|z_i-\bar z_i|^{\alpha' k_i^2/2} \times\prod_{\substack{i,j=1\\ i<j}}^n|z_i-z_j|^{\alpha' k_i\cdot k_j}|z_i-\bar z_j|^{\alpha' k_i\cdot k_j}\,. \tag{6.2.33}$$

For expectation values with $\partial_a X^\mu$s one again sums over contractions, now using the Green's function (6.2.32). Note that $\partial X^\mu(z)\bar\partial X^\nu(\bar z')$ (or $q^\mu(z)\tilde q^\nu(\bar z')$) has a nonzero contraction.

For two points on the boundary, the two terms in the Green's function (6.2.32) are equal, and the Green's function diverges at zero separation even after normal ordering subtracts the first term. For this reason, boundary operators must be defined with *boundary normal ordering*, where the subtraction is doubled:

$$\stackrel{\star}{\ }X^\mu(y_1)X^\nu(y_2)\stackrel{\star}{\ }=X^\mu(y_1)X^\nu(y_2)+2\alpha'\ln|y_1-y_2|\,, \tag{6.2.34}$$

with y denoting a coordinate on the real axis. The combinatorics are the same as for other forms of normal ordering. Boundary normal-ordered expectation values of boundary operators have the same good properties (nonsingularity) as conformal normal-ordered operators in the interior.

An expectation value with exponentials both in the interior and on the boundary, each having the appropriate normal ordering, is given by taking the appropriate limit of the interior result (6.2.33) and dropping the $|z_i-\bar z_i|^{\alpha' k_i^2/2}$ factor for the boundary operators. Explicitly, for exponentials all on the boundary,

$$\left\langle\prod_{i=1}^n\stackrel{\star}{\ }e^{ik_i\cdot X(y_i)}\stackrel{\star}{\ }\right\rangle_{D_2} = iC^X_{D_2}(2\pi)^d\delta^d(\textstyle\sum_i k_i)\displaystyle\prod_{\substack{i,j=1\\ i<j}}^n|y_i-y_j|^{2\alpha' k_i\cdot k_j}\,. \tag{6.2.35}$$

More generally,

$$\left\langle\prod_{i=1}^n\stackrel{\star}{\ }e^{ik_i\cdot X(y_i)}\stackrel{\star}{\ }\prod_{j=1}^p\partial_y X^{\mu_j}(y_j')\right\rangle_{D_2} = iC^X_{D_2}(2\pi)^d\delta^d(\textstyle\sum_i k_i) \times\displaystyle\prod_{\substack{i,j=1\\ i<j}}^n|y_{ij}|^{2\alpha' k_i\cdot k_j}\left\langle\prod_{j=1}^p\left[v^{\mu_j}(y_j')+q^{\mu_j}(y_j')\right]\right\rangle_{D_2}\,, \tag{6.2.36}$$

where now

$$v^\mu(y)=-2i\alpha'\sum_{i=1}^n\frac{k_i^\mu}{y-y_i} \tag{6.2.37}$$

and the qs are contracted using $-2\alpha'(y-y')^{-2}\eta^{\mu\nu}$.

The projective plane

The method of images gives the Green's function as

$$G'(\sigma_1,\sigma_2) = -\frac{\alpha'}{2}\ln|z_1 - z_2|^2 - \frac{\alpha'}{2}\ln|1 + z_1\bar{z}_2|^2 \,. \tag{6.2.38}$$

Now

$$\left\langle \prod_{i=1}^{n} :e^{ik_i\cdot X(z_i,\bar{z}_i)}: \right\rangle_{D_2} = iC_{D_2}^{X}(2\pi)^d\delta^d(\textstyle\sum_i k_i)\displaystyle\prod_{i=1}^{n}|1+z_i\bar{z}_i|^{\alpha' k_i^2/2} \times \prod_{\substack{i,j=1\\ i<j}}^{n}|z_i - z_j|^{\alpha' k_i\cdot k_j}|1+z_i\bar{z}_j|^{\alpha' k_i\cdot k_j} \,, \tag{6.2.39}$$

and so on for more general expectation values. There is no boundary because the involution has no fixed points — there is no way for a point z to approach its image $-\bar{z}^{-1}$.

6.3 The *bc* CFT

The sphere

The path integral for the ghosts has already been set up in section 5.3. According to the Riemann–Roch theorem, the simplest nonvanishing expectation value will be

$$\langle\, c(z_1)c(z_2)c(z_3)\tilde{c}(\bar{z}_4)\tilde{c}(\bar{z}_5)\tilde{c}(\bar{z}_6)\,\rangle_{S_2} \,. \tag{6.3.1}$$

Up to normalization (a functional determinant), the result (5.3.18) was just the zero-mode determinant,

$$\det \mathsf{C}^a_{0j}(\sigma_i) \,. \tag{6.3.2}$$

The six CKVs were found in section 6.1. In a complex basis they are

$$\mathsf{C}^z = 1, z, z^2 \,, \quad \mathsf{C}^{\bar{z}} = 0 \,, \tag{6.3.3a}$$

$$\mathsf{C}^z = 0 \,, \quad \mathsf{C}^{\bar{z}} = 1, \bar{z}, \bar{z}^2 \,. \tag{6.3.3b}$$

In this basis the determinant splits into two 3×3 blocks, and the expectation value becomes

$$C^{\mathrm{g}}_{S_2}\det\begin{vmatrix} 1 & 1 & 1 \\ z_1 & z_2 & z_3 \\ z_1^2 & z_2^2 & z_3^2 \end{vmatrix}\det\begin{vmatrix} 1 & 1 & 1 \\ \bar{z}_4 & \bar{z}_5 & \bar{z}_6 \\ \bar{z}_4^2 & \bar{z}_5^2 & \bar{z}_6^2 \end{vmatrix} = C^{\mathrm{g}}_{S_2} z_{12}z_{13}z_{23}\bar{z}_{45}\bar{z}_{46}\bar{z}_{56} \,. \tag{6.3.4}$$

The constant $C^{\mathrm{g}}_{S_2}$ includes the functional determinant and also a finite-dimensional Jacobian (independent of the positions) that arises because the basis (6.3.3) is not orthonormal.

This is the only *bc* path integral that we will actually need for the tree-level amplitudes, but for completeness we give

$$\left\langle \prod_{i=1}^{p+3} c(z_i)\prod_{j=1}^{p} b(z'_j)\cdot(\text{anti}) \right\rangle_{S_2} = C^{\mathrm{g}}_{S_2}\frac{z_{p+1,p+2}z_{p+1,p+3}z_{p+2,p+3}}{(z_1 - z'_1)\ldots(z_p - z'_p)}\cdot(\text{anti}) \pm \text{permutations} \,, \tag{6.3.5}$$

obtained by contracting the bs with cs. The antiholomorphic part 'anti' has the same form. We are being a little careless with the *overall* sign; again, for the Faddeev–Popov determinant we will in any case take an absolute value.

To give an alternative derivation using holomorphicity, we again consider first the conservation law, derived by inserting into the amplitude (6.3.5) the ghost number contour integral $\oint_C dz\, j(z)/2\pi i$, where C encircles all the vertex operators and $j = -:bc:$ is again the ghost number current. From the OPE with b and c, this just counts the number of c minus the number of b fields, giving $n_c - n_b$ times the original amplitude. Now pull the contour into the u-patch using the conformal transformation (2.5.17),

$$\oint_C \frac{dz}{2\pi i}j_z = \oint_C \frac{du}{2\pi i}j_u + 3 \;\to\; 3 \,. \tag{6.3.6}$$

The shift $+3$ comes about because j is not a tensor, having a z^{-3} term in its OPE with T. In the last step we use holomorphicity in the u-patch. The nonzero amplitudes therefore have $n_c - n_b = 3$ and similarly $n_{\tilde{c}} - n_{\tilde{b}} = 3$, in agreement with the Riemann–Roch theorem.

From the OPE, the ghost expectation value (6.3.1) is holomorphic in each variable, and it must have a zero when two identical anticommuting fields come together. Thus it must be of the form

$$z_{12}z_{13}z_{23}\bar{z}_{45}\bar{z}_{46}\bar{z}_{56}F(z_1,z_2,z_3)\tilde{F}(\bar{z}_4,\bar{z}_5,\bar{z}_6) \,, \tag{6.3.7}$$

with F holomorphic and $\tilde{F}$ antiholomorphic functions of the positions. As $z_1 \to \infty$ this goes as $z_1^2 F$. However, c is a tensor of weight -1, so the amplitude can be no larger than z_1^2 at infinity. Thus $F(z_i)$ must be independent of z_1, and so also of z_2 and z_3; arguing similarly for $\tilde{F}$ we obtain again the result (6.3.4). The same argument gives for the general case (6.3.5) the result

$$C^{\mathrm{g}}_{S_2}\prod_{\substack{i,i'=1\\ i<i'}}^{p+3} z_{ii'}\prod_{\substack{j,j'=1\\ j<j'}}^{p} z'_{jj'}\prod_{i=1}^{p+3}\prod_{j=1}^{p}(z_i - z'_j)^{-1}\cdot(\text{anti}) \,, \tag{6.3.8}$$

which has the correct poles, zeros, and behavior at infinity. The permutations (6.3.5) evidently sum up to give this single term.

We have considered the $\lambda=2$ theory that is relevant to the ghosts of the bosonic string, but either method generalizes readily to any λ.

The disk

The simplest way to obtain the bc amplitudes on the disk is with the doubling trick. As in eq. (2.7.30), we can represent both the holomorphic and antiholomorphic fields in the upper half-plane by holomorphic fields in the whole plane, using

$$\tilde{b}(\bar{z}) = b(z') \,, \quad \tilde{c}(\bar{z}) = c(z') \,, \quad z' = \bar{z} \,, \ \mathrm{Im}\, z > 0 \,. \tag{6.3.9}$$

The expectation value of the holomorphic fields follows as on the sphere,

$$\langle c(z_1)c(z_2)c(z_3)\rangle_{D_2} = C^{\mathrm{g}}_{D_2} z_{12}z_{13}z_{23} \tag{6.3.10}$$

for all z. Then for example

$$\begin{aligned}\langle c(z_1)c(z_2)\tilde{c}(\bar{z}_3)\rangle_{D_2} &= \langle c(z_1)c(z_2)c(z_3')\rangle_{D_2} \\ &= C^{\mathrm{g}}_{D_2} z_{12}(z_1-\bar{z}_3)(z_2-\bar{z}_3) \,.\end{aligned} \tag{6.3.11}$$

More general correlators are obtained in the same way.

The projective plane

The doubling trick again can be used. The involution $z' = -\bar{z}^{-1}$ implies that

$$\tilde{b}(\bar{z}) = \left(\frac{\partial z'}{\partial \bar{z}}\right)^2 b(z') = z'^4 b(z') \,, \quad \tilde{c}(\bar{z}) = \left(\frac{\partial z'}{\partial \bar{z}}\right)^{-1} c(z') = z'^{-2}c(z') \,. \tag{6.3.12}$$

Again

$$\langle c(z_1)c(z_2)c(z_3)\rangle_{RP_2} = C^{\mathrm{g}}_{RP_2} z_{12}z_{13}z_{23} \tag{6.3.13}$$

and so

$$\begin{aligned}\langle c(z_1)c(z_2)\tilde{c}(\bar{z}_3)\rangle_{RP_2} &= z_3'^{-2}\langle c(z_1)c(z_2)c(z_3')\rangle_{RP_2} \\ &= C^{\mathrm{g}}_{RP_2} z_{12}(1+z_1\bar{z}_3)(1+z_2\bar{z}_3) \,.\end{aligned} \tag{6.3.14}$$

6.4 The Veneziano amplitude

Open string amplitudes are slightly simpler than closed string amplitudes, so we begin with these.

We represent the disk as the upper half-plane so the boundary coordinate y is real. There are no moduli. After fixing the metric, the CKG $PSL(2,\mathbf{R})$ can be used to fix the three vertex operators to arbitrary positions y_1, y_2, y_3 on the boundary, except that this group does not change the cyclic ordering of the vertex operators so we must sum over the two orderings. For three open string tachyons on the disk, the general expression (5.3.9) for the string S-matrix thus reduces to

$$\begin{aligned}S_{D_2}(k_1;k_2;k_3) = g_{\mathrm{o}}^3 e^{-\lambda} \left\langle :c^1 e^{ik_1\cdot X}(y_1): :c^1 e^{ik_2\cdot X}(y_2): :c^1 e^{ik_3\cdot X}(y_3):\right\rangle_{D_2} \\ + (k_2 \leftrightarrow k_3) \,,\end{aligned} \tag{6.4.1}$$

each fixed coordinate integration being replaced by the corresponding c-ghost. Each vertex operator includes a factor of g_{o}, the open string coupling. The factor $e^{-\lambda}$ is from the Euler number term in the action. Of course g_{o} and e^{λ} are related, $g_{\mathrm{o}}^2 \propto e^{\lambda}$, but we will determine the constant of proportionality as we go along.

The expectation values in (6.4.1) were found in the previous sections (again, we take the absolute value for the ghosts), giving

$$\begin{aligned}S_{D_2}(k_1;k_2;k_3) = 2ig_{\mathrm{o}}^3 C_{D_2}(2\pi)^{26}\delta^{26}(\textstyle\sum_i k_i) \\ \times |y_{12}|^{1+2\alpha' k_1\cdot k_2}|y_{13}|^{1+2\alpha' k_1\cdot k_3}|y_{23}|^{1+2\alpha' k_2\cdot k_3} \,,\end{aligned} \tag{6.4.2}$$

where $C_{D_2} = e^{-\lambda}C^X_{D_2}C^{\mathrm{g}}_{D_2}$. Momentum conservation and the mass-shell condition $k_i^2 = 1/\alpha'$ imply that

$$2\alpha' k_1\cdot k_2 = \alpha'(k_3^2 - k_1^2 - k_2^2) = -1 \tag{6.4.3}$$

and the same for the other $k_i \cdot k_j$, so this reduces to

$$S_{D_2}(k_1;k_2;k_3) = 2ig_{\mathrm{o}}^3 C_{D_2}(2\pi)^{26}\delta^{26}(\textstyle\sum_i k_i) \,. \tag{6.4.4}$$

This is independent of the gauge choice y_i, which is of course a general property of the Faddeev–Popov procedure. The Weyl invariance is crucial here — if the vertex operators were taken off the mass shell, the amplitude would depend on the choice of y_i.

We could have used independence of the y_i to determine the Faddeev–Popov determinant without calculation. Using the mass-shell condition and momentum conservation, the X^μ expectation value is proportional to $|y_{12}y_{13}y_{23}|^{-1}$, and so the measure must be reciprocal to this. This same measure then applies for $n > 3$ vertex operators, because in all cases three positions are fixed.

The four-tachyon amplitude is obtained in the same way,

$$\begin{aligned}&S_{D_2}(k_1;k_2;k_3;k_4) \\ &= g_{\mathrm{o}}^4 e^{-\lambda}\int_{-\infty}^{\infty} dy_4 \left\langle \prod_{i=1}^{3} :c^1(y_i)e^{ik_i\cdot X(y_i)}: :e^{ik_4\cdot X(y_4)}:\right\rangle + (k_2 \leftrightarrow k_3) \\ &= ig_{\mathrm{o}}^4 C_{D_2}(2\pi)^{26}\delta^{26}(\textstyle\sum_i k_i)|y_{12}y_{13}y_{23}|\displaystyle\int_{-\infty}^{\infty} dy_4 \prod_{i<j}|y_{ij}|^{2\alpha' k_i\cdot k_j} \\ &\qquad + (k_2 \leftrightarrow k_3) \,.\end{aligned} \tag{6.4.5}$$

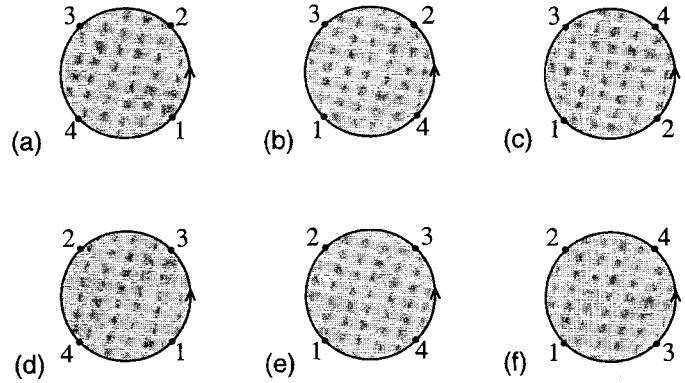


Fig. 6.2. The six cyclically inequivalent orderings of four open string vertex operators on the disk. The coordinate y increases in the direction of the arrow, except that at point 3 it jumps from $+\infty$ to $-\infty$.

This is again independent of $y_{1,2,3}$ after a change of variables (a Möbius transformation) on y_4. It is convenient to take $y_1 = 0$, $y_2 = 1$, and $y_3 \to \infty$. The amplitude is conventionally written in terms of the *Mandelstam variables*

$$s = -(k_1 + k_2)^2 \,, \quad t = -(k_1 + k_3)^2 \,, \quad u = -(k_1 + k_4)^2 \,. \tag{6.4.6}$$

These are not independent: momentum conservation and the mass-shell condition imply that

$$s + t + u = \sum_i m_i^2 = -\frac{4}{\alpha'} \,. \tag{6.4.7}$$

Using $2\alpha' k_i \cdot k_j = -2 + \alpha'(k_i + k_j)^2$, the amplitude becomes

$$S_{D_2}(k_1;k_2;k_3;k_4) = ig_o^4 C_{D_2}(2\pi)^{26}\delta^{26}(\textstyle\sum_i k_i) \times \left[\int_{-\infty}^{\infty} dy_4\, |y_4|^{-\alpha' u - 2} |1 - y_4|^{-\alpha' t - 2} \;+\; (t \to s)\right] . \tag{6.4.8}$$

The integral splits into three ranges, $-\infty < y_4 < 0$, $0 < y_4 < 1$, and $1 < y_4 < \infty$. For these three ranges the vertex operators are ordered as in figures 6.2(a), (b), and (c) respectively. Möbius invariance can be used to take each of these ranges into any other, so they give contributions that are equal up to interchange of vertex operators. The $(t \leftrightarrow s)$ term gives figures 6.2(d), (e), and (f). In all,

$$S_{D_2}(k_1;k_2;k_3;k_4) = 2ig_o^4 C_{D_2}(2\pi)^{26}\delta^{26}(\textstyle\sum_i k_i)\big[I(s,t) + I(t,u) + I(u,s)\big] \,, \tag{6.4.9}$$

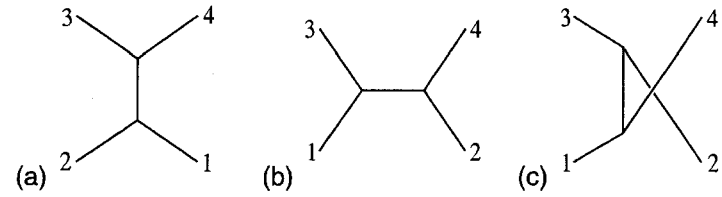


Fig. 6.3. Processes giving poles in the (a) s-, (b) t-, and (c) u-channels.

where

$$I(s,t) = \int_0^1 dy\, y^{-\alpha' s - 2}(1-y)^{-\alpha' t - 2} \,. \tag{6.4.10}$$

The three terms come from figures 6.2(c), (f), 6.2(b), (d), and 6.2(a), (e) respectively.

The integral $I(s,t)$ converges if $\alpha' s < -1$ and $\alpha' t < -1$. As $\alpha' s \to -1$, the integral diverges at $y = 0$. To study the divergence, take a neighborhood of $y = 0$ and approximate the integrand:

$$\begin{aligned} I(s,t) &= \int_0^r y^{-\alpha' s - 2} \;+\; \text{terms analytic at } \alpha' s = -1 \\ &= -\frac{r^{-\alpha' s - 1}}{\alpha' s + 1} \;+\; \text{terms analytic at } \alpha' s = -1 \\ &= -\frac{1}{\alpha' s + 1} \;+\; \text{terms analytic at } \alpha' s = -1 \,. \end{aligned} \tag{6.4.11}$$

In (6.4.11), we have evaluated the integral in the convergent region. We see that the divergence is a pole at $s = -1/\alpha'$, the mass-squared of the open string tachyon. The variable s is just the square of the center-of-mass energy for scattering $1 + 2 \to 3 + 4$, so this pole is a resonance due to an intermediate tachyon state. Again, it is an artifact of the bosonic string that this lightest string state is tachyonic, and not relevant to the discussion. The pole is due to the process shown in figure 6.3(a), in which tachyons 1 and 2 join to become a single tachyon, which then splits into tachyons 3 and 4.

Because the singularity at $\alpha' s = -1$ is just a pole, $I(s,t)$ can be analytically continued past this point into the region $\alpha' s > -1$. The amplitude is defined via this analytic continuation. The divergence of the amplitude *at* the pole is an essential physical feature of the amplitude, a resonance corresponding to propagation of the intermediate string state over long spacetime distances. The divergence of the integral *past* the pole is not; it is just an artifact of this particular integral representation of the amplitude. The continuation poses no problem. In fact, we will see that every

string divergence is of this same basic form, so this one kind of analytic continuation removes all divergences — except of course for the poles themselves. The pole is on the real axis, so we need to define it more precisely. The correct ϵ prescription for a Minkowski process is

$$\frac{1}{\alpha' s+1} \equiv \frac{1}{\alpha' s+1+i\epsilon} \equiv \mathrm{P}\frac{1}{\alpha' s+1} - i\pi\delta(\alpha' s+1)\,, \tag{6.4.12}$$

where P denotes the principal value. Unitarity (which we will develop more systematically in chapter 9) requires this pole to be present and determines its coefficient in terms of the amplitude for two tachyons to scatter into one:

$$S_{D_2}(k_1;k_2;k_3;k_4) = i\int \frac{d^{26}k}{(2\pi)^{26}} \frac{S_{D_2}(k_1;k_2;k)S_{D_2}(-k;k_3;k_4)}{-k^2+\alpha'^{-1}+i\epsilon} + \text{terms analytic at } k^2 = 1/\alpha'\,. \tag{6.4.13}$$

Gathering together the factors in the four-tachyon amplitude, including an equal contribution to the pole from $I(u,s)$, and using the three-tachyon result (6.4.4), the condition (6.4.13) gives

$$C_{D_2} = e^{-\lambda} C^X_{D_2} C^{\mathrm{g}}_{D_2} = \frac{1}{\alpha' g_{\mathrm{o}}^2}\,. \tag{6.4.14}$$

The three-tachyon amplitude is then

$$S_{D_2}(k_1;k_2;k_3) = \frac{2ig_{\mathrm{o}}}{\alpha'}(2\pi)^{26}\delta^{26}(\textstyle\sum_i k_i)\,. \tag{6.4.15}$$

The various functional determinants have dropped out. Using unitarity, all normalizations can be expressed in terms of the constant g_{o} appearing in the vertex operators. The determinants can in fact be computed by careful regularization and renormalization, and the relative normalizations of different topologies agree with those from unitarity.

Continuing past the pole, we encounter further singularities. Taylor expanding the integrand at $y = 0$ gives

$$I(s,t) = \int_0^r dy\,\left[y^{-\alpha' s-2} + (\alpha' t+2)y^{-\alpha' s-1} + \ldots\right]\,. \tag{6.4.16}$$

The second term gives a pole at $\alpha' s = 0$,

$$I(s,t) = \frac{u-t}{2s} + \text{terms analytic at } \alpha' s = 0\,. \tag{6.4.17}$$

From the further terms in the Taylor expansion, the amplitude has poles at

$$\alpha' s = -1, 0, 1, 2, \ldots\,. \tag{6.4.18}$$

These are precisely the positions of the open string states. The integral $I(s,t)$ also has poles in the variable t at the same positions (6.4.18), coming from the endpoint $y = 1$. These are due to the process of figure 6.3(b). The other two terms in the amplitude (6.4.9) give further contributions to the s- and t-channel poles, and also give poles in the u-channel, figure 6.3(c). Because the residue at $s = 0$ in (6.4.17) is odd in $u - t$, this pole actually cancels that from $I(s,u)$, as do all poles at even multiples of $1/\alpha'$. This will not be true for the more general open string theories to be introduced in the next section.

Define the Euler beta function

$$B(a,b) = \int_0^1 dy\, y^{a-1}(1-y)^{b-1}\,, \tag{6.4.19}$$

so that

$$I(s,t) = B(-\alpha_{\mathrm{o}}(s), -\alpha_{\mathrm{o}}(t))\,, \quad \alpha_{\mathrm{o}}(x) = 1 + \alpha' x\,. \tag{6.4.20}$$

This can be expressed in terms of gamma functions. Defining $y = v/w$ for fixed w gives

$$w^{a+b-1}B(a,b) = \int_0^w dv\, v^{a-1}(w-v)^{b-1}\,. \tag{6.4.21}$$

Multiplying both sides by e^{-w}, integrating $\int_0^\infty dw$, and regrouping gives

$$\Gamma(a+b)B(a,b) = \int_0^\infty dv\, v^{a-1}e^{-v}\int_0^\infty d(w-v)(w-v)^{b-1}e^{-(w-v)} = \Gamma(a)\Gamma(b)\,. \tag{6.4.22}$$

The four-tachyon amplitude is then

$$S_{D_2}(k_1;k_2;k_3;k_4) = \frac{2ig_{\mathrm{o}}^2}{\alpha'}(2\pi)^{26}\delta^{26}(\textstyle\sum_i k_i) \times\Big[B(-\alpha_{\mathrm{o}}(s),-\alpha_{\mathrm{o}}(t)) + B(-\alpha_{\mathrm{o}}(s),-\alpha_{\mathrm{o}}(u)) + B(-\alpha_{\mathrm{o}}(t),-\alpha_{\mathrm{o}}(u))\Big], \tag{6.4.23}$$

where

$$B(-\alpha_{\mathrm{o}}(x),-\alpha_{\mathrm{o}}(y)) = \frac{\Gamma(-\alpha' x-1)\Gamma(-\alpha' y-1)}{\Gamma(-\alpha' x-\alpha' y-2)}\,. \tag{6.4.24}$$

This is the *Veneziano amplitude*, originally written to model certain features of strong interaction phenomenology.

The high energy behavior of the Veneziano amplitude is important. There are two regions of interest, the *Regge limit*,

$$s \to \infty\,, \quad t \text{ fixed}\,, \tag{6.4.25}$$

and the *hard scattering limit*,

$$s \to \infty\,, \quad t/s \text{ fixed}\,. \tag{6.4.26}$$

If we consider the scattering process $1+2 \to 3+4$ (so that k_1^0 and k_2^0 are positive and k_3^0 and k_4^0 negative), then in the 1-2 center-of-mass frame,

$$s = E^2 , \quad t = (4m^2 - E^2)\sin^2\frac{\theta}{2} , \quad u = (4m^2 - E^2)\cos^2\frac{\theta}{2} , \qquad (6.4.27)$$

where E is the center-of-mass energy and θ is the angle between particle 1 and particle 3. The Regge limit is high energy and small angle, while the hard scattering limit is high energy and fixed angle. Using Stirling's approximation, $\Gamma(x+1) \approx x^x e^{-x}(2\pi/x)^{1/2}$, the behavior in the Regge region is

$$S_{D_2}(k_1;k_2;k_3;k_4) \approx s^{\alpha_0(t)}\Gamma(-\alpha_0(t)) , \qquad (6.4.28)$$

where again $\alpha_0(t) = \alpha' t + 1$. That is, the amplitude varies as a power of s, the power being t-dependent. This is *Regge behavior*. At the poles of the gamma function, the amplitude is an integer power of s, corresponding to exchange of a string of integer spin $\alpha_0(t)$.

In the hard scattering limit,

$$S_{D_2}(k_1;k_2;k_3;k_4) \approx \exp\left[-\alpha'(s\ln s + t\ln t + u\ln u)\right] = \exp[-\alpha' s f(\theta)] , \qquad (6.4.29)$$

where

$$f(\theta) = -\sin^2\frac{\theta}{2}\ln\sin^2\frac{\theta}{2} - \cos^2\frac{\theta}{2}\ln\cos^2\frac{\theta}{2} \qquad (6.4.30)$$

is positive. The result (6.4.29) is notable. High energy, fixed angle scattering probes the internal structure of the objects being scattered. Rutherford discovered the atomic nucleus with hard alpha–atom scattering. Hard electron–nucleon scattering at SLAC revealed the quark constituents of the nucleon. In quantum field theory, hard scattering amplitudes fall as a power of s. Even a composite object like the nucleon, if its constituents are pointlike, has power law amplitudes. The exponential falloff (6.4.29) is very much softer. The result (6.4.29) suggests a smooth object of size $\alpha'^{1/2}$, as one would expect.

We started with the three-particle amplitude, skipping over the zero-, one-, and two-particle amplitudes. We will discuss these amplitudes, and their interpretation, in section 6.6.

6.5 Chan–Paton factors and gauge interactions

In this section we will consider the interactions of the massless vector state of the open string. To make the discussion a bit more interesting, we first introduce a generalization of the open string theory.

At the end of chapter 3 we introduced a very large class of bosonic string theories, but in this first look at the interactions we are focusing

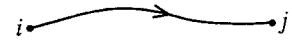


Fig. 6.4. Open string with Chan–Paton degrees of freedom.

on the simplest case of 26 flat dimensions. One can think of this in terms of symmetry: this theory has the maximal 26-dimensional Poincaré invariance. In the closed bosonic string this is the unique theory with this symmetry. An outline of the proof is as follows. The world-sheet Noether currents for spacetime translations have components of weights $(1,0)$ and $(0,1)$. By an argument given in section 2.9, these currents are then holomorphic in z or $\bar{z}$. We have seen in the calculations in this chapter that this is enough to determine all the expectation values.

In open string theory, however, there is a generalization. The open string has boundaries, endpoints. In quantum systems with distinguished points it is natural to have degrees of freedom residing at those points in addition to the fields propagating in the bulk. At each end of the open string let us add a new degree of freedom, known as a *Chan–Paton* degree of freedom, which can be in one of n states. A basis of string states is then

$$|N;k;ij\rangle , \qquad (6.5.1)$$

where i and j denote the states of the left- and right-hand endpoints, running from 1 to n. The energy-momentum tensor is defined to be the same as before, with no dependence on the new degrees of freedom. Conformal invariance is therefore automatic. Poincaré invariance is automatic as well, as the Chan–Paton degrees of freedom are invariant. Although these new degrees of freedom have trivial world-sheet dynamics, they will have a profound effect on the spacetime physics.

In string theories of the strong interaction, the motivation for this was to introduce $SU(3)$ flavor quantum numbers: the endpoints are like quarks and antiquarks, connected by a color-electric flux tube. Now we are motivating it in the general framework of considering all possibilities with given symmetries. We will give a new interpretation to the Chan–Paton degrees of freedom in chapter 8, and a possible further refinement in chapter 14.

There are now n^2 scalar tachyons, n^2 massless vector bosons, and so forth. The n^2 Hermitian matrices λ^a_{ij}, normalized to

$$\mathrm{Tr}(\lambda^a\lambda^b) = \delta^{ab} , \qquad (6.5.2)$$

are a complete set of states for the two endpoints. These are the representation matrices of $U(n)$, so one might guess that the massless vectors are

associated with a $U(n)$ guage symmetry; we will soon see that this is the case.

Define the basis

$$|N;k;a\rangle = \sum_{i,j=1}^{n} |N;k;ij\rangle \lambda^a_{ij} \,. \tag{6.5.3}$$

Now consider the four-tachyon amplitude shown in figure 6.2(a), in which the vertex operators are arranged in the cyclic order 1234. Because the Chan–Paton degrees of freedom do not appear in the Hamiltonian, their state does not evolve between the vertex operators: the right-hand endpoint of tachyon 1 must be in the same state as the left-hand endpoint of tachyon 2, and so forth. Thus, the amplitude 6.2(a) will now contain an additional factor of

$$\mathrm{Tr}\,(\lambda^{a_1}\lambda^{a_2}\lambda^{a_3}\lambda^{a_4}) \tag{6.5.4}$$

from the overlap of the Chan–Paton wavefunctions for each tachyon. This rule generalizes to an arbitrary amplitude: each vertex operator now contains a *Chan–Paton factor* λ^a_{ij} from the wavefunction of the endpoint degrees of freedom, and the amplitude for each world-sheet is multiplied by a trace of the Chan–Paton factors around each boundary.

The three-tachyon amplitude becomes

$$S_{D_2}(k_1,a_1;k_2,a_2;k_3,a_3) = \frac{ig_o}{\alpha'}(2\pi)^{26}\delta^{26}(\textstyle\sum_i k_i)\mathrm{Tr}\,(\lambda^{a_1}\lambda^{a_2}\lambda^{a_3} + \lambda^{a_1}\lambda^{a_3}\lambda^{a_2}) \,, \tag{6.5.5}$$

the two cyclic orderings now having different Chan–Paton traces. The four-tachyon amplitude is

$$\begin{aligned} S_{D_2}(k_1,a_1;k_2,a_2;k_3,a_3;k_4,a_4) &= \frac{ig_o^2}{\alpha'}(2\pi)^{26}\delta^{26}(\textstyle\sum_i k_i) \\ &\times\Big[\mathrm{Tr}\,(\lambda^{a_1}\lambda^{a_2}\lambda^{a_4}\lambda^{a_3} + \lambda^{a_1}\lambda^{a_3}\lambda^{a_4}\lambda^{a_2})B(-\alpha_o(s),-\alpha_o(t)) \\ &+ \mathrm{Tr}\,(\lambda^{a_1}\lambda^{a_3}\lambda^{a_2}\lambda^{a_4} + \lambda^{a_1}\lambda^{a_4}\lambda^{a_2}\lambda^{a_3})B(-\alpha_o(s),-\alpha_o(u)) \\ &+ \mathrm{Tr}\,(\lambda^{a_1}\lambda^{a_2}\lambda^{a_3}\lambda^{a_4} + \lambda^{a_1}\lambda^{a_4}\lambda^{a_3}\lambda^{a_2})B(-\alpha_o(t),-\alpha_o(u))\Big] \,. \end{aligned} \tag{6.5.6}$$

Considering again the unitarity relation (6.4.13), the pole at $s = -\alpha'$ acquires a factor of

$$\frac{1}{4}\mathrm{Tr}\left(\{\lambda^{a_1},\lambda^{a_2}\}\{\lambda^{a_3},\lambda^{a_4}\}\right) \tag{6.5.7}$$

on the left and a factor of

$$\frac{1}{4}\sum_a \mathrm{Tr}\left(\{\lambda^{a_1},\lambda^{a_2}\}\lambda^a\right)\mathrm{Tr}\left(\{\lambda^{a_3},\lambda^{a_4}\}\lambda^a\right) \,, \tag{6.5.8}$$

on the right, the sum being over the Chan–Paton wavefunction of the intermediate state. The completeness of the λ^a_{ij} and the normalization (6.5.2) imply that

$$\mathrm{Tr}\,(A\lambda^a)\,\mathrm{Tr}\,(B\lambda^a) = \mathrm{Tr}\,(AB) \tag{6.5.9}$$

for any matrices A and B, and so the amplitudes are still unitary.

Gauge interactions

The amplitude for a gauge boson and two tachyons is

$$\begin{aligned} S_{D_2}(k_1,a_1,e_1;k_2,a_2;k_3,a_3) &= -ig'_o g_o^2 e^{-\lambda} e_{1\mu} \\ &\times\left\langle :c^1\dot{X}^\mu e^{ik_1\cdot X}(y_1): :c^1 e^{ik_2\cdot X}(y_2): :c^1 e^{ik_3\cdot X}(y_3):\right\rangle_{D_2} \mathrm{Tr}\,(\lambda^{a_1}\lambda^{a_2}\lambda^{a_3}) \\ &+ (k_2,a_2)\leftrightarrow(k_3,a_3) \,. \end{aligned} \tag{6.5.10}$$

We have used the gauge boson vertex operator (3.6.26), but for now are allowing an independent normalization constant g'_o. Using the results of section 6.2, the X path integral is

$$\begin{aligned} \left\langle :\dot{X}^\mu e^{ik_1\cdot X}(y_1): :e^{ik_2\cdot X}(y_2): :e^{ik_3\cdot X}(y_3):\right\rangle_{D_2} &= -2i\alpha'\left(\frac{k_2^\mu}{y_{12}} + \frac{k_3^\mu}{y_{13}}\right) \\ \times iC^X_{D_2}(2\pi)^{26}\delta^{26}(\textstyle\sum_i k_i)|y_{12}|^{2\alpha' k_1\cdot k_2}|y_{13}|^{2\alpha' k_1\cdot k_3}|y_{23}|^{2\alpha' k_2\cdot k_3} \,. \end{aligned} \tag{6.5.11}$$

Using momentum conservation, the mass-shell conditions, and the physical state condition $k_1\cdot e_1 = 0$, the amplitude becomes

$$\begin{aligned} S_{D_2}(k_1,a_1,e_1;k_2,a_2;k_3,a_3) \\ = -ig'_o e_1\cdot k_{23}(2\pi)^{26}\delta^{26}(\textstyle\sum_i k_i)\mathrm{Tr}\left(\lambda^{a_1}[\lambda^{a_2},\lambda^{a_3}]\right) \,, \end{aligned} \tag{6.5.12}$$

where $k_{ij} \equiv k_i - k_j$. This is again independent of the vertex operator positions.

The $s = 0$ pole in the four-tachyon amplitude no longer vanishes. The terms that canceled now have the Chan–Paton factors in different orders, so the pole is proportional to

$$\mathrm{Tr}\left([\lambda^{a_1},\lambda^{a_2}]\,[\lambda^{a_3},\lambda^{a_4}]\right) \,. \tag{6.5.13}$$

Relating the coefficient of this pole to the amplitude (6.5.12) by unitarity, one obtains

$$g'_o = (2\alpha')^{-1/2} g_o \,. \tag{6.5.14}$$

This is the same relative normalization as from the state–operator mapping: there is only one independent coupling constant.

For the three-gauge-boson coupling, a similar calculation gives

$$S_{D_2}(k_1,a_1,e_1;k_2,a_2,e_2;k_3,a_3,e_3)$$
$$= ig_o'(2\pi)^{26}\delta^{26}(\textstyle\sum_i k_i)\Big(e_1\cdot k_{23}\, e_2\cdot e_3 + e_2\cdot k_{31}\, e_3\cdot e_1 + e_3\cdot k_{12}\, e_1\cdot e_2$$
$$+\frac{\alpha'}{2}e_1\cdot k_{23}\, e_2\cdot k_{31}\, e_3\cdot k_{12}\Big)\mathrm{Tr}\left(\lambda^{a_1}[\lambda^{a_2},\lambda^{a_3}]\right) . \qquad (6.5.15)$$

Up to first order in the momenta, the amplitudes we have found are reproduced by the spacetime action

$$S = \frac{1}{g_o'^2}\int d^{26}x\left[-\frac{1}{2}\mathrm{Tr}\,(D_\mu\varphi D^\mu\varphi) + \frac{1}{2\alpha'}\mathrm{Tr}\,(\varphi^2) + \frac{2^{1/2}}{3\alpha'^{1/2}}\mathrm{Tr}\,(\varphi^3)\right.$$
$$\left. - \frac{1}{4}\mathrm{Tr}\,(F_{\mu\nu}F^{\mu\nu})\right] , \qquad (6.5.16)$$

where the tachyon field φ and the Yang–Mills vector potential A_μ are written as $n\times n$ matrices, e. g. $A_\mu = A_\mu^a\lambda^a$. Also, $D_\mu\varphi = \partial_\mu\varphi - i[A_\mu,\varphi]$ and $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu - i[A_\mu,A_\nu]$.

This is the action for a $U(n)$ gauge field coupled to a scalar in the adjoint representation. Adding in the Chan–Paton factors has produced just the gauge-invariant expressions needed. The gauge invariance is automatic because the decoupling of unphysical states is guaranteed in string perturbation theory, as we will develop further.

At momenta k small compared to the string scale, the only open string states are the massless gauge bosons. As discussed in section 3.7, it is in this limit that the physics should reduce to an effective field theory of the massless states. We have therefore been somewhat illogical in including the tachyon in the action (6.5.16); we did so for illustration, but now let us focus on the gauge bosons. The four-gauge-boson amplitude has a form analogous to the Veneziano amplitude but with additional structure from the polarization tensors. Expanding it in powers of $\alpha'k^2$, the first term, which survives in the *zero-slope limit* $\alpha'\to 0$, is the sum of pole terms in s, t, and u plus a constant. Consistency guarantees that it is precisely the four-gauge-boson amplitude obtained in field theory from the Yang–Mills Lagrangian $F_{\mu\nu}F^{\mu\nu}$. Note, however, the term of order $\alpha'k^3$ in the three-gauge-boson amplitude (6.5.15). This implies a higher derivative term

$$-\frac{2i\alpha'}{3g_o'^2}\mathrm{Tr}\,(F_\mu{}^\nu F_\nu{}^\omega F_\omega{}^\mu) \qquad (6.5.17)$$

in the Lagrangian. Similarly, expanding the four-point amplitude reveals an infinite sum of higher order interactions (beyond (6.5.17), these do not contribute to the three-gauge-boson amplitude for kinematic reasons). String loop amplitudes also reduce, in the low energy region, to the loops

obtained from the effective Lagrangian. By the usual logic of effective actions, the higher derivative terms are less important at low energy. The scale where they become important, $\alpha'k^2\approx 1$, is just where new physics (massive string states) appears and the effective action is no longer applicable.

If we have a cutoff, why do we need renormalization theory? Renormalization theory still has content and in fact this is its real interpretation: it means that low energy physics is independent of the details of the high energy theory, except for the parameters in the effective Lagrangian. This is a mixed blessing: it means that we can use ordinary quantum field theory to make predictions at accelerator energies without knowing the form of the Planck scale theory, but it also means that we cannot probe the Planck scale theory with physics at particle accelerators.

The string spectrum and amplitudes have an obvious global $U(n)$ symmetry,

$$\lambda^a \to U\lambda^a U^\dagger , \qquad (6.5.18)$$

which leaves the Chan–Paton traces and the norms of states invariant. From the detailed form of the amplitudes, we have learned that this is actually a *local* symmetry in spacetime. We will see that this promotion of a global world-sheet symmetry to a local spacetime symmetry is a rather general phenomenon in string theory.

All open string states transform as the $n\times n$ adjoint representation under the $U(n)$ symmetry. Incidentally, $U(n)$ is not a simple Lie algebra: $U(n) = SU(n)\times U(1)$. The $U(1)$ gauge bosons, $\lambda_{ij} = \delta_{ij}/n^{1/2}$, decouple from the amplitudes (6.5.12) and (6.5.15). The adjoint representation of $U(1)$ is trivial, so all string states are neutral under the $U(1)$ symmetry.

The unoriented string

It is interesting to generalize to the unoriented string. Consider first the theory without Chan–Paton factors. In addition to Möbius invariance, the X^μ CFT on the sphere or disk is invariant under the orientation-reversing symmetry $\sigma^1\to\pi-\sigma^1$ in the open string or $\sigma^1\to 2\pi-\sigma^1$ in the closed string. This *world-sheet parity* symmetry is generated by an operator Ω. From the mode expansions it follows that

$$\Omega\alpha_n^\mu\Omega^{-1} = (-1)^n\alpha_n^\mu \qquad (6.5.19)$$

in the open string and

$$\Omega\alpha_n^\mu\Omega^{-1} = \tilde\alpha_n^\mu \qquad (6.5.20)$$

in the closed string. The symmetry extends to the ghosts as well, but to avoid distraction we will not discuss them explicitly, since they only contribute a fixed factor to the tree-level amplitudes.

The tachyon vertex operators are even under world-sheet parity in either the closed or open string (this is obvious for the integrated vertex operators, without ghosts; the fixed operators must then transform in the same way), determining the sign of the operator. All states can then be classified by their parity eigenvalue $\omega = \pm 1$. The relation (6.5.19) in the open string implies that

$$\Omega|N;k\rangle = \omega_N|N;k\rangle, \quad \omega_N = (-1)^{1+\alpha' m^2}. \tag{6.5.21}$$

World-sheet parity is multiplicatively conserved. For example, the three-tachyon amplitude is nonzero, consistent with $(+1)^3 = 1$. On the other hand, the massless vector has $\omega = -1$ and so we would expect the vector–tachyon–tachyon amplitude (6.5.10) and three-vector amplitude (6.5.15) to vanish in the absence of Chan–Paton degrees of freedom, as indeed they do (the λ^a are replaced by 1 and the commutators vanish). The different cyclic orderings, related to one another by world-sheet parity, cancel.

Given a consistent oriented string theory, we can make a new *unoriented* string theory by restricting the spectrum to the states of $\omega = +1$. States with odd $\alpha' m^2$ remain, while states with even $\alpha' m^2$, including the photon, are absent. The conservation of ω guarantees that if all external states have $\omega = +1$, then the intermediate states in tree-level amplitudes will also have $\omega = +1$. Unitarity of the unoriented theory thus follows from that of the oriented theory, at least at tree level (and in fact to all orders, as we will outline in chapter 9).

The main point of interest in the unoriented theory is the treatment of the Chan–Paton factors. Since we have identified these with the respective endpoints of the open string, world-sheet parity must reverse them,

$$\Omega|N;k;ij\rangle = \omega_N|N;k;ji\rangle \ . \tag{6.5.22}$$

Again, this is a symmetry of all amplitudes in the oriented theory. To form the unoriented theory, we again restrict the spectrum to world-sheet parity eigenvalue $\omega = +1$. Take a basis for λ^a_{ij} in which each matrix is either symmetric, $s^a = +1$, or antisymmetric, $s^a = -1$. Then

$$\Omega|N;k;a\rangle = \omega_N s^a|N;k;a\rangle \ . \tag{6.5.23}$$

The world-sheet parity eigenvalue is $\omega = \omega_N s^a$, and the unoriented spectrum is

$$\alpha' m^2 \text{ even:} \quad \lambda^a \text{ antisymmetric} \ , \tag{6.5.24a}$$
$$\alpha' m^2 \text{ odd:} \quad \lambda^a \text{ symmetric} \ . \tag{6.5.24b}$$

For the massless gauge bosons the Chan–Paton factors are the $n \times n$ antisymmetric matrices and so the gauge group is $SO(n)$. The states at even mass levels transform as the adjoint representation of the orthogonal group $SO(n)$, and the states at odd mass levels transform as the traceless symmetric tensor plus singlet representations.

The oriented theory has a larger set of orientation-reversing symmetries, obtained as a combination of Ω and a $U(n)$ rotation,

$$\Omega_\gamma|N;k;ij\rangle = \omega_N \gamma_{jj'}|N;k;j'i'\rangle\gamma^{-1}_{i'i} \ . \tag{6.5.25}$$

We can form more general unoriented theories by restricting the spectrum to $\omega_\gamma = +1$, which is again consistent with the interactions. Acting twice with Ω_γ gives

$$\Omega^2_\gamma|N;k;ij\rangle = [\gamma(\gamma^T)^{-1}]_{ii'}|N;k;i'j'\rangle(\gamma^{-1}\gamma^T)_{j'j} \ . \tag{6.5.26}$$

We will insist that $\Omega^2_\gamma = 1$ for reasons to be explained below. This then implies that

$$\gamma^T = \pm\gamma \ . \tag{6.5.27}$$

That is, γ is symmetric or antisymmetric.

A general change of Chan–Paton basis,

$$|N;k;ij\rangle' = U^{-1}_{ii'}|N;k;i'j'\rangle U_{j'j} \ , \tag{6.5.28}$$

transforms γ to

$$\gamma' = U^T\gamma U \ . \tag{6.5.29}$$

In the symmetric case, it is always possible to find a basis such that $\gamma = 1$, giving the theory already considered above. In the antisymmetric case there is a basis in which

$$\gamma = M \equiv i\begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix} \ . \tag{6.5.30}$$

Here I is the $k \times k$ identity matrix and $n = 2k$ must be even because γ is an invertible antisymmetric matrix. We take a basis for the Chan–Paton wavefunctions such that $M(\lambda^a)^T M = s^{a'}\lambda^a$ with $s^{a'} = \pm 1$. Then the world-sheet parity eigenvalue is $\omega_\gamma = \omega_N s^{a'}$, and the unoriented spectrum is

$$\alpha' m^2 \text{ even:} \quad M(\lambda^a)^T M = -\lambda^a \ , \tag{6.5.31a}$$
$$\alpha' m^2 \text{ odd:} \quad M(\lambda^a)^T M = +\lambda^a \ . \tag{6.5.31b}$$

At the even mass levels, including the gauge bosons, this defines the adjoint representation of the symplectic group $Sp(k)$.

The argument that we must have $\Omega^2_\gamma = 1$ to construct the unoriented theory is as follows. Since $\Omega^2 = 1$, it must be that Ω^2_γ acts only on the

Chan–Paton states, not the oscillators. In fact, from eq. (6.5.26), it acts on the Chan–Paton wavefunction as

$$\lambda \to \gamma(\gamma^T)^{-1}\lambda\gamma^{-1}\gamma^T = \lambda \,, \tag{6.5.32}$$

where the last equality must hold in the unoriented theory, since all states in this theory are invariant under Ω_γ and so also under Ω_γ^2. Now we assert that the allowed Chan–Paton wavefunctions must form a complete set. The point is that two open strings, by the splitting–joining interaction of figure 3.4(c), can exchange endpoints. In this way one can get to a complete set, that is, to any Chan–Paton state $|ij\rangle$ (this argument is slightly heuristic, but true). By Schur's lemma, if eq. (6.5.32) holds for a complete set, then $\gamma(\gamma^T)^{-1} = \gamma^{-1}\gamma^T = 1$ and so $\Omega_\gamma^2 = 1$.

We might try to obtain other gauge groups by taking different sets of λ^a. In fact, the oriented $U(n)$ and unoriented $SO(n)$ and $Sp(k)$ theories constructed above are the only possibilities. Generalization of the completeness argument in eqs. (6.5.7)–(6.5.9) shows these to be the most general solutions of the unitarity conditions. In particular, exceptional Lie algebras, which are of interest in grand unification and which will play a major role in volume two, cannot be obtained with Chan–Paton factors (in perturbation theory). In closed string theory there is another mechanism that gives rise to gauge bosons, and it allows other groups.

6.6 Closed string tree amplitudes

The discussion of closed string amplitudes is parallel to the above. The amplitude for three closed string tachyons is

$$S_{S_2}(k_1;k_2;k_3) = g_c^3 e^{-2\lambda}\left\langle \prod_{i=1}^{3} :\tilde{c}ce^{ik_i\cdot X}(z_i,\bar{z}_i): \right\rangle_{S_2} . \tag{6.6.1}$$

In this case the CKG $PSL(2,\mathbf{C})$ (the Möbius group) can be used to fix the three vertex operators to arbitrary positions $z_{1,2,3}$. Taking the expectation values from section 6.2, the result is again independent of the vertex operator positions,

$$S_{S_2}(k_1;k_2;k_3) = ig_c^3 C_{S_2}(2\pi)^{26}\delta^{26}(\textstyle\sum_i k_i) \,, \tag{6.6.2}$$

where $C_{S_2} = e^{-2\lambda}C_{S_2}^X C_{S_2}^{\mathrm{g}}$.

For four closed string tachyons,

$$S_{S_2}(k_1;k_2;k_3;k_4) = g_c^4 e^{-2\lambda}\int_{\mathbf{C}} d^2z_4 \left\langle \prod_{i=1}^{3} :\tilde{c}ce^{ik_i\cdot X}(z_i,\bar{z}_i): :e^{ik_4\cdot X}(z_4,\bar{z}_4): \right\rangle_{S_2} , \tag{6.6.3}$$

where the integral runs over the complex plane $\mathbf{C}$. Evaluating the expectation value and setting $z_1 = 0$, $z_2 = 1$, $z_3 = \infty$, this becomes

$$S_{S_2}(k_1;k_2;k_3;k_4) = ig_c^4 C_{S_2}(2\pi)^{26}\delta^{26}(\textstyle\sum_i k_i)J(s,t,u) \,, \tag{6.6.4}$$

where

$$J(s,t,u) = \int_{\mathbf{C}} d^2z_4\, |z_4|^{-\alpha' u/2-4}|1-z_4|^{-\alpha' t/2-4} . \tag{6.6.5}$$

Here $s+t+u = -16/\alpha'$, but we indicate the dependence of J on all three variables to emphasize its symmetry among them. This amplitude converges when $s,t,u < -4/\alpha'$. It has poles in the variable u from $z_4 \to 0$, in the variable t from $z_4 \to 1$, and in the variable s from $z_4 \to \infty$. The poles are at the values

$$\alpha' s \,,\ \alpha' t \,,\ \alpha' u = -4, 0, 4, 8, \ldots \,, \tag{6.6.6}$$

which are the masses-squared of the closed string states. The pole at $\alpha' s = -4$ is

$$ig_c^4 C_{S_2}\int_{|z_4|>1/\epsilon} d^2z_4\, |z_4|^{\alpha' s/2} \sim -\frac{8\pi i g_c^4 C_{S_2}}{\alpha' s + 4} . \tag{6.6.7}$$

Unitarity gives

$$C_{S_2} = \frac{8\pi}{\alpha' g_c^2} \,, \tag{6.6.8}$$

and so

$$S_{S_2}(k_1;k_2;k_3) = \frac{8\pi i g_c}{\alpha'}(2\pi)^{26}\delta^{26}(\textstyle\sum_i k_i) . \tag{6.6.9}$$

Like the Veneziano amplitude, the amplitude for four closed string tachyons can be expressed in terms of gamma functions (exercise 6.10):

$$S_{S_2}(k_1;k_2;k_3;k_4) = \frac{8\pi i g_c^2}{\alpha'}(2\pi)^{26}\delta^{26}(\textstyle\sum_i k_i)C(-\alpha_c(t),-\alpha_c(u)) \,, \tag{6.6.10}$$

where $\alpha_c(x) = 1 + \alpha' x/4$ and

$$\begin{aligned} C(a,b) &= \int_{\mathbf{C}} d^2z\, |z|^{2a-2}|1-z|^{2b-2} \\ &= 2\pi\frac{\Gamma(a)\Gamma(b)\Gamma(c)}{\Gamma(a+b)\Gamma(a+c)\Gamma(b+c)} \,, \quad a+b+c = 1 . \end{aligned} \tag{6.6.11}$$

This is the *Virasoro–Shapiro amplitude*. There is just a single term, with poles in the s-, t-, and u-channels coming from the gamma functions in the numerator. Like the Veneziano amplitude, the Virasoro–Shapiro amplitude has Regge behavior in the Regge limit,

$$S_{S_2}(k_1;k_2;k_3;k_4) \propto s^{2\alpha_c(t)}\frac{\Gamma(-\alpha_c(t))}{\Gamma(1+\alpha_c(t))} \,, \tag{6.6.12}$$

and exponential behavior in the hard scattering limit,

$$S_{S_2}(k_1;k_2;k_3;k_4) \propto \exp\left[-\frac{\alpha'}{2}(s\ln s + t\ln t + u\ln u)\right] . \tag{6.6.13}$$

On the sphere, the amplitude for a massless closed string and two closed string tachyons is

$$\begin{aligned} S_{S_2}(k_1,e_1;k_2;k_3) &= g_c^2 g_c' e^{-2\lambda} e_{1\mu\nu} \left\langle :\tilde{c}c\partial X^\mu \bar{\partial} X^\nu e^{ik_1\cdot X}(z_1,\bar{z}_1): \right. \\ &\qquad \left. :\tilde{c}ce^{ik_2\cdot X}(z_2,\bar{z}_2): :\tilde{c}ce^{ik_3\cdot X}(z_3,\bar{z}_3):\right\rangle_{S_2} \\ &= -\frac{\pi i\alpha'}{2} g_c' e_{1\mu\nu} k_{23}^\mu k_{23}^\nu (2\pi)^{26}\delta^{26}(\textstyle\sum_i k_i) , \end{aligned} \tag{6.6.14}$$

where $e_{1\mu\nu}e_1^{\mu\nu} = 1$. Expanding the Virasoro–Shapiro amplitude (6.6.10) on the $s = 0$ pole and using unitarity determines

$$g_c' = \frac{2}{\alpha'} g_c , \tag{6.6.15}$$

again in agreement with the state–operator mapping with overall constant g_c. The amplitude (6.6.14) would be obtained in field theory from the spacetime action $S + S_T$, where S is the action (3.7.20) for the massless fields, and where

$$S_T = -\frac{1}{2}\int d^{26}x\,(-G)^{1/2} e^{-2\tilde{\Phi}}\left(G^{\mu\nu}\partial_\mu T\partial_\nu T - \frac{4}{\alpha'}T^2\right) , \tag{6.6.16}$$

is the action for the closed string tachyon T. For example, the amplitude for a graviton of polarization $e_{\mu\nu}$ is obtained from this action by expanding

$$\tilde{G}_{\mu\nu} = \eta_{\mu\nu} + 2\kappa e_{\mu\nu} e^{ik\cdot x} . \tag{6.6.17}$$

Note that this is the Einstein metric, whose action (3.7.25) is independent of the dilaton. The normalization of the fluctuation is determined by that of the graviton kinetic term in the spacetime action. Specifically, if one takes $\tilde{G}_{\mu\nu} - \eta_{\mu\nu} = 2\kappa e_{\mu\nu} f(x)$ with $e_{\mu\nu}e^{\mu\nu} = 1$, the effective action for f has the canonical normalization $\frac{1}{2}$ for a real scalar. The field theory amplitude matches the string result (6.6.14) and relates the normalization of the vertex operators to the gravitational coupling,

$$\kappa = \pi\alpha' g_c' = 2\pi g_c . \tag{6.6.18}$$

The amplitude for three massless closed strings is

$$S_{S_2}(k_1,e_1;k_2,e_2;k_3,e_3) = \frac{i\kappa}{2}(2\pi)^{26}\delta^{26}(\textstyle\sum_i k_i) e_{1\mu\nu} e_{2\alpha\beta} e_{3\gamma\delta} T^{\mu\alpha\gamma} T^{\nu\beta\delta} , \tag{6.6.19}$$

where

$$T^{\mu\alpha\gamma} = k_{23}^\mu \eta^{\alpha\gamma} + k_{31}^\alpha \eta^{\gamma\mu} + k_{12}^\gamma \eta^{\mu\alpha} + \frac{\alpha'}{8} k_{23}^\mu k_{31}^\alpha k_{12}^\gamma . \tag{6.6.20}$$

The order k^2 terms in this amplitude correspond to the spacetime action (3.7.25), while the k^4 and k^6 terms come from a variety of higher derivative interactions, including terms quadratic and cubic in the spacetime curvature. These higher corrections to the action can also be determined by calculating higher loop corrections to the world-sheet beta functions (3.7.14).

The tensor structure of the closed string amplitude (6.6.19) is just two copies of that in the open string amplitude (6.5.15), if one sets $\alpha' = 2$ in the closed string and $\alpha' = \frac{1}{2}$ in the open. The same holds for the amplitude (6.6.14). This is a consequence of the factorization of free-field expectation values on the sphere into holomorphic and antiholomorphic parts. A similar factorization holds for four or more closed strings *before* integration over the vertex operator positions. Further, by careful treatment of the contour of integration it is possible to find relations between the integrated amplitudes. For the four-tachyon amplitudes the integrals above are related

$$J(s,t,u,\alpha') = -2\sin\pi\alpha_c(t)\, I(s,t,4\alpha')\, I(t,u,4\alpha') , \tag{6.6.21}$$

after use of the gamma function identity $\Gamma(x)\Gamma(1-x)\sin(\pi x) = -\pi$; we now indicate the explicit dependence of the integrals on α'. For the general integral appearing in four-point closed string amplitudes there is a relation

$$\begin{aligned} &\int_C d^2z\, z^{a-1+m_1}\bar{z}^{a-1+n_1}(1-z)^{b-1+m_2}(1-\bar{z})^{a-1+n_2} \\ &\quad = \sin[\pi(b+n_2)]B(a+m_1,b+m_2)B(b+n_2,1-a-b-n_1-n_2) . \end{aligned} \tag{6.6.22}$$

This implies a corresponding relation between four-point open and closed string amplitudes

$$A_c(s,t,u,\alpha',g_c) = \frac{\pi i g_c^2\alpha'}{g_o^4}\sin[\pi\alpha_c(t)] A_o(s,t,\tfrac{1}{4}\alpha',g_o) A_o(t,u,\tfrac{1}{4}\alpha',g_o)^* , \tag{6.6.23}$$

where the open string amplitudes include just one of the six cyclic permutations, with poles in the indicated channels.

Consistency

In chapter 9 we will discuss the convergence and gauge invariance of tree-level amplitudes in a general way, but as an introduction we will now use the OPE to see how these work for the lowest levels. Consider the operator product

$$\begin{aligned} :e^{ik_1\cdot X(z_1,\bar{z}_1)}: :e^{ik_4\cdot X(z_4,\bar{z}_4)}: &= |z_{14}|^{\alpha' k_1\cdot k_4} :\Big(1 + iz_{14}k_1\cdot\partial X + i\bar{z}_{14}k_1\cdot\bar{\partial}X \\ &\quad - z_{14}\bar{z}_{14}k_1\cdot\partial X k_1\cdot\bar{\partial}X + \ldots\Big) e^{i(k_1+k_4)\cdot X}(z_4,\bar{z}_4): . \end{aligned} \tag{6.6.24}$$

This appears in the amplitude with z_{14} integrated. The integral converges as $z_{14} \to 0$ when

$$\alpha' k_1 \cdot k_4 = \frac{\alpha'}{2}(k_1 + k_4)^2 - 4 > -2 \tag{6.6.25}$$

and has a pole at that point. The coefficient of the pole is just the tachyon vertex operator. Thus, if a pair of tachyons in any amplitude has total momentum $(k_1 + k_4)^2 = 4/\alpha'$, there will be a pole proportional to the amplitude with one fewer tachyon as required by unitarity. A pole in momentum space corresponds to long distance in spacetime, so this is a process in which two tachyons scatter into one, which then propagates and interacts with the remaining particles. Carrying the OPE further, the $O(z_{14})$ and $O(\bar{z}_{14})$ terms do not produce poles because the angular integration gives zero residue, the $O(z_{14}\bar{z}_{14})$ term gives a massless pole, and so on.

Now let us look a little more closely at the way the local spacetime symmetries are maintained in the string amplitudes. The various amplitudes we have calculated all vanish if any of the polarizations are of the form $e_\mu = k_\mu$, or $e_{\mu\nu} = \zeta_\mu k_\nu + k_\mu \tilde{\zeta}_\nu$ with $k\cdot\zeta = k\cdot\tilde{\zeta} = 0$. The corresponding spacetime actions are thus invariant under the Yang–Mills, coordinate, and antisymmetric tensor symmetries. As discussed in section 3.6, the vertex operator for a longitudinal polarization is the sum of a total derivative and a term that vanishes by the equations of motion. Upon integration, the total derivative vanishes but the equation of motion term might have a source at one of the other vertex operators in the path integral. The operator product (6.6.24) vanishes rapidly when $k_1 \cdot k_4$ is large. Using this property, there will be for any pair of vertex operators a kinematic region in which all possible contact terms are suppressed. The amplitude for any null polarization then vanishes *identically* in this region, and since all amplitudes are analytic except for poles (and branch points at higher order) the amplitudes for null polarizations must vanish everywhere. We see that the singularities required by unitarity, as well as any possible divergences or violations of spacetime gauge invariance, arise from the limits $z \to 0, 1, \infty$, where two vertex operators come together. For the sphere with four marked points (vertex operators) these are the boundaries of moduli space. The analytic continuation argument used here is known for historic reasons as the *canceled propagator argument.*

Closed strings on D_2 and RP_2

The lowest order closed–open interactions come from the disk with both closed and open vertex operators. The low energy effective action for these can be deduced from general considerations. With a trivial closed string background we found the usual gauge kinetic term,

$$-\frac{1}{4g_o'^2}\int d^{26}x \,\mathrm{Tr}\,(F_{\mu\nu}F^{\mu\nu})\,. \tag{6.6.26}$$

Obviously the metric must couple to this in a covariant way. In addition, the coupling of the dilaton can be deduced. Recall that $g_o'^2 \propto e^{\Phi_0}$, with Φ_0 the expectation value of the dilaton. So we should replace this (inside the integral) with $\Phi = \Phi_0 + \tilde{\Phi}$. The action

$$-\frac{1}{4g_o'^2}\int d^{26}x\,(-G)^{1/2}e^{-\tilde{\Phi}}\mathrm{Tr}\,(F_{\mu\nu}F^{\mu\nu}) \tag{6.6.27}$$

thus incorporates all interactions not involving derivatives of the closed string fields; indices are now raised and lowered with $G_{\mu\nu}$. This action reflects the general principle that the effective action from Euler number χ is weighted by $g_c^{-\chi} \propto e^{-\chi\Phi}$.

The disk and projective plane also make a contribution to purely closed string interactions. The amplitudes for n closed strings are of order $g_o^{-2}g_c^n \sim g_c^{n-1}$, one power of g_c higher than the sphere. Closed string loop amplitudes, to be considered in the next chapter, are of order g_c^2 times the sphere, because emitting and reabsorbing a closed string adds two factors of g_c. Thus the disk and projective plane are 'half-loop order.'

Of particular interest are the amplitudes on the disk and the projective plane with a single closed string vertex operator. Fixing the position of the vertex operator removes only two of the three CKVs, the residual gauge symmetry consisting of rotations about the vertex operator position. Thus we have to divide the amplitude by the volume of this residual CKG. We have not shown how to do this explicitly, but will work it out for the torus in the next chapter. For the disk with one closed string this is a finite factor and the result is nonzero. The amplitude is a numerical factor times $g_c g_o^{-2}$, which in turn is a pure number, times powers of α' as required by dimensional analysis. We will not work out these numerical factors here, but will obtain them in an indirect way in chapter 8.

Thus there is an amplitude for a single closed string to appear from the vacuum, either through the disk or the projective plane, necessarily with zero momentum. Such an amplitude is known as a *tadpole.* In other words, the background closed string fields are corrected from their original values at order g_c. Again we can write an effective action, which is simply

$$-\Lambda\int d^{26}x\,(-G)^{1/2}e^{-\tilde{\Phi}}\,, \tag{6.6.28}$$

the Φ dependence deduced as above. This is a potential for the dilaton. We will consider it further in the next chapter.

The amplitude for a single closed string on the sphere, on the other hand, is zero. The residual CKV is a noncompact subgroup of $PSL(2,\mathbf{C})$ and so

one has to divide by an infinite volume. A nonzero result would have been a logical inconsistency, a zeroth order correction to the background fields. Similarly the amplitude for two closed strings on the sphere (a zeroth order correction to the mass) vanishes, as do the corresponding disk amplitudes, one or two open strings. The amplitudes with no vertex operators at all are also meaningful — they just calculate the term of order $\tilde{\Phi}^0$ in the Taylor expansion of the action (6.6.28). The disk amplitude with no vertex operators is thus nonvanishing; this requires a somewhat formal treatment of the conformal Killing volume.

6.7 General results

In this section we obtain some general results concerning CFT on the sphere and disk.

Möbius invariance

We have seen that the sphere has a group of globally defined conformal transformations, the Möbius group $PSL(2,\mathbf{C})$,

$$z' = \frac{\alpha z + \beta}{\gamma z + \delta} \tag{6.7.1}$$

for complex $\alpha, \beta, \gamma, \delta$ with $\alpha\delta - \beta\gamma = 1$. This is the most general conformal transformation that is one-to-one on all of S_2, the complex z-plane plus the point at infinity. Expectation values must be invariant under any Möbius transformation:

$$\langle \mathscr{A}_i(z_1,\bar{z}_1)\dots\mathscr{A}_k(z_n,\bar{z}_n)\rangle_{S_2} = \langle \mathscr{A}'_i(z_1,\bar{z}_1)\dots\mathscr{A}'_k(z_n,\bar{z}_n)\rangle_{S_2} . \tag{6.7.2}$$

We will consider the consequences of this symmetry for expectation values with one, two, three, or four local operators.

For a single operator of weight $(h_i,\tilde{h}_i)$, the rescaling plus rotation $z' = \gamma z$ gives

$$\langle \mathscr{A}_i(0,0)\rangle_{S_2} = \langle \mathscr{A}'_i(0,0)\rangle_{S_2} = \gamma^{-h_i}\bar{\gamma}^{-\tilde{h}_i}\langle \mathscr{A}_i(0,0)\rangle_{S_2} . \tag{6.7.3}$$

The one-point function therefore vanishes unless $h_i = \tilde{h}_i = 0$. This is another way to see that the one-point string amplitude vanishes on the sphere, because the matter factor is the expectation value of a $(1,1)$ operator.

For $n = 2$, we can use a translation plus $z' = \gamma z$ to bring any pair of operators to the points 0 and 1, giving

$$\langle \mathscr{A}_i(z_1,\bar{z}_1)\mathscr{A}_j(z_2,\bar{z}_2)\rangle_{S_2} = z_{12}^{-h_i-h_j}\bar{z}_{12}^{-\tilde{h}_i-\tilde{h}_j}\langle \mathscr{A}_i(1,1)\mathscr{A}_j(0,0)\rangle_{S_2} , \tag{6.7.4}$$

so the position dependence is completely determined. Single-valuedness implies that $J_i + J_j \in \mathbf{Z}$, where $J_i = h_i - \tilde{h}_i$. There is a further constraint on the two-point function from the conformal transformation $z' = z + \epsilon(z - z_1)(z - z_2) + O(\epsilon^2)$, which leaves z_1 and z_2 fixed. For general operators this is complicated, but for *tensor fields* $\mathscr{O}_p$ and $\mathscr{O}_q$ it simply implies that

$$\langle \mathscr{O}_p(z_1,\bar{z}_1)\mathscr{O}_q(z_2,\bar{z}_2)\rangle_{S_2} = 0 \quad \text{unless } h_p = h_q,\ \ \tilde{h}_p = \tilde{h}_q . \tag{6.7.5}$$

Any three points $z_{1,2,3}$ can be brought to given positions by a Möbius transformation. For $n \geq 3$, Möbius invariance therefore reduces the expectation value from a function of n complex variables to a function of $n-3$ complex variables. Again the result takes a simple form only for tensor fields. For example, for three tensor fields one finds

$$\left\langle \prod_{i=1}^{3} \mathscr{O}_{p_i}(z_i,\bar{z}_i) \right\rangle_{S_2} = C_{p_1p_2p_3} \prod_{\substack{i,j=1\\ i<j}}^{3} z_{ij}^{h-2(h_i+h_j)}\bar{z}_{ij}^{\tilde{h}-2(\tilde{h}_i+\tilde{h}_j)} , \tag{6.7.6}$$

where $C_{p_1p_2p_3}$ is independent of position and $h = h_1 + h_2 + h_3$. For four primary fields,

$$\left\langle \prod_{i=1}^{4} \mathscr{O}_{p_i}(z_i,\bar{z}_i) \right\rangle_{S_2} = C_{p_1p_2p_3p_4}(z_c,\bar{z}_c)\,(z_{12}z_{34})^h(\bar{z}_{12}\bar{z}_{34})^{\tilde{h}} \times \prod_{\substack{i,j=1\\ i<j}}^{4} z_{ij}^{-h_i-h_j}\bar{z}_{ij}^{-\tilde{h}_i-\tilde{h}_j} , \tag{6.7.7}$$

where $h = \sum_i h_i$, $\tilde{h} = \sum_i \tilde{h}_i$, and $z_c = z_{12}z_{34}/z_{13}z_{24}$ is the Möbius invariant cross-ratio. The function $C_{p_1p_2p_3p_4}(z_c,\bar{z}_c)$ is not determined by conformal invariance, so we are reduced from a function of four variables to an unknown function of one variable.

On the disk represented as the upper half-plane, only the Möbius transformations with $\alpha,\beta,\gamma,\delta$ real remain, forming the group $PSL(2,\mathbf{R})$. The extension of the above is left as an exercise. One learns much more by considering the full conformal algebra. We will see in chapter 15 that it determines all expectation values in terms of those of the tensor fields.

Path integrals and matrix elements

The path integrals we have considered can be related to operator expressions. Consider the path integral on the sphere with two operators, one at the origin and one at infinity:

$$\langle \mathscr{A}'_i(\infty,\infty)\mathscr{A}_j(0,0)\rangle_{S_2} . \tag{6.7.8}$$

The prime indicates the u-frame, which we have to take for the operator at infinity; by a slight abuse of notation we still give the position in terms of z. Using the state–operator mapping we can replace the disk $|z| < 1$ containing $\mathscr{A}_j$ at $z = 0$ by the state $\Psi_{\mathscr{A}_j}$ on the circle $|z| = 1$. We can

also replace the disk $|z| > 1$ ($|u| < 1$) containing $\mathscr{A}_i$ at $u = 0$ by the state $\Psi_{\mathscr{A}_i}$ on the circle $|z| = 1$. All that is left is the integral over the fields ϕ_b on the circle, so the expectation value (6.7.8) becomes

$$\int [d\phi_b]\Psi_{\mathscr{A}_i}[\phi_b^\Omega]\Psi_{\mathscr{A}_j}[\phi_b] \ ; \tag{6.7.9}$$

here $\phi_b^\Omega(\sigma) = \phi_b(2\pi - \sigma)$, arising from the mapping $zu = 1$.

This convolution of wavefunctions resembles an inner product, so we define

$$\langle\!\langle i|j\rangle\!\rangle = \langle \mathscr{A}_i'(\infty,\infty)\mathscr{A}_j(0,0)\rangle_{S_2} \ . \tag{6.7.10}$$

This is essentially the inner product introduced by Zamolodchikov. For tensor operators in unitary theories, a Möbius transformation shows that this is the same as the coefficient of 1 in the OPE,

$$\mathscr{O}_i(z,\bar z)\mathscr{O}_j(0,0) = \frac{\langle\!\langle i|j\rangle\!\rangle}{z^{h_i+h_j}\bar z^{\tilde h_i+\tilde h_j}\langle 1\rangle_{S_2}} + \dots \ . \tag{6.7.11}$$

We have abbreviated $|\mathscr{A}_i\rangle$ to $|i\rangle$. The $\langle\!\langle\ |\ \rangle\!\rangle$ product is not the same as the quantum mechanical inner product $\langle\ |\ \rangle$ introduced in chapter 4. The latter is Hermitean, whereas the former includes no complex conjugation and so is bilinear, up to a sign if i and j are anticommuting. That is,

$$\langle\!\langle i|j\rangle\!\rangle = \pm\langle\!\langle j|i\rangle\!\rangle \ , \tag{6.7.12}$$

since all we have done is to interchange the two operators and rename $z \leftrightarrow u$. There is a simple relation between the two inner products, which we will develop later.

We will sometimes write $\mathscr{G}_{ij}$ for $\langle\!\langle i|j\rangle\!\rangle$, and $\mathscr{G}^{ij}$ for the inverse matrix, where i, j run over a complete set. The matrices $\mathscr{G}_{ij}$ and $\mathscr{G}^{ij}$ will be used to raise and lower indices, $\mathscr{A}^i = \mathscr{G}^{ij}\mathscr{A}_j$, $\mathscr{A}_i = \mathscr{G}_{ij}\mathscr{A}^j$.

Operators in the path integral translate in the usual way into operators in Hilbert space. For example,

$$\langle \mathscr{A}_i'(\infty,\infty)\mathscr{A}_k(1,1)\mathscr{A}_j(0,0)\rangle_{S_2} = \langle\!\langle i|\hat{\mathscr{A}}_k(1,1)|j\rangle\!\rangle \ , \tag{6.7.13}$$

where we reintroduce the hat to emphasize that we are in a Hilbert space formalism. Using the OPE, the left-hand side becomes

$$\sum_l c_{kj}^l \langle \mathscr{A}_i'(\infty,\infty)\mathscr{A}_l(0,0)\rangle_{S_2} = c_{ikj} \ . \tag{6.7.14}$$

Thus the three-point expectation values on the sphere, the OPE coefficients with all indices lowered, and the matrix elements of general local operators are all the same thing. By a Möbius transformation (rescaling of z) we have also

$$\langle \mathscr{A}_i'(\infty,\infty)\mathscr{A}_k(z_1,\bar z_1)\mathscr{A}_j(0,0)\rangle_{S_2} = z_1^{h_i-h_k-h_j}\bar z_1^{\tilde h_i-\tilde h_k-\tilde h_j}c_{ikj} \ . \tag{6.7.15}$$

The four-point function translates into an operator expression

$$\langle \mathscr{A}_i'(\infty,\infty)\mathscr{A}_k(z_1,\bar z_1)\mathscr{A}_l(z_2,\bar z_2)\mathscr{A}_j(0,0)\rangle_{S_2} = \langle\!\langle i|\mathrm{T}[\hat{\mathscr{A}}_k(z_1,\bar z_1)\hat{\mathscr{A}}_l(z_2,\bar z_2)]|j\rangle\!\rangle \ , \tag{6.7.16}$$

where T denotes radial ordering. Let $|z_1| > |z_2|$ and insert a complete set of states,

$$1 = |m\rangle\mathscr{G}^{mn}\langle\!\langle n| \ . \tag{6.7.17}$$

The four-point amplitude (6.7.16) becomes

$$\sum_m z_1^{h_i-h_k-h_m}\bar z_1^{\tilde h_i-\tilde h_k-\tilde h_m}z_2^{h_m-h_l-h_j}\bar z_2^{\tilde h_m-\tilde h_l-\tilde h_j}c_{ikm}c^m{}_{lj} \ . \tag{6.7.18}$$

Thus, the operator product coefficients determine not only the three-point expectation values on the sphere, but also the four-point and, by the same construction, arbitrary n-point amplitudes. For $|z_1 - z_2| > |z_1|$, which overlaps the region $|z_1| > |z_2|$ where the expansion (6.7.18) is valid, we can translate $\mathscr{A}_k$ to the origin and give a similar expansion in terms of $c_{ilm}c^m{}_{kj}$. The equality of these two expansions is associativity of the OPE, figure 2.8.

Operator calculations

The Hilbert space expressions give us one more way to calculate expectations values. We take as an example the case of four exponential operators

$$\left\langle :e^{ik_4\cdot X(\infty,\infty)}: :e^{ik_1\cdot X(z_1,\bar z_1)}: :e^{ik_2\cdot X(z_2,\bar z_2)}: :e^{ik_3\cdot X(0,0)}: \right\rangle_{S_2}$$
$$= \langle\!\langle 0;k_4|\mathrm{T}[\,{}^\circ_\circ e^{ik_1\cdot X_1}{}^\circ_\circ\, {}^\circ_\circ e^{ik_2\cdot X_2}{}^\circ_\circ]|0;k_3\rangle\!\rangle \ . \tag{6.7.19}$$

We have used the result (2.7.11) that : : ordered operators are the same as $^\circ_\circ\ ^\circ_\circ$ ordered operators for this CFT. We have abbreviated $X^\mu(z_i,\bar z_i)$ as X_i^μ and to avoid clutter are omitting the hats on operators. By definition

$${}^\circ_\circ e^{ik\cdot X}{}^\circ_\circ = e^{ik\cdot X_C}e^{ik\cdot X_A} \ , \tag{6.7.20}$$

where

$$X_C^\mu(z,\bar z) = x^\mu - i\left(\frac{\alpha'}{2}\right)^{1/2}\sum_{m=1}^\infty \frac{1}{m}(\alpha_{-m}^\mu z^m + \tilde\alpha_{-m}^\mu \bar z^m) \ , \tag{6.7.21a}$$

$$X_A^\mu(z,\bar z) = -i\frac{\alpha'}{2}p^\mu \ln|z|^2 + i\left(\frac{\alpha'}{2}\right)^{1/2}\sum_{m=1}^\infty \frac{1}{m}\left(\frac{\alpha_m^\mu}{z^m} + \frac{\tilde\alpha_m^\mu}{\bar z^m}\right) \ . \tag{6.7.21b}$$

For $|z_1| > |z_2|$ the matrix element (6.7.19) becomes

$$\langle\!\langle 0;k_4|e^{ik_1\cdot X_{1C}}e^{ik_1\cdot X_{1A}}e^{ik_2\cdot X_{2C}}e^{ik_2\cdot X_{2A}}|0;k_3\rangle\!\rangle \ . \tag{6.7.22}$$

To evaluate this use the Campbell–Baker–Hausdorff (CBH) formula

$$e^{ik_1\cdot X_{1A}}\,e^{ik_2\cdot X_{2C}} = e^{ik_2\cdot X_{2C}}e^{ik_1\cdot X_{1A}}\,e^{-[k_1\cdot X_{1A},k_2\cdot X_{2C}]} = e^{ik_2\cdot X_{2C}}e^{ik_1\cdot X_{1A}}|z_{12}|^{\alpha' k_1\cdot k_2}\,. \tag{6.7.23}$$

The expectation value (6.7.22) becomes

$$\begin{aligned}&|z_{12}|^{\alpha' k_1\cdot k_2}\langle\!\langle 0;k_4|e^{ik_1\cdot X_{1C}+ik_2\cdot X_{2C}}\,e^{ik_1\cdot X_{1A}+ik_2\cdot X_{2A}}|0;k_3\rangle\\ &= |z_{12}|^{\alpha' k_1\cdot k_2}\langle\!\langle 0;k_4|e^{i(k_1+k_2)\cdot x}\,e^{\alpha'(k_1\ln|z_1|+k_2\ln|z_2|)\cdot p}|0;k_3\rangle\\ &= |z_{12}|^{\alpha' k_1\cdot k_2}|z_1|^{\alpha' k_1\cdot k_3}|z_2|^{\alpha' k_2\cdot k_3}\langle\!\langle 0;k_1+k_2+k_4|0;k_3\rangle\\ &= iC^X_{S_2}(2\pi)^d\delta^d(\textstyle\sum_i k_i)|z_{12}|^{\alpha' k_1\cdot k_2}|z_1|^{\alpha' k_1\cdot k_3}|z_2|^{\alpha' k_2\cdot k_3}\,,\end{aligned} \tag{6.7.24}$$

where we have used the two-point expectation value to normalize the last line. This is the familiar result (6.2.31), obtained by two other methods in section 6.2, after one includes in the latter a factor $|z_4|^{\alpha' k_4^2}$ from the change of frame and takes $z_4\to\infty$. All other free-field results can be obtained by this same oscillator method.

Relation between inner products

Nondegenerate bilinear and Hermitean inner products can always be related to one another by an appropriate antilinear operation on the bra. Let us consider an example. From the free-field expectation values, we have

$$\langle\!\langle 0;k|0;l\rangle = iC^X_{S_2}(2\pi)^d\delta^d(k+l)\,, \tag{6.7.25}$$

using the fact that $|0;k\rangle$ maps to $e^{ik\cdot X}$. Compare this with the inner product from the X^μ CFT,

$$\langle 0;k|0;l\rangle = (2\pi)^d\delta^d(l-k)\,. \tag{6.7.26}$$

These differ only by $k\to -k$ from conjugating $e^{ik\cdot X}$, and normalization,

$$\langle\!\langle 0;k| = iC^X_{S_2}\langle 0;-k|\,. \tag{6.7.27}$$

For more general operators, there is a natural notion of conjugation in CFT. In Euclidean quantum mechanics, Hermitean conjugation inverts Euclidean time, eq. (A.1.37), so the natural operation of Euclidean conjugation is conjugation × time-reversal: an operator that is Hermitean in Minkowski space is also Hermitean under this combined operation. In CFT, we make the same definition, but also must include a time-reversal on the conformal frame,

$$\overline{\mathscr{A}(p)} = \mathscr{A}'(p')^\dagger\,. \tag{6.7.28}$$

Here p and p' are related by radial time-reversal, $z' = \bar z^{-1}$, and the unprimed operator is in the z-frame and the primed operator in the u-

frame. To see how this works, consider a holomorphic operator of weight h, whose Laurent expansion is

$$\mathcal{O}(z) = i^h\sum_{n=-\infty}^{\infty}\frac{\mathcal{O}_n}{z^{n+h}}\,. \tag{6.7.29}$$

Its simple adjoint is

$$\mathcal{O}(z)^\dagger = i^{-h}\sum_{n=-\infty}^{\infty}\frac{\mathcal{O}_n^\dagger}{\bar z^{n+h}}\,. \tag{6.7.30}$$

Then the Euclidean adjoint is

$$\overline{\mathcal{O}(z)} = i^{-h}(-z^{-2})^h\sum_{n=-\infty}^{\infty}\frac{\mathcal{O}_n^\dagger}{z^{-n-h}} = i^h\sum_{n=-\infty}^{\infty}\frac{\mathcal{O}_{-n}^\dagger}{z^{n+h}}\,. \tag{6.7.31}$$

For an operator that is Hermitean in Minkowski time, $\mathcal{O}_{-n}^\dagger = \mathcal{O}_n$, and so this operator is also Hermitean[3] under $\bar{\mathcal{O}}$. The Euclidean adjoint (6.7.28) conjugates all *explicit* factors of i, but leaves z and $\bar z$ indices unchanged. This is its whole effect, other than an overall factor $(-1)^{N_a(N_a-1)/2}$ from reversing the order of anticommuting fields; here N_a is the total number of anticommuting fields in the operator.

This is the natural conformally invariant operation of conjugation, so it must be that

$$\langle\!\langle\overline{\mathscr{A}_i}| = K\langle\mathscr{A}_i| \tag{6.7.32}$$

for some constant K. For a direct demonstration, the best we have come up with is

$$\begin{aligned}\langle\!\langle\bar i\,|\,j\rangle &\equiv \langle\!\langle 1|\overline{\mathscr{A}'_i(\infty,\infty)}|j\rangle = K\langle 1|\overline{\mathscr{A}'_i(\infty,\infty)}|j\rangle\\ &= K\langle 1|\mathscr{A}_i(0,0)^\dagger|j\rangle = K\langle j|\mathscr{A}_i(0,0)|1\rangle^*\\ &= K\langle j|i\rangle^* = K\langle i|j\rangle\,.\end{aligned} \tag{6.7.33}$$

The only step here that is not either a definition or obvious is the assumption of proportionality, $\langle\!\langle 1| = K\langle 1|$. This must hold because $|1\rangle$ is the unique $SL(2,\mathbf{C})$-invariant state.

For the X CFT we have seen that $K^X = iC^X_{S_2}$. For the ghost CFT, the Laurent expansion of the amplitude (6.3.4) gives $\langle\!\langle 0|\tilde c_0 c_0|0\rangle = -C^{\rm g}_{S_2}$, where $|0\rangle = \tilde c_1 c_1|1\rangle$. The Hermitean inner product was defined in eq. (4.3.18) as $\langle 0|\tilde c_0 c_0|0\rangle = i$, so $K^{\rm g} = iC^{\rm g}_{S_2}$.

Except for the is, which are absent in a unitary CFT, one can set $K=1$ by adding an Euler number term to the action. In a Hermitean basis of operators, the two inner products are then identical, and the distinction

[3] Because its Laurent expansion has no factor of i, the Faddeev–Popov c ghost is actually anti-Hermitean under the Euclidean adjoint. This is due to an inconvenient conflict between conventions, and has no significance.

between them is often ignored. Notice, however, that a vertex operator of definite nonzero momentum cannot be Hermitean.

Again, this can all be extended to the open string, with the disk in place of the sphere.

Exercises

6.1 Verify that the expectation value (6.2.31) is smooth in the u-patch.

6.2 For the linear dilaton CFT, $\Phi = V_\mu X^\mu$, the X^μ zero-mode path integral diverges. This reflects the fact that the coupling is diverging in some direction. Consider therefore imaginary $\Phi = ib_\mu X^\mu$; this is unphysical but has technical applications (one could instead take complex momenta).
(a) Generalize the calculation that gave (6.2.17) to this case. Show that in the end the only changes are in the Weyl dependence (reflecting a change in the dimension of the exponential) and in the spacetime momentum conservation.
(b) Generalize the calculation of the same quantity using holomorphicity. Show that the result is the same as in the regular X^μ CFT but with an extra *background charge* operator $\exp(-2ib \cdot X)$ taken to infinity.

6.3 For one b and four c fields, show that expressions (6.3.5) and (6.3.8) are equal. Ignore the antiholomorphic fields.

6.4 (a) Write the amplitude for n open string tachyons on the disk as a generalization of eq. (6.4.5).
(b) Show by a change of variables that it is independent of the positions of the fixed vertex operators.

6.5 (a) Show that the residue of the pole in $I(s,t)$ at $\alpha' s = J - 1$ is a polynomial in $u - t$ of degree J, corresponding to intermediate particles of spin up to J.
(b) Consider the amplitude $I(s,t)$ without assuming $D = 26$. In the center-of-mass frame, write the residue of the pole at $\alpha' s = 1$ in terms of

$$k_1^i k_1^j P^J_{ij,kl} k_3^k k_3^l \ ,$$

where P^0 projects onto spin 0 (the unit matrix) and P^2 projects onto spin 2 (traceless symmetric tensors). Show that the coefficient of P^0 is positive for $D < 26$, zero at $D = 26$, and negative at $D > 26$. Compare with the expected string spectrum at this level, and with the discussion of this level in section 4.1.

6.6 (a) In the hard scattering limit, evaluate the integral in the Veneziano amplitude (6.4.5) using the saddle point approximation. Take the naive saddle point; the justification requires consideration of the analytic continuations needed to define the integral.
(b) Do the same for the Virasoro–Shapiro amplitude.

6.7 (a) Derive the amplitude (6.5.12).
(b) Verify the relation (6.5.14) between vertex operator normalizations. Show that this agrees with the state–operator mapping.

6.8 Derive the three-gauge-boson amplitude (6.5.15).

6.9 (a) Obtain the four-gauge-boson amplitude. To reduce the rather extensive algebra, consider only those terms in which the polarizations appear as $e_1 \cdot e_2\, e_3 \cdot e_4$.
(b) Compare the behavior at small $\alpha' k_i \cdot k_j$ with the same amplitude in Yang–Mills theory.

6.10 Carry out the integral in eq. (6.6.5) to obtain the Virasoro–Shapiro amplitude (6.6.10). The relation

$$|z|^{-2a} = \frac{1}{\Gamma(a)} \int_0^\infty dt\, t^{a-1} \exp(-tz\bar{z})$$

is useful. (Reference: Green, Schwarz, & Witten (1987), section 7.2; note that their d^2z is $\frac{1}{2}$ times ours.)

6.11 (a) Verify the amplitude (6.6.14).
(b) Verify the relation (6.6.15) between vertex operator normalizations.
(c) Complete the field theory calculation and verify the relation (6.6.18) between g_c and the gravitational coupling.

6.12 Calculate the disk amplitude with one closed string tachyon and two open string tachyons.

6.13 (a) Find the $PSL(2,\mathbf{C})$ transformation that takes three given points $z_{1,2,3}$ into chosen positions $\hat{z}_{1,2,3}$.
(b) Verify the Möbius invariance results (6.7.3)–(6.7.7). Show that to derive (6.7.5) it is sufficient that L_1 and $\tilde{L}_1$ annihilate the operators.

6.14 (a) Find the $PSL(2,\mathbf{R})$ transformation that takes three given points $y_{1,2,3}$ into chosen positions $\hat{y}_{1,2,3}$.
(b) Generalize the Möbius invariance results (6.7.3)–(6.7.7) to operators on the boundary of a disk.
(c) Find similar results for the disk with one interior, two interior, and one interior and one boundary operators.

6.15 (a) Generalize the operator calculation (6.7.24) to a product of n exponentials.
(b) Generalize it to the expectation value (6.2.25).

7

One-loop amplitudes

After describing the relevant surfaces, we focus on the torus, first on the CFTs on this surface and then on the string amplitudes. We then generalize to the open and unoriented theories. The most important issue to be understood is the absence of short-distance (UV) divergences.

7.1 Riemann surfaces

There are four Riemann surfaces with Euler number zero.

The torus

The torus T^2, discussed in section 5.1, is the only closed oriented surface with Euler number zero. We describe it as the complex plane with metric $ds^2 = dwd\bar{w}$ and identifications

$$w \cong w + 2\pi \cong w + 2\pi\tau \,. \tag{7.1.1}$$

There are two moduli, the real and imaginary parts of $\tau = \tau_1 + i\tau_2$, and two CKVs, the translations. In terms of the real coordinates $w = \sigma^1 + i\sigma^2$,

$$(\sigma^1, \sigma^2) \cong (\sigma^1 + 2\pi, \sigma^2) \cong (\sigma^1 + 2\pi\tau_1, \sigma^2 + 2\pi\tau_2) \,, \tag{7.1.2}$$

so that one can think of the torus as a cylinder of circumference 2π and length $2\pi\tau_2$ with the ends rotated by an angle of $2\pi\tau_1$ and then sewn together.

In terms of the coordinate $z = \exp(-iw)$, the identification $w \cong w + 2\pi$ is automatic, while $w \cong w + 2\pi\tau$ becomes

$$z \cong z\exp(-2\pi i\tau) \,. \tag{7.1.3}$$

A fundamental region is the annulus

$$1 \le |z| \le \exp(2\pi\tau_2) \,. \tag{7.1.4}$$

The torus is formed by rotating the outer circle by $2\pi\tau_1$ and then sewing the inner and outer circles together. We will use the w-coordinate unless otherwise noted.

The cylinder (annulus)

The cylinder C_2 is the region

$$0 \le \mathrm{Re}\, w \le \pi \,, \quad w \cong w + 2\pi i t \,. \tag{7.1.5}$$

That is, it is a strip of width π and length $2\pi t$, with the ends joined. There is a single modulus t, which runs over the whole range $0 < t < \infty$. Unlike the torus there is no modular group, the long-cylinder limit $t \to 0$ being quite different from the long-strip limit $t \to \infty$. There is a single CKV, the translation parallel to the boundary.

The cylinder can be obtained from the torus with imaginary $\tau = it$ by identifying under the involution

$$w' = -\bar{w} \,, \tag{7.1.6}$$

which is a reflection through the imaginary axis. The lines $\sigma^1 = 0, \pi$ are fixed by this reflection and so become boundaries. The other coordinate is periodically identified,

$$(\sigma^1, 0) \cong (\sigma^1, 2\pi t) \,. \tag{7.1.7}$$

The Klein bottle

The Klein bottle K_2 can be regarded as the complex plane with identification

$$w \cong w + 2\pi \cong -\bar{w} + 2\pi i t \tag{7.1.8}$$

or

$$(\sigma^1, \sigma^2) \cong (\sigma^1 + 2\pi, \sigma^2) \cong (-\sigma^1, \sigma^2 + 2\pi t) \,. \tag{7.1.9}$$

This is a cylinder of circumference 2π and length $2\pi\tau_2$, with the ends sewn together after a parity-reversal Ω. The single modulus t runs over $0 < t < \infty$, and there is no modular group. Translation in the σ^2 direction is the only CKV. The Klein bottle can be obtained from the torus with modulus $\tau = 2it$ by identifying under

$$w' = -\bar{w} + 2\pi i t \,. \tag{7.1.10}$$

The Klein bottle can be thought of as a sphere with two cross-caps, but we postpone this description until section 7.4.

The Möbius strip

The Möbius strip M_2 is sewn from a strip with a twist by Ω,

$$0 \le \mathrm{Re}\, w \le \pi\,, \quad w \cong -\bar{w} + \pi + 2\pi i t\,. \tag{7.1.11}$$

Again the modulus t runs over $0 < t < \infty$ and the only CKV is the σ^2-translation. It can be obtained from the torus with $\tau = 2it$ by identifying under the *two* involutions

$$w' = -\bar{w} \quad \text{and} \quad w' = w + \pi(2it + 1)\,. \tag{7.1.12}$$

The Möbius strip can be thought of as a disk with a cross-cap, but again we defer this description.

7.2 CFT on the torus

Scalar correlators

As in the case of the sphere, we start with the Green's function (6.2.7), satisfying

$$\frac{2}{\alpha'}\bar{\partial}\partial G'(w,\bar{w};w',\bar{w}') = -2\pi\delta^2(w-w') + \frac{1}{4\pi\tau_2}\,. \tag{7.2.1}$$

The Green's function is periodic in both directions on the torus, and aside from the source and the background charge term it would be the sum of holomorphic and antiholomorphic functions. These properties identify it as being associated with the theta functions, whose properties are given at the end of this section. In particular we guess

$$G'(w,\bar{w};w',\bar{w}') \sim -\frac{\alpha'}{2}\ln\left|\vartheta_1\left(\frac{w-w'}{2\pi}\middle|\tau\right)\right|^2\,. \tag{7.2.2}$$

The function ϑ_1 vanishes linearly when its first argument goes to zero, giving the correct behavior as $w \to w'$. Elsewhere the logarithm is the sum of holomorphic and antiholomorphic functions and so is annihilated by $\bar{\partial}\partial$. However, this function is not quite doubly periodic owing to the quasiperiodicity (7.2.32b) — it changes by $-\alpha'[\mathrm{Im}(w-w') + \pi\tau_2]$ under $w \to w + 2\pi\tau$. Also, the background charge is missing. Both properties are easily corrected:

$$G'(w,\bar{w};w',\bar{w}') = -\frac{\alpha'}{2}\ln\left|\vartheta_1\left(\frac{w-w'}{2\pi}\middle|\tau\right)\right|^2 + \alpha'\frac{[\mathrm{Im}(w-w')]^2}{4\pi\tau_2} + k(\tau,\bar{\tau})\,. \tag{7.2.3}$$

The function $k(\tau,\bar{\tau})$ is determined by orthogonality to X_0, but as in the case of the sphere it drops out due to spacetime momentum conservation.

The expectation value of a product of vertex operators is given, in parallel to the earlier result (6.2.13) on the sphere, by

$$\left\langle \prod_{i=1}^{n} :e^{ik_i\cdot X(z_i,\bar{z}_i)}: \right\rangle_{T^2} = iC^X_{T^2}(\tau)(2\pi)^d\delta^d(\textstyle\sum_i k_i) \times \prod_{i<j}\left|\frac{2\pi}{\partial_v\vartheta_1(0|\tau)}\vartheta_1\left(\frac{w_{ij}}{2\pi}\middle|\tau\right)\exp\left[-\frac{(\mathrm{Im}\, w_{ij})^2}{4\pi\tau_2}\right]\right|^{\alpha' k_i\cdot k_j}\,. \tag{7.2.4}$$

The factor $2\pi/\partial_v\vartheta_1$ comes from the renormalized self-contractions.

The scalar partition function

The overall normalization of the sphere was absorbed into the string coupling constant, but we cannot do this for the torus, in particular because there is a nontrivial τ-dependence. In fact, we will see that there is a great deal of physics in the amplitude with no vertex operators at all. Consider the path integral with no vertex operators, $\langle 1\rangle_{T^2(\tau)} \equiv Z(\tau)$. We can think of the torus with modulus τ as formed by taking a field theory on a circle, evolving for Euclidean time $2\pi\tau_2$, translating in σ^1 by $2\pi\tau_1$, and then identifying the ends. In operator language this gives a trace,

$$\begin{aligned} Z(\tau) &= \mathrm{Tr}\left[\exp(2\pi i\tau_1 P - 2\pi\tau_2 H)\right] \\ &= (q\bar{q})^{-d/24}\mathrm{Tr}\left(q^{L_0}\bar{q}^{\tilde{L}_0}\right)\,. \end{aligned} \tag{7.2.5}$$

Here $q = \exp(2\pi i\tau)$, the momentum $P = L_0 - \tilde{L}_0$ generates translations of σ^1, and the Hamiltonian $H = L_0 + \tilde{L}_0 - \frac{1}{24}(c+\tilde{c})$ generates translations of σ^2 as in eq. (2.6.10). Such a trace, weighted by the exponential of the Hamiltonian and other conserved quantities, is termed a *partition function* as in statistical mechanics.

The trace breaks up into a sum over occupation numbers $N_{\mu n}$ and $\tilde{N}_{\mu n}$ for each μ and n and an integral over momentum k^μ. The Virasoro generators similarly break up into a sum, and $Z(\tau)$ becomes

$$V_d(q\bar{q})^{-d/24}\int\frac{d^dk}{(2\pi)^d}\exp(-\pi\tau_2\alpha' k^2)\prod_{\mu,n}\sum_{N_{\mu n},\tilde{N}_{\mu n}=0}^{\infty} q^{nN_{\mu n}}\bar{q}^{n\tilde{N}_{\mu n}}\,. \tag{7.2.6}$$

The factor of spacetime volume V_d comes from the continuum normalization of the momentum, $\sum_k$ becoming $V_d(2\pi)^{-d}\int d^dk$. The various sums are geometric,

$$\sum_{N=0}^{\infty} q^{nN} = (1-q^n)^{-1}\,, \tag{7.2.7}$$

and so we obtain

$$Z(\tau) = iV_d Z_X(\tau)^d \,, \tag{7.2.8}$$

where

$$Z_X(\tau) = (4\pi^2\alpha'\tau_2)^{-1/2}|\eta(\tau)|^{-2} \,. \tag{7.2.9}$$

Here

$$\eta(\tau) = q^{1/24}\prod_{n=1}^{\infty}(1-q^n) \tag{7.2.10}$$

is the Dedekind eta function, discussed further at the end of the section. The i comes from the rotation $k^0 \to ik^d$ needed to define the divergent integral. Also, the earlier $C^X_{T^2}(\tau)$ is $Z_X(\tau)^d$.

The result (7.2.8) is obviously invariant under $\tau \to \tau + 1$. Invariance of Z_X under $\tau \to -1/\tau$ follows from eq. (7.2.44), $\eta(-1/\tau) = (-i\tau)^{1/2}\eta(\tau)$, thus generating the full modular group. Similarly the expectation value with vertex operators is modular-covariant, taking into account the Weyl transformations of the operators.

We have evaluated the path integral by translating into an operator expression, but as on the sphere there are two other methods. We leave to the reader the direct path integral evaluation. Fourier transforming in σ^1, it just breaks up into a product of harmonic oscillator path integrals.

The evaluation using holomorphicity is less direct. We include it for illustration because it is used at higher genus and for more complicated CFTs. The idea is to get a differential equation with respect to τ. Consider the torus with modulus τ and make a small change in the metric, $\delta g_{ww} = \epsilon^*$. The new metric is

$$\begin{aligned} ds^2 &= dwd\bar{w} + \epsilon^* dw^2 + \epsilon d\bar{w}^2 \\ &= (1+\epsilon^*+\epsilon)\, d[w + \epsilon(\bar{w}-w)]\, d[\bar{w} + \epsilon^*(w-\bar{w})] + O(\epsilon^2) \,. \end{aligned} \tag{7.2.11}$$

That is, the metric is Weyl-equivalent to $dw'd\bar{w}'$ with $w' = w + \epsilon(\bar{w} - w)$. This has the periodicities

$$w' \cong w' + 2\pi \cong w' + 2\pi(\tau - 2i\tau_2\epsilon) \,. \tag{7.2.12}$$

This change in the metric is therefore equivalent to a change

$$\delta\tau = -2i\tau_2\epsilon \tag{7.2.13}$$

in the modulus.

Under a change in the metric, the change in the path integral is

$$\begin{aligned} \delta Z(\tau) &= -\frac{1}{2\pi}\int d^2w \left[\delta g_{\bar{w}\bar{w}} \langle\, T_{ww}(w)\,\rangle + \delta g_{ww} \langle\, T_{\bar{w}\bar{w}}(\bar{w})\,\rangle\right] \\ &= -2\pi i\left[\delta\tau \langle\, T_{ww}(0)\,\rangle - \delta\bar{\tau} \langle\, T_{\bar{w}\bar{w}}(0)\,\rangle\right] \,. \end{aligned} \tag{7.2.14}$$

In the second line we have used translation invariance to carry out the integral over w. All expectation values are understood to be on the torus with modulus τ. To get the expectation value of the energy-momentum tensor, we use the OPE

$$\partial_w X^\mu(w)\partial_w X_\mu(0) = -\frac{\alpha' d}{2w^2} - \alpha' T_{ww}(0) + O(w^2) \,. \tag{7.2.15}$$

Now,

$$\begin{aligned} Z(\tau)^{-1} \langle\, \partial_w X^\mu(w)\partial_w X_\mu(0)\,\rangle &= d\partial_w\partial_{w'} G'(w,\bar{w};w',\bar{w}')|_{w'=0} \\ &= \frac{\alpha' d}{2}\frac{\vartheta_1\partial_w^2\vartheta_1 - \partial_w\vartheta_1\partial_w\vartheta_1}{\vartheta_1^2} + \frac{\alpha' d}{8\pi\tau_2} \,, \end{aligned} \tag{7.2.16}$$

where the arguments of all the theta functions are $(w/2\pi, \tau)$. This does indeed have a double pole at $w = 0$. Carefully expanding both numerator and denominator to order w^2 (this is simplified by the fact that ϑ_1 is odd), the order w^0 term in the expectation value (7.2.16) is

$$\frac{\alpha' d}{6}\frac{\partial_w^3\vartheta_1}{\partial_w\vartheta_1} + \frac{\alpha' d}{8\pi\tau_2} \,. \tag{7.2.17}$$

The OPE (7.2.15) then gives

$$\langle\, T_{ww}(0)\,\rangle = \left(-\frac{d}{6}\frac{\partial_w^3\vartheta_1(0|\tau)}{\partial_w\vartheta_1(0|\tau)} - \frac{d}{8\pi\tau_2}\right) Z(\tau) \,, \tag{7.2.18}$$

and the variation (7.2.14) gives the differential equation

$$\partial_\tau \ln Z(\tau) = \frac{\pi i d}{3}\frac{\partial_w^3\vartheta_1(0|\tau)}{\partial_w\vartheta_1(0|\tau)} + \frac{id}{4\tau_2} \,. \tag{7.2.19}$$

To proceed further, we use the fact, which is easily verified from the infinite sum, that the theta functions satisfy

$$\partial_w^2\vartheta_1\left(\frac{w}{2\pi}\middle|\tau\right) = \frac{i}{\pi}\partial_\tau\vartheta_1\left(\frac{w}{2\pi}\middle|\tau\right) \,. \tag{7.2.20}$$

Rewrite (7.2.19) as

$$\partial_\tau \ln Z(\tau) = -\frac{d}{3}\partial_\tau \ln \partial_w\vartheta_1(0|\tau) + \frac{id}{4\tau_2} \,. \tag{7.2.21}$$

Together with the conjugate equation, this determines

$$Z(\tau) = |\partial_w\vartheta_1(0|\tau)|^{-2d/3}\tau_2^{-d/2} \tag{7.2.22}$$

up to a numerical coefficient that must be fixed by other means. Noting the relation (7.2.43), this is the same as the earlier result (7.2.8).

The bc CFT

For the ghost CFT, the partition function is again determined easily in terms of a trace over states. There are two sets of raising operators b_{-n} and c_{-n} on both the right and left, but now these are anticommuting so the occupation numbers can only be zero or one — a single oscillator at mode n thus gives $(1+q^n)$. In addition there are four ground states, giving

$$\begin{aligned}\mathrm{Tr}\left[\exp(2\pi i\tau_1 P - 2\pi\tau_2 H)\right] &= (q\bar{q})^{13/12}\mathrm{Tr}\left(q^{L_0}\bar{q}^{\tilde{L}_0}\right) \\ &= 4(q\bar{q})^{1/12}\prod_{n=1}^{\infty}|1+q^n|^4 \,. \end{aligned} \tag{7.2.23}$$

Recall, however, that for anticommuting fields the simple trace corresponds to a path integral with antiperiodic boundary conditions in the time direction. In order to calculate the Faddeev–Popov determinant, the ghosts must have the same periodicity as the original coordinate transformations, which were periodic. Thus we need

$$\begin{aligned} Z(\tau) &= \mathrm{Tr}\left[(-1)^F \exp(2\pi i\tau_1 P - 2\pi\tau_2 H)\right] \\ &= 0 \,, \end{aligned} \tag{7.2.24}$$

where $(-1)^F$ anticommutes with all the ghost fields. The trace vanishes because the states $|\downarrow\downarrow\rangle$ and $|\uparrow\uparrow\rangle$ have opposite ghost number mod 2 from $|\uparrow\downarrow\rangle$ and $|\downarrow\uparrow\rangle$. From the path integral point of view, $Z(\tau)$ vanishes due to ghost zero modes from the moduli and CKVs. These must be saturated with appropriate insertions, the simplest nonvanishing amplitude (and the one we will need) being

$$\left\langle c(w_1)b(w_2)\tilde{c}(\bar{w}_3)\tilde{b}(\bar{w}_4)\right\rangle \,. \tag{7.2.25}$$

In the operator calculation we write each field in terms of its mode expansion. Only the $n=0$ terms contribute, the expectation values of the others vanishing as above. Then (7.2.25) becomes

$$\mathrm{Tr}\left[(-1)^F c_0 b_0 \tilde{c}_0 \tilde{b}_0 \exp(2\pi i\tau_1 P - 2\pi\tau_2 H)\right] \,. \tag{7.2.26}$$

The operator $c_0 b_0 \tilde{c}_0 \tilde{b}_0$ projects onto the single ground state $|\downarrow\downarrow\rangle$, so the result is

$$(q\bar{q})^{1/12}\prod_{n=1}^{\infty}|1-q^n|^4 = |\eta(\tau)|^4 \,. \tag{7.2.27}$$

Note the sign change in the infinite product, from the $(-1)^F$ in the trace. The result (7.2.27) is independent of the positions of the ghost fields because the CKV and quadratic differentials on the torus are constants.

Again, we do not try to keep track of the overall sign, setting it to be positive.

General CFTs

For general CFTs,

$$Z(\tau) = \sum_i q^{h_i - c/24}\bar{q}^{\tilde{h}_i - \tilde{c}/24}(-1)^{F_i} \,, \tag{7.2.28}$$

where i runs over all states of the CFT, and F_i is again the world-sheet fermion number. Invariance under $\tau \to \tau + 1$ requires $h_i - \tilde{h}_i - \frac{1}{24}(c - \tilde{c})$ to be an integer. Looking at the unit operator, this requires $c - \tilde{c}$ to be a multiple of 24. For general operators one then has the requirement that the spin be an integer,

$$h_i - \tilde{h}_i \in \mathbf{Z} \,. \tag{7.2.29}$$

Incidentally, CFTs for which $c - \tilde{c}$ is not a multiple of 24 are certainly of interest. These cannot be separately modular-invariant, but with appropriate projections can be combined into a modular-invariant theory.

Invariance under $\tau \to -1/\tau$ puts a further constraint on the spectrum. Let us here note just one general consequence, which relates the states of large h to those of lowest h. Consider the partition function for $\tau = i\ell$ as $\ell \to 0$. The convergence factor $q = \exp(-2\pi\ell)$ is approaching unity, so the partition function is determined by the density of states at high weight. Using modular invariance, this is equal to the partition function at $\tau = i/\ell$. In the latter case, $q = \exp(-2\pi/\ell)$ is going to zero, so the state of lowest weight dominates the sum. This is the unit state, with $L_0 = \tilde{L}_0 = 0$, giving

$$Z(i\ell) \overset{\ell\to 0}{\approx} \exp\left[\frac{\pi(c+\tilde{c})}{12\ell}\right] \,. \tag{7.2.30}$$

The density of states at high weight is thus determined by the central charge. This generalizes the free-boson result, where $c = \tilde{c} = d$ counts the number of free bosons.[1]

[1] To be precise, the result (7.2.30) holds in compact unitary CFTs. In nonunitary theories there are usually operators of negative weight, whose contribution will dominate that of the unit operator. In the linear dilaton theory, the central charge depends on V^μ but the density of states does not: here the exception occurs because of complications due to the noncompactness of X^μ combined with lack of translation invariance.

Theta functions

The basic *theta function* is

$$\vartheta(v,\tau) = \sum_{n=-\infty}^{\infty} \exp(\pi i n^2\tau + 2\pi i n v) . \tag{7.2.31}$$

It has the periodicity properties

$$\vartheta(v+1,\tau) = \vartheta(v,\tau) , \tag{7.2.32a}$$
$$\vartheta(v+\tau,\tau) = \exp(-\pi i\tau - 2\pi i v)\vartheta(v,\tau) , \tag{7.2.32b}$$

and these in fact determine it up to normalization. Under modular transformations,

$$\vartheta(v,\tau+1) = \vartheta(v+1/2,\tau) , \tag{7.2.33a}$$
$$\vartheta(v/\tau,-1/\tau) = (-i\tau)^{1/2}\exp(\pi i v^2/\tau)\vartheta(v,\tau) . \tag{7.2.33b}$$

The theta function has a unique zero, up to the periodicity (7.2.32), at $v = \frac{1}{2}(1+\tau)$. It can also be written as an infinite product,

$$\vartheta(v,\tau) = \prod_{m=1}^{\infty}(1-q^m)(1+zq^{m-1/2})(1+z^{-1}q^{m-1/2}) , \tag{7.2.34}$$

in terms of

$$q = \exp(2\pi i\tau), \qquad z = \exp(2\pi i v) . \tag{7.2.35}$$

We often need the asymptotic behavior of the theta function as $q \to 0$ or as $q \to 1$. The $q \to 0$ behavior can be read immediately from either the infinite sum or infinite product form. The $q \to 1$ behavior is not manifest in either form, because an infinite number of terms contribute. It can be obtained from the $\tau \to -1/\tau$ modular transformation, which relates these two limits by $q \to \exp(4\pi^2/\ln q)$.

It is also useful to define the *theta function with characteristics,*

$$\vartheta\begin{bmatrix} a \\ b \end{bmatrix}(v,\tau) = \exp\left[\pi i a^2\tau + 2\pi i a(v+b)\right]\vartheta(v+a\tau+b,\tau)$$
$$= \sum_{n=-\infty}^{\infty}\exp\left[\pi i(n+a)^2\tau + 2\pi i(n+a)(v+b)\right] . \tag{7.2.36}$$

Other common notations are

$$\vartheta_{00}(v,\tau) = \vartheta_3(v|\tau) = \vartheta\begin{bmatrix} 0 \\ 0 \end{bmatrix}(v,\tau) = \sum_{n=-\infty}^{\infty} q^{n^2/2}z^n , \tag{7.2.37a}$$

$$\vartheta_{01}(v,\tau) = \vartheta_4(v|\tau) = \vartheta\begin{bmatrix} 0 \\ 1/2 \end{bmatrix}(v,\tau) = \sum_{n=-\infty}^{\infty}(-1)^n q^{n^2/2}z^n , \tag{7.2.37b}$$

$$\vartheta_{10}(v,\tau) = \vartheta_2(v|\tau) = \vartheta\begin{bmatrix} 1/2 \\ 0 \end{bmatrix}(v,\tau) = \sum_{n=-\infty}^{\infty} q^{(n-1/2)^2/2}z^{n-1/2} , \tag{7.2.37c}$$

$$\vartheta_{11}(v,\tau) = -\vartheta_1(v|\tau) = \vartheta\begin{bmatrix} 1/2 \\ 1/2 \end{bmatrix}(v,\tau)$$
$$= -i\sum_{n=-\infty}^{\infty}(-1)^n q^{(n-1/2)^2/2}z^{n-1/2} . \tag{7.2.37d}$$

These have the product representations

$$\vartheta_{00}(v,\tau) = \prod_{m=1}^{\infty}(1-q^m)(1+zq^{m-1/2})(1+z^{-1}q^{m-1/2}) , \tag{7.2.38a}$$

$$\vartheta_{01}(v,\tau) = \prod_{m=1}^{\infty}(1-q^m)(1-zq^{m-1/2})(1-z^{-1}q^{m-1/2}) , \tag{7.2.38b}$$

$$\vartheta_{10}(v,\tau) = 2\exp(\pi i\tau/4)\cos\pi v\prod_{m=1}^{\infty}(1-q^m)(1+zq^m)(1+z^{-1}q^m) , \tag{7.2.38c}$$

$$\vartheta_{11}(v,\tau) = -2\exp(\pi i\tau/4)\sin\pi v\prod_{m=1}^{\infty}(1-q^m)(1-zq^m)(1-z^{-1}q^m) . \tag{7.2.38d}$$

Their modular transformations are

$$\vartheta_{00}(v,\tau+1) = \vartheta_{01}(v,\tau) , \tag{7.2.39a}$$
$$\vartheta_{01}(v,\tau+1) = \vartheta_{00}(v,\tau) , \tag{7.2.39b}$$
$$\vartheta_{10}(v,\tau+1) = \exp(\pi i/4)\vartheta_{10}(v,\tau) , \tag{7.2.39c}$$
$$\vartheta_{11}(v,\tau+1) = \exp(\pi i/4)\vartheta_{11}(v,\tau) , \tag{7.2.39d}$$

and

$$\vartheta_{00}(v/\tau,-1/\tau) = (-i\tau)^{1/2}\exp(\pi i v^2/\tau)\vartheta_{00}(v,\tau) , \tag{7.2.40a}$$
$$\vartheta_{01}(v/\tau,-1/\tau) = (-i\tau)^{1/2}\exp(\pi i v^2/\tau)\vartheta_{10}(v,\tau) , \tag{7.2.40b}$$
$$\vartheta_{10}(v/\tau,-1/\tau) = (-i\tau)^{1/2}\exp(\pi i v^2/\tau)\vartheta_{01}(v,\tau) , \tag{7.2.40c}$$
$$\vartheta_{11}(v/\tau,-1/\tau) = -(-i\tau)^{1/2}\exp(\pi i v^2/\tau)\vartheta_{11}(v,\tau) . \tag{7.2.40d}$$

We will encounter these functions more extensively in the superstring.

The theta function satisfies Jacobi's 'abstruse identity,' a special case of a quartic identity of Riemann,

$$\vartheta_{00}^4(0,\tau) - \vartheta_{01}^4(0,\tau) - \vartheta_{10}^4(0,\tau) = 0 . \tag{7.2.41}$$

Note also that

$$\vartheta_{11}(0,\tau) = 0 . \tag{7.2.42}$$

Finally, the Dedekind eta function is

$$\eta(\tau) = q^{1/24}\prod_{m=1}^{\infty}(1-q^m) = \left[\frac{\partial_v \vartheta_{11}(0,\tau)}{-2\pi}\right]^{1/3} . \tag{7.2.43}$$

It has the modular transformations

$$\eta(\tau+1) = \exp(i\pi/12)\,\eta(\tau) , \tag{7.2.44a}$$

$$\eta(-1/\tau) = (-i\tau)^{1/2}\,\eta(\tau) . \tag{7.2.44b}$$

7.3 The torus amplitude

We now apply to the torus the general result (5.3.9) expressing the string amplitude as an integral with ghost and vertex operator insertions. The two CKVs require one vertex operator to be fixed, so

$$S_{T^2}(1;2;\ldots;n) = \frac{1}{2}\int_{F_0} d\tau d\bar\tau \left\langle B\tilde B\, \tilde c c \mathscr{V}_1(w_1,\bar w_1)\prod_{i=2}^{n}\int dw_i d\bar w_i\, \mathscr{V}_i(w_i,\bar w_i)\right\rangle_{T^2} . \tag{7.3.1}$$

The fundamental region F_0 and the factor of $\frac{1}{2}$ from $w \to -w$ were discussed in section 5.1. As usual for complex variables, $d\tau d\bar\tau = 2d\tau_1 d\tau_2$, where $\tau = \tau_1 + i\tau_2$. The ghost insertion for $d\tau$ is

$$\begin{aligned} B = \frac{1}{4\pi}(b,\partial_\tau g) &= \frac{1}{2\pi}\int d^2w\, b_{ww}(w)\partial_\tau g_{\bar w\bar w} \\ &= \frac{i}{4\pi\tau_2}\int d^2w\, b_{ww}(w) \\ &\to 2\pi i b_{ww}(0) . \end{aligned} \tag{7.3.2}$$

We have used the result (7.2.13) for the τ-derivative of the metric. In the final line we have used the fact that the ghost path integral (7.2.27) is independent of position to carry out the w-integration. (Recall that $d^2w = 2d\sigma^1 d\sigma^2$ and that $g_{w\bar w} = \frac{1}{2}$.)

The CKG consists of the translations of the torus. Since this group has finite volume, we need not fix it: we can rewrite the amplitude so that all vertex operators are integrated and the volume of the CKG is divided out. It is easy to do this by hand here. The CKVs are constant, so the expectation value in eq. (7.3.1) is independent of where the c ghosts are placed: we can move them away from w_1, putting them at some fixed position. For the vertex operators, translation invariance implies that the amplitude is unchanged if we replace $w_i \to w_i + w$. Average over translations

$$\int \frac{dwd\bar w}{2(2\pi)^2\tau_2} , \tag{7.3.3}$$

the denominator being the area of the torus, which is the volume of the CKG. Using eqs. (7.3.2) and (7.3.3), the amplitude becomes

$$S_{T^2}(1;2;\ldots;n) = \int_{F_0}\frac{d\tau d\bar\tau}{4\tau_2}\left\langle b(0)\tilde b(0)\tilde c(0)c(0)\prod_{i=1}^{n}\int dw_i d\bar w_i\, \mathscr{V}_i(w_i,\bar w_i)\right\rangle_{T^2} . \tag{7.3.4}$$

All vertex operators are now on an equal footing.

The form (7.3.4) could also have been derived directly, without the intermediate step of fixing a vertex operator, and is valid even without vertex operators,

$$Z_{T^2} = \int_{F_0}\frac{d\tau d\bar\tau}{4\tau_2}\left\langle b(0)\tilde b(0)\tilde c(0)c(0)\right\rangle_{T^2} . \tag{7.3.5}$$

This vacuum amplitude is quite interesting and will be our main focus. Not only does it reveal the essential difference between the ultraviolet behaviors of string theory and field theory, but it has an important physical interpretation of its own.

For 26 flat dimensions, the matter and ghost path integrals (7.3.5) were evaluated in the previous section, with the result

$$Z_{T^2} = iV_{26}\int_{F_0}\frac{d\tau d\bar\tau}{4\tau_2}(4\pi^2\alpha'\tau_2)^{-13}|\eta(\tau)|^{-48} . \tag{7.3.6}$$

The amplitude has the important property of modular invariance. The product $\tau_2|\eta(\tau)|^4$ is invariant by eq. (7.2.44), and it is easy to check that

$$\frac{d\tau d\bar\tau}{\tau_2^2} \tag{7.3.7}$$

is modular-invariant as well. Note the exponent -48, from the 24 left- and right-moving oscillators. The contribution of the ghosts cancels two sets of oscillators, leaving only the contribution of the transverse modes. The amplitude with n tachyons is given by including the vertex operator expectation value (7.2.4) and integrating over positions.

For a general CFT the path integral on the torus is expressed as a trace in eq. (7.2.28). As long as there are $d \geq 2$ noncompact flat dimensions the ghosts still cancel two sets of bosonic operators, and the vacuum amplitude is

$$Z_{T^2} = V_d\int_{F_0}\frac{d\tau d\bar\tau}{4\tau_2}\int\frac{d^dk}{(2\pi)^d}\exp(-\pi\tau_2\alpha' k^2)\sum_{i\in\mathscr{H}^\perp} q^{h_i-1}\bar q^{\tilde h_i-1} \tag{7.3.8a}$$

$$= iV_d\int_{F_0}\frac{d\tau d\bar\tau}{4\tau_2}(4\pi^2\alpha'\tau_2)^{-d/2}\sum_{i\in\mathscr{H}^\perp} q^{h_i-1}\bar q^{\tilde h_i-1} . \tag{7.3.8b}$$

Here $\mathscr{H}^\perp$ is the closed string Hilbert space excluding the ghosts, the $\mu = 0, 1$ oscillators, and the noncompact momenta.

In order to understand the physics of these amplitudes it is useful to compare with the corresponding quantity in field theory, the sum over all particle paths with the topology of a circle. This is given by

$$Z_{S_1}(m^2) = V_d \int \frac{d^d k}{(2\pi)^d} \int_0^\infty \frac{dl}{2l} \exp[-(k^2 + m^2)l/2]$$
$$= iV_d \int_0^\infty \frac{dl}{2l} (2\pi l)^{-d/2} \exp(-m^2 l/2) \,. \tag{7.3.9}$$

This can be derived by gauge-fixing the point-particle path integral, which was given as exercise 5.2. The result is intuitive: l is the modulus for the circle, $\frac{1}{2}(k^2 + m^2)$ is the world-line Hamiltonian, and the $2l$ in the denominator removes the overcounting from translation and reversal of the world-line coordinate.

Let us now take this point-particle result, which is for a particle of given m^2, and sum over the actual spectrum of the string. As shown in chapter 4, the physical spectrum is in one-to-one correspondence with $\mathscr{H}^\perp$, where the mass is related to the transverse weights by

$$m^2 = \frac{2}{\alpha'}(h + \tilde{h} - 2) \tag{7.3.10}$$

and with the constraint $h = \tilde{h}$. It is useful to write the latter constraint in integral form,

$$\delta_{h,\tilde{h}} = \int_{-\pi}^{\pi} \frac{d\theta}{2\pi} \exp\left[i(h - \tilde{h})\theta\right] \,. \tag{7.3.11}$$

In writing this we are assuming that the difference $h - \tilde{h}$ is an integer. This is true in 26 flat dimensions, and as we have discussed in the previous section it is necessary in general for modular invariance. Then

$$\sum_{i \in \mathscr{H}^\perp} Z_{S_1}(m_i^2) = iV_d \int_0^\infty \frac{dl}{2l} \int_{-\pi}^{\pi} \frac{d\theta}{2\pi} (2\pi l)^{-d/2}$$
$$\times \sum_{i \in \mathscr{H}^\perp} \exp\left[-(h_i + \tilde{h}_i - 2)l/\alpha' + i(h_i - \tilde{h}_i)\theta\right]$$
$$= iV_d \int_R \frac{d\tau d\bar{\tau}}{4\tau_2} (4\pi^2 \alpha' \tau_2)^{-d/2} \sum_{i \in \mathscr{H}^\perp} q^{h_i - 1} \bar{q}^{\tilde{h}_i - 1} \,, \tag{7.3.12}$$

where $\theta + il/\alpha' = 2\pi\tau$. Here the region of integration R is

$$R: \quad \tau_2 > 0 \,, \quad |\tau_1| < \frac{1}{2} \,. \tag{7.3.13}$$

Now let us interpret these results. The amplitude (7.3.9) for a single point particle diverges as $l \to 0$. This is the usual UV divergence of quantum field theory (note that l has units of spacetime length-squared). Summing this over the string spectrum as in eq. (7.3.12) only makes things worse, as all the states of the string contribute with the same sign. However, compare the sum (7.3.12) to the actual string amplitude (7.3.8). There is great similarity but also one remarkable difference. The integrands are identical but the regions of integration are different: in the string it is the fundamental region

$$F_0: \quad |\tau| > 1 \,, \quad |\tau_1| < \frac{1}{2} \,, \tag{7.3.14}$$

and in field theory the larger region R. In the string amplitude, the UV divergent region is simply *absent*. We can also see this from the point of view of the momentum integral (7.3.8a). Since τ_2 is bounded below, this integral (after contour rotation) is a convergent Gaussian.

Another possible divergence comes from the limit $\tau_2 \to \infty$, where the torus becomes very long. In this region, the string amplitude in 26 flat dimensions has the expansion

$$2\pi i V_{26} \int^\infty \frac{d\tau_2}{2\tau_2} (4\pi^2 \alpha' \tau_2)^{-13} \left[\exp(4\pi\tau_2) + 24^2 + \ldots\right] \,, \tag{7.3.15}$$

the asymptotic behavior being controlled by the lightest string states. The first term in the series diverges, and the interpretation is clear from the field theory point of view. The series is in order of increasing mass-squared. The first term is from the tachyon and diverges due to the positive exponential in the path sum (7.3.9). This divergence is therefore an artifact of a theory with a tachyon, and will not afflict more realistic string theories. Incidentally, this path sum can be defined by analytic continuation from positive mass-squared, but there is still a pathology — the continued energy density is complex. This signifies an instability, the tachyon field rolling down its inverted potential. With a general CFT, eq. (7.3.12) becomes

$$2\pi i V_d \int^\infty \frac{d\tau_2}{2\tau_2} (4\pi^2 \alpha' \tau_2)^{-d/2} \sum_i \exp(-\pi\alpha' m_i^2 \tau_2) \,. \tag{7.3.16}$$

Because the tachyon is always present in bosonic string theory this again diverges; in the absence of tachyons it will converge.

The torus illustrates a general principle, which holds for all string amplitudes: there is no UV region of moduli space that might give rise to high energy divergences. All limits are controlled by the lightest states, the long-distance physics. For the torus with vertex operators the τ-integration is still cut off as above, but there are more limits to consider as vertex operators approach one another. The same general principle applies to these as well, as we will develop further in chapter 9.

One might try to remove the UV divergence from field theory by cutting off the l integral. Similarly, one might try to make an analogous modification of string theory, for example replacing the usual fundamental region F_0 with the region $|\tau_1| \leq \frac{1}{2}$, $\tau_2 > 1$ whose lower edge is straight. However, in either case this would spoil the consistency of the theory: unphysical negative norm states would no longer decouple. We have seen in the general discussion in section 5.4 that the coupling of a BRST-null state is proportional to a total derivative on moduli space. For the fundamental region F_0, the apparent boundaries I and I′ and II and II′ respectively (shown in fig. 5.2) are identified and the surface terms cancel. Modifying the region of integration introduces real boundaries. Surface terms on the modified moduli space will no longer cancel, so there is nonzero amplitude for null states and an inconsistent quantum theory. This is directly analogous to what happens in gravity if we try to make the theory finite by some brute force cutoff on the short-distance physics: it is exceedingly difficult to do this without spoiling the local spacetime symmetry and making the theory inconsistent. String theory manages in a subtle way to soften the short-distance behavior and eliminate the divergences without loss of spacetime gauge invariance.

Physics of the vacuum amplitude

Besides serving as a simple illustration of the behavior of string amplitudes, the one-loop vacuum amplitude has an interesting physical interpretation. In point-particle theory, 'vacuum' paths consist of any number n of disconnected circles. Including a factor of $1/n!$ for permutation symmetry and summing in n leads to

$$Z_{\text{vac}}(m^2) = \exp\left[Z_{S_1}(m^2)\right]. \tag{7.3.17}$$

Translating into canonical field theory,

$$\begin{aligned} Z_{\text{vac}}(m^2) &= \langle 0|\exp(-iHT)|0\rangle \\ &= \exp(-i\rho_0 V_d)\,, \end{aligned} \tag{7.3.18}$$

where ρ_0 is the vacuum energy density:

$$\rho_0 = \frac{i}{V_d} Z_{S_1}(m^2)\,. \tag{7.3.19}$$

The l integral in $Z_{S_1}(m^2)$ diverges as $l \to 0$. In renormalizable field theory this divergence is canceled by counterterms or supersymmetry. To get a rough insight into the physics we will cut off the integral at $l \geq \epsilon$ and then take $\epsilon \to 0$, but drop divergent terms. With this prescription,

$$\int_0^\infty \frac{dl}{2l} \exp[-(k^2+m^2)l/2] \to -\frac{1}{2}\ln(k^2+m^2) \tag{7.3.20}$$

and

$$i\int_0^\infty \frac{dl}{2l}\int_{-\infty}^{\infty}\frac{dk^0}{2\pi}\exp[-(k^2+m^2)l/2] \to \frac{\omega_{\mathbf{k}}}{2}\,, \tag{7.3.21}$$

where $\omega_{\mathbf{k}}^2 = \mathbf{k}\cdot\mathbf{k} + m^2$. Using the latter form, the vacuum energy density becomes

$$\rho_0 = \int \frac{d^{d-1}\mathbf{k}}{(2\pi)^{d-1}}\,\frac{\omega_{\mathbf{k}}}{2}\,, \tag{7.3.22}$$

which is just the sum of the zero-point energies of the modes of the field. This is in spacetime, of course — we encountered the same kind of sum but on the world-sheet in section 1.3.

The description of a quantum field theory as a sum over particle paths is not as familiar as the description in terms of a sum over field histories, but it is equivalent. In particular, the free path integral in field theory can be found in most modern field theory textbooks:

$$\begin{aligned} \ln Z_{\text{vac}}(m^2) &= -\frac{1}{2}\text{Tr}\ln(-\partial^2+m^2) \\ &= -\frac{V_d}{2}\int\frac{d^dk}{(2\pi)^d}\ln(k^2+m^2). \end{aligned} \tag{7.3.23}$$

Using eq. (7.3.20), this is the same as the path sum result (7.3.9).

In older quantum field theory texts, one often reads that vacuum amplitudes such as eq. (7.3.18) are irrelevant because they give an overall phase that divides out of any expectation value. This is true if one is considering scattering experiments in a fixed background and ignoring gravity, but there are at least two circumstances in which the vacuum energy density is quite important. The first is in comparing the energy densities of *different* states to determine which is the true ground state of a theory. It is very likely, for example, that the breaking of the electroweak $SU(2)\times U(1)$ symmetry is determined in part by such quantum corrections to the vacuum energy density. This was the original motivation that led to the Coleman–Weinberg formula (7.3.23). The generalization to a theory with particles of arbitrary spin is

$$\rho_0 = \frac{i}{V_d}\sum_i (-1)^{\mathbf{F}_i} Z_{S_1}(m_i^2). \tag{7.3.24}$$

The sum runs over all physical particle states. Each polarization counts separately, giving a factor of $2s_i + 1$ for a particle of spin s_i in four dimensions. Here $\mathbf{F}_i$ is the spacetime fermion number, defined mod 2, so fermions contribute with the opposite sign.

The second circumstance is when one considers the coupling to gravity. The vacuum energy gives a source term in Einstein's equation, the

cosmological constant, and so has observable effects. In fact, this cosmological constant is a great challenge, because the fact that spacetime is approximately flat and static means that its value is very small,

$$|\rho_0| \lesssim 10^{-44}\,\mathrm{GeV}^4\,. \tag{7.3.25}$$

If one considers only the contributions to the vacuum energy from the vacuum fluctuations of the known particles up to the currently explored energy (roughly the electroweak scale m_{ew}), the zero-point energy is already of order

$$m_{\mathrm{ew}}^4 \approx 10^8\,\mathrm{GeV}^4\,, \tag{7.3.26}$$

52 orders of magnitude too large. The Higgs field potential energy and the QCD vacuum energy are also far too large. Finding a mechanism that would account for the cancellation of the net cosmological constant to great accuracy has proven very difficult. For example, in a supersymmetric theory, the contributions of degenerate fermions and bosons in the sum (7.3.24) cancel. But supersymmetry is not seen in nature, so it must be a broken symmetry. The cancellation is then imperfect, leaving again a remainder of at least m_{ew}^4. The cosmological constant problem is one of the most nagging difficulties, and therefore probably one of the best clues, in trying to find a unified theory with gravity.

What about the cosmological constant problem in string theory? At string tree level we had a consistent theory with a flat metric, so the cosmological constant was zero. In fact we arranged this by hand, when we took 26 dimensions. One sees from the spacetime action (3.7.20) that there would otherwise be a tree-level potential energy proportional to $D-26$. The one-loop vacuum energy density in bosonic string theory is nonzero, and is necessarily of the order of the string scale (ignoring the tachyon divergence). In four dimensions this would correspond to 10^{72} GeV^4, which is again far too large. In a supersymmetric string theory, there will be a certain amount of cancellation, but again one expects a remainder of at least m_{ew}^4 in a realistic theory with supersymmetry breaking.

The cosmological constant problem is telling us that there is something that we still do not understand about the vacuum, in field theory and string theory equally. We will return to this problem at various points.

7.4 Open and unoriented one-loop graphs

The cylinder

The results for the torus are readily extended to the other surfaces of Euler number zero. For example, the vacuum amplitude from the cylinder

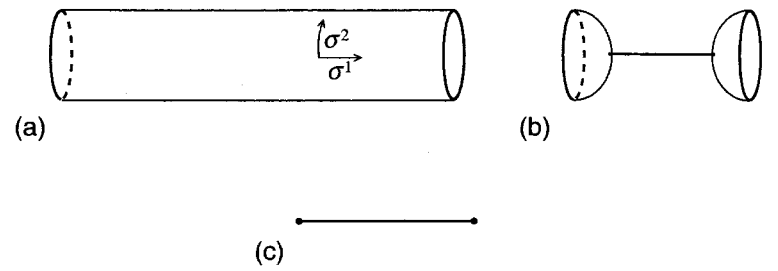


Fig. 7.1. (a) Cylinder in the limit of small t. (b) The amplitude separated into disk tadpole amplitudes and a closed string propagator. (c) Analogous field theory graph. The heavy circles represent the tadpoles.

in the oriented theory is

$$\begin{aligned} Z_{C_2} &= \int_0^\infty \frac{dt}{2t}\,\mathrm{Tr}'_{\mathrm{o}}[\exp(-2\pi t L_0)] \\ &= iV_d \int_0^\infty \frac{dt}{2t}(8\pi^2\alpha' t)^{-d/2} \sum_{i\in\mathscr{H}_{\mathrm{o}}^{\perp}} \exp[-2\pi t(h_i-1)] \\ &\to iV_{26}\,n^2 \int_0^\infty \frac{dt}{2t}(8\pi^2\alpha' t)^{-13}\eta(it)^{-24}\,. \end{aligned} \tag{7.4.1}$$

This can be obtained either by working out the path integral measure in terms of the ghost zero modes, as we did for the torus, or by guessing (correctly) that we should again sum the point-particle result (7.3.9) over the open string spectrum. In the first line, the trace is over the full open string CFT, except that for omission of the ghost zero modes as denoted by the prime. In the final line the trace has been carried out for the case of 26 flat dimensions, with n Chan–Paton degrees of freedom. Including tachyon vertex operators is straightforward.

The $t\to\infty$ limit of the cylinder is much like the $\tau_2\to\infty$ limit of the torus. The cylinder looks like a long strip, and the leading asymptotics are given by the lightest open string states. As in the closed string there is a divergence, but only from the open string tachyon.

The $t\to 0$ limit is rather interesting. Unlike the case of the torus, there is no modular group acting to cut off the range of integration, and so the UV divergence of field theory still seems to be present. However, we will see that, as with all divergences in string theory, this should actually be interpreted as a *long-distance* effect. In the $t\to 0$ limit the cylinder is very long as shown in figure 7.1(a). That is, it looks like a closed string appearing from the vacuum, propagating a distance, and then disappearing again into the vacuum. To make this clear, use the

modular transformation (7.2.44)

$$\eta(it) = t^{-1/2}\eta(i/t) \tag{7.4.2}$$

and change variables to $s = \pi/t$, with the result

$$Z_{C_2} = i\frac{V_{26}n^2}{2\pi(8\pi^2\alpha')^{13}}\int_0^\infty ds\,\eta(is/\pi)^{-24}\,. \tag{7.4.3}$$

Scaling down the metric by $1/t$ so that the cylinder has the usual closed string circumference 2π, the length of the cylinder is s. Expanding

$$\begin{aligned}\eta(is/\pi)^{-24} &= \exp(2s)\prod_{n=1}^{\infty}[1-\exp(-2ns)]^{-24}\\ &= \exp(2s)+24+O(\exp(-2s))\,,\end{aligned} \tag{7.4.4}$$

one sees the expected asymptotics from expanding in a complete set of closed string states. In other words, if we think of σ^2 as world-sheet time and σ^1 as world-sheet space, figure 7.1(a) is a very short open string loop. If we reverse the roles of world-sheet space and time, it is a very long closed string world-line beginning and ending on boundary loops. In the Euclidean path integral, either description can be used, and each is useful in a different limit of moduli space.

The leading divergence in the vacuum amplitude is from the closed string tachyon. This is uninteresting and can be defined by analytic continuation,

$$\int_0^\infty ds\,\exp(\beta s) \equiv -\frac{1}{\beta}\,. \tag{7.4.5}$$

The second term, from the massless closed string states, gives a divergence of the form $1/0$ even with the continuation. To see the origin of this divergence, imagine separating the process as shown in figure 7.1(b). As we have discussed in section 6.6, there is a nonzero amplitude with one closed string vertex operator on the disk. This tadpole corresponds to a closed string appearing from or disappearing into the vacuum. In between the two tadpoles is the closed string propagator. In momentum space the massless propagator is proportional to $1/k^2$. Here, momentum conservation requires that the closed string appear from the vacuum with zero momentum, so the propagator diverges.

This same kind of divergence occurs in quantum field theory. Consider a massless scalar field ϕ with a term linear in ϕ in the Lagrangian. There is then a vertex which connects to just a single propagator, and the graph shown in figure 7.1(c) exists. This diverges because the intermediate propagator is

$$\frac{1}{k^2}\Big|_{k^\mu=0}\,. \tag{7.4.6}$$

This is a long-distance (IR) divergence, because poles in propagators are associated with propagation over long spacetime distances.

UV and IR divergences in quantum field theory have very different physical origins. UV divergences usually signify a breakdown of the theory, the need for new physics at some short distance. IR divergences generally mean that we have asked the wrong question, or expanded in the wrong way. So it is here. In normal perturbation theory we expand around $\phi(x) = 0$, or some other constant configuration. For the action

$$-\frac{1}{g^2}\int d^dx\left(\frac{1}{2}\partial_\mu\phi\partial^\mu\phi + g\Lambda\phi\right)\,, \tag{7.4.7}$$

the equation of motion

$$\partial^2\phi = g\Lambda \tag{7.4.8}$$

does not allow $\phi(x) = 0$ as a solution. Instead we must expand around a solution to the equation (7.4.8); any solution is necessarily position-dependent. The resulting amplitudes are free of divergences. This is true even though the right-hand side is a perturbation (we have included the factor of g appropriate to the disk), because it is a singular perturbation. In particular, the corrected background breaks some of the Poincaré symmetry of the zeroth order solution.

The situation is just the same in string theory. The disk tadpole is a source

$$-\Lambda\int d^{26}x\,(-G)^{1/2}e^{-\tilde{\Phi}} \tag{7.4.9}$$

for both the dilaton and metric. Expanding around a solution to the corrected field equations (with the metric and dilaton no longer constant) leads to finite amplitudes. The details are a bit intricate, and will be discussed somewhat further in chapter 9. Incidentally, in supersymmetric string theories if the tree-level background is invariant under supersymmetry, it usually receives no loop corrections.

The pole (7.4.6) is the same kind of divergence encountered in the tree-level amplitudes, a resonance corresponding to propagation over long spacetime distances. If we add open string vertex operators to each end of the cylinder so that it represents a one-loop open string scattering amplitude, then the momentum k^μ flowing from one boundary to the other is in general nonzero. The large-s limit (7.4.4) then includes a factor of

$$\exp(-\alpha' k^2 s/2) \tag{7.4.10}$$

and the divergence becomes a momentum pole representing scattering of open strings into an intermediate closed string. Thus, as claimed in chapter 3, an open string theory must include closed strings as well. It is

curious that the mechanism that removes the UV divergence is different in the closed and open string cases. In the closed string it is an effective cutoff on the modular integration; in the open string it is a reinterpretation of the dangerous limit of moduli space in terms of long distances.

In eq. (7.4.1) the path integral on the torus has been related to a trace over the open string spectrum by cutting open the path integral with σ^2 treated as time. It can also be obtained in terms of the closed string Hilbert space by treating σ^1 as time. Let σ^2 be periodic with period 2π and the boundaries be at $\sigma^1 = 0, s$. The closed string appears in some state $|B\rangle$ at $\sigma^1 = 0$ and then disappears in the same way at $\sigma^1 = s$. Including the measure insertions, the path integral is then proportional to

$$\langle B|c_0 b_0 \exp[-s(L_0 + \tilde{L}_0)]|B\rangle. \tag{7.4.11}$$

The *boundary state* $|B\rangle$ is determined by the condition that $\partial_1 X^\mu$, c^1, and b_{12} vanish on the boundary. In the Hamiltonian form these must annihilate $|B\rangle$, which in terms of the Laurent coefficients is

$$(\alpha_n^\mu + \tilde{\alpha}_{-n}^\mu)|B\rangle = (c_n + \tilde{c}_{-n})|B\rangle = (b_n - \tilde{b}_{-n})|B\rangle = 0\,, \quad \text{all } n\,. \tag{7.4.12}$$

This determines

$$|B\rangle \propto \exp\Big[-\sum_{n=1}^{\infty}(n^{-1}\alpha_n \cdot \tilde{\alpha}_n + b_n\tilde{c}_n + \tilde{b}_n c_n)\Big]|0;0\rangle\,. \tag{7.4.13}$$

Using this in (7.4.11) gives the result (7.4.3), except that the normalization of $|B\rangle$ is undetermined. This representation is useful in analyzing the $t \to 0$ limit and the closed string poles.

By comparing the string and field theory calculations we can determine the disk tadpole Λ, but it will be more convenient to do this in the next chapter as a special case of a more general result.

The Klein bottle

The vacuum amplitude from the Klein bottle is

$$\begin{aligned} Z_{K_2} &= \int_0^\infty \frac{dt}{4t}\,\mathrm{Tr}'_c\Big\{\Omega\exp[-2\pi t(L_0+\tilde{L}_0)]\Big\} \\ &= iV_d\int_0^\infty \frac{dt}{4t}(4\pi^2\alpha' t)^{-d/2}\sum_{i\in\mathscr{H}_c^\perp}\Omega_i\exp[-2\pi t(h_i+\tilde{h}_i-2)]\,, \end{aligned} \tag{7.4.14}$$

where the notation follows the cylinder (7.4.1). Relative to the torus in the oriented theory there is an extra factor of $\frac{1}{2}$ from the projection operator $\frac{1}{2}(1+\Omega)$. For the same reason there is an extra $\frac{1}{2}$ in both the torus and cylinder amplitudes in the unoriented theory. One can also think of this as coming from the extra gauge invariance $w \to \bar{w}$. To evaluate the trace for 26 flat dimensions, note that the only diagonal elements of Ω are those

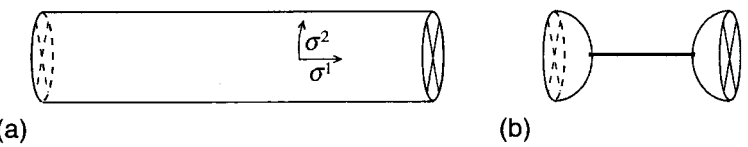


Fig. 7.2. (a) Klein bottle in the limit of small t as a cylinder capped by cross-caps. (b) The amplitude separated into RP_2 tadpole amplitudes and a closed string propagator.

for which the right- and left-movers are in the same state, and these states contribute with $\Omega = +1$. The trace is then effectively over only one side, which is the same as the open string spectrum, except that the weight is doubled because the right- and left-movers make equal contributions. The result is

$$Z_{K_2} \to iV_{26}\int_0^\infty \frac{dt}{4t}(4\pi^2\alpha' t)^{-13}\eta(2it)^{-24}\,. \tag{7.4.15}$$

The modulus t runs over the same range $0 < t < \infty$ as for the cylinder, so the $t \to 0$ divergences are the same. Again the $1/0$ pole has a long-distance interpretation in terms of a closed string pole. To see this consider the regions

$$0 \le \sigma^1 \le 2\pi\,, \quad 0 \le \sigma^2 \le 2\pi t\,, \tag{7.4.16a}$$

$$0 \le \sigma^1 \le \pi\,, \quad 0 \le \sigma^2 \le 4\pi t\,, \tag{7.4.16b}$$

each of which is a fundamental region for the identification

$$w \cong w + 2\pi \cong -\bar{w} + 2\pi i t\,. \tag{7.4.17}$$

In (7.4.16a), the left- and right-hand edges are periodically identified while the upper and lower edges are identified after a parity-reversal, giving an interpretation as a closed string loop weighted by Ω. For (7.4.16b), note that the identifications (7.4.17) imply also that

$$w \cong w + 4\pi i t\,, \quad w + \pi \cong -(\bar{w} + \pi) + 2\pi i t\,. \tag{7.4.18}$$

It follows that the upper and lower edges of the region (7.4.16b) are periodically identified. The left-hand edge is identified with itself after translation by half its length and reflection through the edge, and similarly the right-hand edge: this is the definition of a cross-cap. Thus we have the representation in figure 7.2(a) as a cylinder capped by two cross-caps. Rescaling by $1/2t$, the cylinder has circumference 2π and length $s = \pi/2t$. After a modular transformation, the amplitude becomes

$$Z_{K_2} = i\frac{2^{26}V_{26}}{4\pi(8\pi^2\alpha')^{13}}\int_0^\infty ds\,\eta(is/\pi)^{-24}\,. \tag{7.4.19}$$

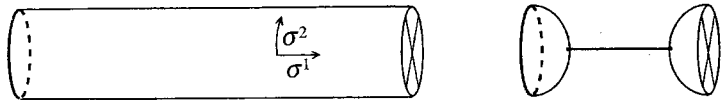


Fig. 7.3. (a) Möbius in the limit of small t as a cylinder capped by one cross-cap. (b) The amplitude separated into D_2 and RP_2 tadpole amplitudes and a closed string propagator.

All the discussion of the cylinder divergence now applies, except that the tadpole is from the projective plane rather than the disk.

The Möbius strip

For the Möbius strip,

$$Z_{M_2} = iV_d \int_0^\infty \frac{dt}{4t}(8\pi^2\alpha' t)^{-d/2} \sum_{i\in\mathscr{H}_o^\perp} \Omega_i \exp[-2\pi t(h_i - 1)] \qquad (7.4.20)$$

This differs from the result on the cylinder only by the Ω and the factor of $\frac{1}{2}$ from the projection operator. In 26 flat dimensions, the effect of the operator Ω in the trace is an extra -1 at the even mass levels plus an appropriate accounting of the Chan–Paton factors. The oscillator trace is thus

$$\exp(2\pi t)\prod_{n=1}^{\infty}[1-(-1)^n\exp(-2\pi nt)]^{-24} = \vartheta_{00}(0,2it)^{-12}\eta(2it)^{-12} \ . \qquad (7.4.21)$$

For the $SO(n)$ theory the $\frac{1}{2}n(n+1)$ symmetric states have $\Omega = +1$ while the $\frac{1}{2}n(n-1)$ antisymmetric states have $\Omega = -1$ for a net contribution of n. For the $Sp(k)$ theory these degeneracies are reversed, giving $-n$ (recall that in our notation $Sp(k)$ corresponds to $n = 2k$ Chan–Paton states). The amplitude is then

$$Z_{M_2} = \pm inV_{26}\int_0^\infty \frac{dt}{4t}(8\pi^2\alpha' t)^{-13}\vartheta_{00}(0,2it)^{-12}\eta(2it)^{-12} \ . \qquad (7.4.22)$$

By the same construction as for the Klein bottle, the Möbius strip can be represented as a cylinder with a boundary at one end and a cross-cap on the other, as in figure 7.3(a). The length of the cylinder is now $s = \pi/4t$. By a modular transformation, the amplitude becomes

$$Z_{M_2} = \pm 2in\frac{2^{13}V_{26}}{4\pi(8\pi^2\alpha')^{13}}\int_0^\infty ds\,\vartheta_{00}(0,2is/\pi)^{-12}\eta(2is/\pi)^{-12} \ . \qquad (7.4.23)$$

As for the annulus this can be written as an operator expression, one of the boundary states in eq. (7.4.11) being replaced by an analogous cross-cap state $|C\rangle$. There is again a $t \to 0$ divergence; it corresponds to the process of figure 7.3(a) with one tadpole from the disk and one from the projective plane.

In the unoriented theory the divergences on these three surfaces combine into

$$i\frac{24V_{26}}{4\pi(8\pi^2\alpha')^{13}}(2^{13} \mp n)^2\int_0^\infty ds \ . \qquad (7.4.24)$$

That is, the total tadpole is proportional to $2^{13} \mp n$. For the gauge group $SO(2^{13}) = SO(8192)$ this vanishes, the tadpoles from the disk and projective plane canceling. For the bosonic string this probably has no special significance, but the analogous cancellation for $SO(32)$ in superstring theory does.

Exercises

7.1 Fill in the steps leading to the expectation value (7.2.4) of exponential operators on the torus. Show that it has the correct transformation under $\tau \to -1/\tau$.

7.2 Derive the torus vacuum amplitude (7.3.6) by regulating and evaluating the determinants, as is done for the harmonic operator in appendix A. Show that a modular transformation just permutes the eigenvalues. [Compare exercise A.3.] (Reference: Polchinski (1986).)

7.3 Derive the Green's function (7.2.3) by carrying out the eigenfunction sum (6.2.7).

7.4 Evaluate $\langle\, \partial_w X^\mu(w)\partial_w X_\mu(0)\,\rangle$ on the torus by representing it as a trace. Show that the result agrees with eq. (7.2.16).

7.5 (a) Obtain the leading behavior of the theta functions (7.2.37) as $\mathrm{Im}\,\tau \to \infty$.
(b) By using the modular transformations, do the same as $\tau \to 0$ along the imaginary axis.
(c) A more exotic question: what if τ approaches a nonzero point on the real axis, along a path parallel to the imaginary axis? [Hint: the answer depends crucially on whether the point is rational or irrational.]

7.6 (a) Verify eqs. (7.3.20) and (7.3.21).
(b) Evaluate $Z_{S_1}(m^2)$ with a cutoff on the l-integral in any dimension d. Show that the counterterms are analytic in m^2. Give the nonanalytic part for $d = 4$. (Reference: Coleman & Weinberg (1973).)

7.7 Obtain the amplitude for n open string tachyons on the cylinder

Indices M, N run over all dimensions $0, \ldots, d$ and indices μ, ν run only over noncompact dimensions $0, \ldots, d-1$. We designate the full D-dimensional metric by G^D_{MN}. One should note that $G_{\mu\nu} \neq G^D_{\mu\nu}$; in d-dimensional actions we raise and lower indices with $G_{\mu\nu}$. For now the fields $G_{\mu\nu}$, G_{dd}, and A_μ are allowed to depend only on the noncompact coordinates x^μ. The form (8.1.2) is the most general metric invariant under translations of x^d. This form still allows d-dimensional reparameterizations $x'^\mu(x^\nu)$ and also reparameterizations

$$x'^d = x^d + \lambda(x^\mu) \,. \tag{8.1.3}$$

Under the latter,

$$A'_\mu = A_\mu - \partial_\mu \lambda \,, \tag{8.1.4}$$

so gauge transformations arise as part of the higher-dimensional coordinate group. This is the *Kaluza–Klein* mechanism.

To see the effect of x^d-dependence, consider a massless scalar ϕ in D dimensions, where for simplicity the metric in the compact direction is $G_{dd} = 1$. The momentum in the periodic dimension is quantized, $p_d = n/R$. Expand the x^d-dependence of ϕ in a complete set,

$$\phi(x^M) = \sum_{n=-\infty}^{\infty} \phi_n(x^\mu) \exp(inx^d/R) \,. \tag{8.1.5}$$

The D-dimensional wave equation $\partial_M \partial^M \phi = 0$ becomes

$$\partial_\mu \partial^\mu \phi_n(x^\mu) = \frac{n^2}{R^2} \phi_n(x^\mu) \,. \tag{8.1.6}$$

The modes ϕ_n of the D-dimensional field thus become an infinite tower of d-dimensional fields, labeled by n. The d-dimensional mass-squared

$$-p^\mu p_\mu = \frac{n^2}{R^2} \tag{8.1.7}$$

is nonzero for all fields with nonvanishing p_d. At energies small compared to R^{-1}, only the x^d-independent fields remain and the physics is d-dimensional. At energies above R^{-1}, one sees the tower of Kaluza–Klein states.

The charge corresponding to the Kaluza–Klein gauge invariance (8.1.3) is the p_d-momentum. In this simple example, all fields carrying the Kaluza–Klein charge are massive. More generally, with higher spin fields and curved backgrounds, there can be *massless* charged fields.

The effective action for the massless fields is always an important object to consider. Define $G_{dd} = e^{2\sigma}$. The Ricci scalar for the metric (8.1.2) is

$$\boldsymbol{R} = \boldsymbol{R}_d - 2e^{-\sigma}\nabla^2 e^\sigma - \frac{1}{4} e^{2\sigma} F_{\mu\nu} F^{\mu\nu} \,, \tag{8.1.8}$$

where $\boldsymbol{R}$ is constructed from G_{MN} and $\boldsymbol{R}_d$ from $G_{\mu\nu}$. The graviton-dilaton action (3.7.20) becomes

$$\begin{aligned} \boldsymbol{S}_1 &= \frac{1}{2\kappa_0^2} \int d^D x (-G_d)^{1/2} e^{-2\Phi} (\boldsymbol{R} + 4\nabla_\mu \Phi \nabla^\mu \Phi) \\ &= \frac{\pi R}{\kappa_0^2} \int d^d x (-G_d)^{1/2} e^{-2\Phi + \sigma} \\ &\qquad \times \left(\boldsymbol{R}_d - 4\partial_\mu \Phi \partial^\mu \sigma + 4\partial_\mu \Phi \partial^\mu \Phi - \frac{1}{4} e^{2\sigma} F_{\mu\nu} F^{\mu\nu} \right) \\ &= \frac{\pi R}{\kappa_0^2} \int d^d x (-G_d)^{1/2} e^{-2\Phi_d} \\ &\qquad \times \left(\boldsymbol{R}_d - \partial_\mu \sigma \partial^\mu \sigma + 4\partial_\mu \Phi_d \partial^\mu \Phi_d - \frac{1}{4} e^{2\sigma} F_{\mu\nu} F^{\mu\nu} \right) , \end{aligned} \tag{8.1.9}$$

giving kinetic terms for all the massless fields. Here G_d denotes the determinant of $G_{\mu\nu}$, and in the last line we have introduced the effective d-dimensional dilaton, $\Phi_d = \Phi - \sigma/2$. The apparent wrong sign of the dilaton kinetic term is illusory because mixing of the graviton with the trace of the metric must also be taken into account. This is most easily done by means of a Weyl transformation as in eq. (3.7.25), and the resulting kinetic term has the correct sign.

The field equations do not determine the radius of the compact dimension:[1] a flat metric and constant dilaton are a solution for any values of Φ and σ. In other words, there is no potential energy for Φ or for σ, and so these fields are massless, much like Goldstone bosons. The different values of Φ and σ label degenerate configurations (or states in the quantum theory), and a state in which these fields are slowly varying has energy only from the gradient. The difference from the Goldstone phenomenon is that the degenerate states are not related to one another by any symmetry, and in fact the physics depends on Φ and σ. In this bosonic theory the degeneracy is accidental and the one-loop energy, discussed in the previous chapter, breaks the degeneracy. In supersymmetric theories, the existence of physically inequivalent but degenerate vacua is quite common and plays an important role in understanding the dynamics. The massless fields, which label the inequivalent vacua, are called *moduli*. In nature,

[1] Of course, only the invariant radius

$$\rho = Re^\sigma$$

distinguishes physically inequivalent solutions. We could set R to some convenient value, say $\alpha'^{1/2}$, or to unity if we use dimensionless variables, but it is convenient to leave it general; often it will be more convenient to set G_{dd} to unity instead. Similarly we could set κ_0^2 to be the appropriate power of α' by an additive shift of Φ, but again it does not hurt to leave it general and we will see in chapter 13 that it is natural to fix the additive normalization of Φ in a different way.

supersymmetry breaking must almost surely give mass to all moduli, else they would mediate infinite-range interactions of roughly gravitational strength. We will see in chapter 14 that in some string theories the dilaton itself is the radius of a hidden dimension, one that is not evident in string perturbation theory.

Defining $A_\mu = R\tilde{A}_\mu$, the covariant derivative is

$$\partial_\mu + ip_d A_\mu = \partial_\mu + in\tilde{A}_\mu \,, \tag{8.1.10}$$

so that the charges are integers. The d-dimensional gauge and gravitational couplings are conventionally defined as follows. The coefficient of $\tilde{F}_{\mu\nu}\tilde{F}^{\mu\nu}$ in the Lagrangian density is defined to be $-1/4g_d^2$, and that of $\boldsymbol{R}_d$ is defined to be $1/2\kappa_d^2$. The gauge coupling is thus determined in terms of the gravitational coupling,

$$g_d^2 = \frac{\kappa_0^2 e^{2\Phi_d}}{\pi R^3 e^{2\sigma}} = \frac{2\kappa_d^2}{\rho^2} \,. \tag{8.1.11}$$

The d-dimensional and D-dimensional gravitational couplings are related

$$\frac{1}{\kappa_d^2} = \frac{2\pi\rho}{\kappa^2} \,, \tag{8.1.12}$$

$2\pi\rho$ being the volume of the compactified dimension.

The antisymmetric tensor also gives rise to a gauge symmetry by a generalization of the Kaluza–Klein mechanism. Separating B_{MN} into $B_{\mu\nu}$ and $A'_\mu = B_{d\mu}$, the gauge parameter ζ_M defined in eq. (3.7.7) separates into a d-dimensional antisymmetric tensor transformation ζ_μ and an ordinary gauge invariance ζ_d. The gauge field is $B_{d\mu}$ and the field strength $H_{d\mu\nu}$. The antisymmetric tensor action becomes

$$\begin{aligned} S_2 &= -\frac{1}{24\kappa_0^2}\int d^D x\,(-G_D)^{1/2} e^{-2\Phi} H_{MNL}H^{MNL} \\ &= -\frac{\pi R}{12\kappa_0^2}\int d^d x\,(-G_d)^{1/2} e^{-2\Phi_d}\left(\tilde{H}_{\mu\nu\lambda}\tilde{H}^{\mu\nu\lambda} + 3e^{-2\sigma}H_{d\mu\nu}H_d^{\mu\nu}\right) . \end{aligned} \tag{8.1.13}$$

We have defined

$$\tilde{H}_{\mu\nu\lambda} = (\partial_\mu B_{\nu\lambda} - \partial_\lambda H_{d\nu\lambda}) + \text{cyclic permutations} \,. \tag{8.1.14}$$

The term proportional to the vector potential arises from the inverse metric G^{MN}. It is known as a *Chern–Simons term*, this signifying the antisymmetrized combination of one gauge potential and any number of field strengths. Such terms appear in many places in supersymmetric theories, and are associated with interesting physical effects. Notice in particular that $\tilde{H}_{\mu\nu\lambda}$ is gauge-invariant because the variation (8.1.4) of A_μ is canceled by a variation

$$B'_{\nu\lambda} = B_{\nu\lambda} - \lambda H_{d\nu\lambda} \,. \tag{8.1.15}$$

There is no way to couple the potential B_{MN} minimally to other fields, and so unlike the Kaluza–Klein case there are no fields charged under the antisymmetric tensor gauge symmetry; this will be different in string theory.

8.2 Toroidal compactification in CFT

Now we consider the conformal field theory of a single periodic scalar field,

$$X \cong X + 2\pi R \,. \tag{8.2.1}$$

To keep the equations uncluttered we drop the superscript from X^d and set $G_{dd} = 1$. The world-sheet action is as in the noncompact theory, $\int d^2z\,\partial X\bar{\partial}X/2\pi\alpha'$, so the equations of motion, operator products, and energy-momentum tensor are unchanged. In particular, the theory is still conformally invariant. The periodicity has two effects. First, string states must be single-valued under the identification (8.2.1). That is, the operator $\exp(2\pi i R p)$ which translates strings once around the periodic dimension must leave states invariant, so the center-of-mass momentum is quantized

$$k = \frac{n}{R} \,, \quad n \in \mathbf{Z} \,. \tag{8.2.2}$$

This is just as in field theory.

The second effect is special to string theory. A closed string may now wind around the compact direction,

$$X(\sigma + 2\pi) = X(\sigma) + 2\pi R w \,, \qquad w \in \mathbf{Z} \,. \tag{8.2.3}$$

The integer w is the *winding number*. States of winding numbers $+1$, 0, and -1 are shown in figure 8.1(a). From the point of view of the world-sheet field theory, strings of nonzero winding number are *topological solitons*, states with a topologically nontrivial field configuration. A consistent string theory must include the winding number states: by the usual splitting–joining process a $w = 0$ string can turn into a $w = +1$, $w = -1$ pair as shown in figure 8.1(b). It is easy to see that winding number is always conserved as in this example.

To describe the states of the closed string CFT, consider the Laurent

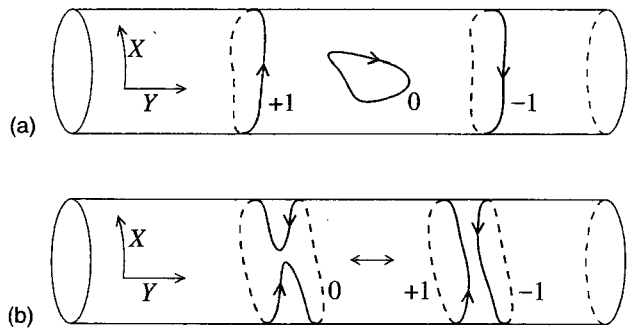


Fig. 8.1. (a) Closed oriented strings of winding number $w = +1, 0, -1$. For illustration, one compact dimension X and one noncompact dimension Y are shown. (b) Transition of a $w = 0$ string into $w = +1$ and $w = -1$ strings.

expansions

$$\partial X(z) = -i\left(\frac{\alpha'}{2}\right)^{1/2} \sum_{m=-\infty}^{\infty} \frac{\alpha_m}{z^{m+1}}\,, \quad \bar\partial X(\bar z) = -i\left(\frac{\alpha'}{2}\right)^{1/2} \sum_{m=-\infty}^{\infty} \frac{\tilde\alpha_m}{\bar z^{m+1}}\,. \tag{8.2.4}$$

The total change in the coordinate X in going around the string is

$$2\pi R w = \oint (dz\,\partial X + d\bar z\,\bar\partial X) = 2\pi(\alpha'/2)^{1/2}(\alpha_0 - \tilde\alpha_0)\,. \tag{8.2.5}$$

The total Noether momentum is

$$p = \frac{1}{2\pi\alpha'} \oint (dz\,\partial X - d\bar z\,\bar\partial X) = (2\alpha')^{-1/2}(\alpha_0 + \tilde\alpha_0)\,. \tag{8.2.6}$$

For a noncompact dimension this would give the usual relation $\alpha_0 = \tilde\alpha_0 = p(\alpha'/2)^{1/2}$, but for the periodic dimension we have

$$p_L \equiv (2/\alpha')^{1/2}\alpha_0 = \frac{n}{R} + \frac{wR}{\alpha'}\,, \tag{8.2.7a}$$

$$p_R \equiv (2/\alpha')^{1/2}\tilde\alpha_0 = \frac{n}{R} - \frac{wR}{\alpha'}\,. \tag{8.2.7b}$$

The Virasoro generators are

$$L_0 = \frac{\alpha' p_L^2}{4} + \sum_{n=1}^{\infty} \alpha_{-n}\alpha_n\,, \tag{8.2.8a}$$

$$\tilde L_0 = \frac{\alpha' p_R^2}{4} + \sum_{n=1}^{\infty} \tilde\alpha_{-n}\tilde\alpha_n\,. \tag{8.2.8b}$$

The partition function

The partition function for X is now

$$\begin{aligned}(q\bar q)^{-1/24}\mathrm{Tr}\left(q^{L_0}\bar q^{\tilde L_0}\right) &= |\eta(\tau)|^{-2} \sum_{n,w=-\infty}^{\infty} q^{\alpha' p_L^2/4}\bar q^{\alpha' p_R^2/4} \\ &= |\eta(\tau)|^{-2} \sum_{n,w=-\infty}^{\infty} \exp\left[-\pi\tau_2\left(\frac{\alpha' n^2}{R^2} + \frac{w^2R^2}{\alpha'}\right) + 2\pi i\tau_1 nw\right]\,. \end{aligned} \tag{8.2.9}$$

The oscillator sum is the same as in the noncompact case, while the momentum integration is replaced by a sum over n and w. Modular invariance is not manifest, but can be made so by using the Poisson resummation formula,

$$\sum_{n=-\infty}^{\infty} \exp(-\pi a n^2 + 2\pi i b n) = a^{-1/2} \sum_{m=-\infty}^{\infty} \exp\left[-\frac{\pi(m-b)^2}{a}\right]\,. \tag{8.2.10}$$

The partition function becomes

$$2\pi R Z_X(\tau) \sum_{m,w=-\infty}^{\infty} \exp\left(-\frac{\pi R^2|m - w\tau|^2}{\alpha'\tau_2}\right)\,. \tag{8.2.11}$$

Here $Z_X(\tau)$ is the modular-invariant expression (7.2.9) from the noncompact theory. The sum is obviously invariant under $\tau \to \tau + 1$, by making a change of variables $m \to m + w$. It is also invariant under $\tau \to -1/\tau$ with $m \to -w$, $w \to m$.

The expression (8.2.11) has a simple path integral interpretation. Summing over all genus one world-sheets in a periodic spacetime, each nontrivial closed curve on the world-sheet can be wound around the compact direction,

$$X(\sigma^1 + 2\pi, \sigma^2) = X(\sigma^1, \sigma^2) + 2\pi w R\,, \tag{8.2.12a}$$

$$X(\sigma^1 + 2\pi\tau_1, \sigma^2 + 2\pi\tau_2) = X(\sigma^1, \sigma^2) + 2\pi m R\,. \tag{8.2.12b}$$

That is, the path integral breaks up into topologically distinct sectors labeled by w and m. The Gaussian path integration can be done by the usual strategy of writing X as a classical solution of the correct periodicity,

$$X_{\mathrm{cl}} = \sigma^1 wR + \sigma^2(m - w\tau_1)R/\tau_2\,, \tag{8.2.13}$$

plus a quantum piece with periodic boundary conditions. The action separates into a classical piece plus a quantum piece. The path integral over the quantum piece is just as in the noncompact case, while the classical action appears as the exponent in eq. (8.2.11). The effect of the

modular transformations on the periodicities is then simply to change the summation variables m and w.

Vertex operators

In order to create a winding state, which has $\alpha_0 \neq \tilde{\alpha}_0$, we need independent variables x_L and x_R, with

$$[x_L, p_L] = [x_R, p_R] = i \,. \tag{8.2.14}$$

The field X then splits into holomorphic and antiholomorphic parts,

$$X(z,\bar{z}) = X_L(z) + X_R(\bar{z}) \,, \tag{8.2.15}$$

with

$$X_L(z) = x_L - i\frac{\alpha'}{2}p_L \ln z + i\left(\frac{\alpha'}{2}\right)^{1/2} \sum_{\substack{m=-\infty \\ m\neq 0}}^{\infty} \frac{\alpha_m}{mz^m} \,, \tag{8.2.16a}$$

$$X_R(\bar{z}) = x_R - i\frac{\alpha'}{2}p_R \ln \bar{z} + i\left(\frac{\alpha'}{2}\right)^{1/2} \sum_{\substack{m=-\infty \\ m\neq 0}}^{\infty} \frac{\tilde{\alpha}_m}{m\bar{z}^m} \,. \tag{8.2.16b}$$

If we restrict attention to states with $k_L = k_R$ and to operators constructed only from the sum $X_R(z)+X_L(\bar{z})$, this reduces to the CFT for a noncompact dimension.

Let us now discuss the OPE and vertex operators, first in a slightly heuristic way and then filling in the details. It is easy to guess what the operator products should be, since the usual XX operator product $-(\alpha'/2)\ln(z_{12}\bar{z}_{12})$ separates into a sum of holomorphic plus antiholomorphic functions. Thus,

$$X_L(z_1)X_L(z_2) \sim -\frac{\alpha'}{2}\ln z_{12} \,, \quad X_R(\bar{z}_1)X_R(\bar{z}_2) \sim -\frac{\alpha'}{2}\ln \bar{z}_{12} \,, \tag{8.2.17a}$$

$$X_L(z_1)X_R(\bar{z}_2) \sim 0 \,. \tag{8.2.17b}$$

The operator corresponding to the state $|0;k_L,k_R\rangle$ would be

$$\mathscr{V}_{k_Lk_R}(z,\bar{z}) = :e^{ik_LX_L(z)+ik_RX_R(\bar{z})}: \tag{8.2.18}$$

with OPE

$$\mathscr{V}_{k_Lk_R}(z_1,\bar{z}_1)\mathscr{V}_{k'_Lk'_R}(z_2,\bar{z}_2) \sim z_{12}^{\alpha' k_Lk'_L/2}\bar{z}_{12}^{\alpha' k_Rk'_R/2}\mathscr{V}_{(k+k')_L(k+k')_R}(z_2,\bar{z}_2) \,. \tag{8.2.19}$$

One encounters branch cuts in various expressions, which is not surprising because the field X is no longer single-valued. It is important, however, that the full vertex operator OPE is single-valued: the net phase when z_1 circles z_2 is

$$\exp[\pi i\alpha'(k_Lk'_L - k_Rk'_R)] = \exp[2\pi i(nw' + wn')] = 1 \,. \tag{8.2.20}$$

This is as it must be in order for the string amplitudes to be well defined.

A technicality

The above is essentially correct, but we have been cavalier with the location of the branch cut in the logarithm. One can see a problem in the OPE (8.2.19) if one interchanges the points z_1 and z_2 and the momenta k and k'. The left-hand side is symmetric, but the right-hand side changes by $\exp[\pi i(nw'+wn')]$ and so changes sign if $nw'+wn'$ is odd. One can also see the problem as follows. From the mode expansion one may derive the equal time ($|z_1| = |z_2|$) commutator

$$[X_L(z_1), X_L(z_2)] = \frac{\pi i\alpha'}{2}\mathrm{sign}(\sigma_1^1 - \sigma_2^1) \,. \tag{8.2.21}$$

The CBH formula then shows that if we define the operators simply as creation–annihilation ordered, then for $nw' + wn'$ odd the operators $\mathscr{V}_{k_Lk_R}$ and $\mathscr{V}_{k'_Lk'_R}$ anticommute rather than commute; the invisible branch cut (8.2.20) is separated into two visible ones. A correct oscillator expression for the vertex operator (there is some arbitrariness in the phase) is

$$\mathscr{V}_{k_Lk_R}(z,\bar{z}) = \exp[\pi i(k_L - k_R)(p_L + p_R)\alpha'/4] \,{}^{\circ}_{\circ} e^{ik_LX_L(z)+ik_RX_R(\bar{z})}{}^{\circ}_{\circ} \,, \tag{8.2.22}$$

where as usual the ps are momentum operators and the ks are numbers (the momenta carried by the given vertex operator). When $\mathscr{V}_{k_Lk_R}$ and $\mathscr{V}_{k'_Lk'_R}$ are commuted past each other, the additional factors, known as *cocycles*, give an additional phase

$$\exp\Big\{\pi i\Big[(k_L - k_R)(k'_L + k'_R) - (k'_L - k'_R)(k_L + k_R)\Big]\alpha'/4\Big\} = \exp[\pi i(nw' - wn')] \,, \tag{8.2.23}$$

removing the branch cuts. For most purposes one can ignore this complication and work with the simpler expressions of the previous paragraph; the cocycle affects only the relative signs of certain amplitudes.

The general X^μ path integral (6.2.18) factorizes in an obvious way into holomorphic times antiholomorphic, so that one can simply replace

$$\prod_{\substack{i,j=1 \\ i<j}}^{n} |z_{ij}|^{\alpha' k_ik_j} \to \prod_{\substack{i,j=1 \\ i<j}}^{n} z_{ij}^{\alpha' k_{Li}k_{Lj}/2}\bar{z}_{ij}^{\alpha' k_{Ri}k_{Rj}/2} \,, \tag{8.2.24}$$

and also replace the $2\pi\delta(\sum k)$ from the noncompact integral over x_0 by

$$2\pi R\delta_{\Sigma_i n_i,0}\delta_{\Sigma_i w_i,0} \,. \tag{8.2.25}$$

Again this is naive; the precise expression (8.2.22) gives some additional signs. The reader can readily calculate an amplitude with four vertex operators, which is similar to the Virasoro–Shapiro amplitude.

DDF operators

We have just seen that the exponential operator can be separated into holomorphic and antiholomorphic parts. This has many applications. We now use it to tie up a loose end from the end of chapter 4, the DDF operators.

We will work with light-cone coordinates $X^{\pm} = 2^{-1/2}(X^0 \pm X^1)$, whose holomorphic parts have the OPE (we drop the subscript L)

$$X^+(z)X^-(0) \sim \frac{\alpha'}{2} \ln z_{12} , \quad X^+(z)X^+(0) \sim X^-(z)X^-(0) \sim 0 . \tag{8.2.26}$$

It follows that the operators

$$V^i(nk_0, z) = \partial X^i(z) e^{ink_0 X^+(z)} (2/\alpha')^{1/2} \tag{8.2.27}$$

with i transverse are $(1,0)$ primary fields with OPE

$$V^i(nk_0, z)\, V^j(mk_0, 0) \sim -\frac{\delta^{ij}}{z^2} e^{i(n+m)k_0 X^+(0)} - \frac{ink_0 \delta^{ij}}{z} \partial X^+(0) e^{i(n+m)k_0 X^+(0)} . \tag{8.2.28}$$

Define the DDF operators by

$$A_n^i = \oint \frac{dz}{2\pi} V^i(nk_0, z) . \tag{8.2.29}$$

The residue of the z^{-1} term in the OPE (8.2.28) is a total derivative except when $n + m = 0$, and so the DDF operators satisfy an oscillator algebra

$$[A_m^i, A_n^j] = m\delta^{ij}\delta_{m,-n} \frac{\alpha' k_0 p^+}{2} , \tag{8.2.30}$$

which is an oscillator algebra, up to normalization.

Consider the action of the DDF operators on a state of given momentum q. The OPE of $V^i(nk_0, z)$ with the vertex operator for this state involves $z^{-\alpha' nk_0 q^-/2}$ and so is single-valued if we fix $k_0 = 2/\alpha' q^-$. The contour integral (8.2.29) then defines a sensible operator in this sector. The DDF operators, being integrals of $(1,0)$ tensors, commute with the Virasoro generators and so take physical states into physical states. It follows that we can build physical states by acting with the DDF operators on the oscillator ground state. The states obtained in this way are in one-to-one correspondence with the light-cone states, and so are a complete set of physical states. The DDF operators do not involve X^- and so these states have no α^-_m excitations. They are therefore the same states given by the more general but less explicit construction (4.4.19).

8.3 Closed strings and T-duality

Now that we are looking at string theory we take $D = 26$, with only X^{25} periodic at first. The mass-shell (L_0 and $\tilde{L}_0$) conditions are

$$\begin{aligned} m^2 = -k^\mu k_\mu &= (k_L^{25})^2 + \frac{4}{\alpha'}(N-1) \\ &= (k_R^{25})^2 + \frac{4}{\alpha'}(\tilde{N}-1) , \end{aligned} \tag{8.3.1}$$

or

$$m^2 = \frac{n^2}{R^2} + \frac{w^2 R^2}{\alpha'^2} + \frac{2}{\alpha'}(N + \tilde{N} - 2) , \tag{8.3.2a}$$

$$0 = nw + N - \tilde{N} . \tag{8.3.2b}$$

One recognizes four contributions to the mass-squared: those from the compact momentum, the potential energy of the winding string, the oscillators, and the zero-point energy. The discussion of the physical spectrum in chapter 4 was valid for an arbitrary compactification, so we get the correct counting of states by considering only the transverse oscillators $M = 2, \ldots, 25$.

Let us first recover the field theory results for the massless spectrum. At a generic value of R, the only way a state can be massless is if $n = w = 0$ and $N = \tilde{N} = 1$. These are the same 24^2 states as in the noncompact theory, but it is useful now to separate them according to whether the oscillations are in the spacetime directions μ or the internal direction 25:

$$\begin{aligned} &\alpha^\mu_{-1}\tilde{\alpha}^\nu_{-1}|0;k\rangle , \quad (\alpha^\mu_{-1}\tilde{\alpha}^{25}_{-1} + \alpha^{25}_{-1}\tilde{\alpha}^\mu_{-1})|0;k\rangle , \\ &(\alpha^\mu_{-1}\tilde{\alpha}^{25}_{-1} - \alpha^{25}_{-1}\tilde{\alpha}^\mu_{-1})|0;k\rangle , \quad \alpha^{25}_{-1}\tilde{\alpha}^{25}_{-1}|0;k\rangle . \end{aligned} \tag{8.3.3a}$$

The first of these breaks up further into a 25-dimensional graviton plus dilaton plus antisymmetric tensor. The second, a graviton with one spacetime and one internal index, is the Kaluza–Klein vector. The third state is the vector from the antisymmetric tensor. The final state is a scalar, the modulus for the radius of the compact direction. Its vertex operator, $:\partial X^{25}\bar{\partial}X^{25}e^{ik\cdot X}:$, is a perturbation of the metric $G_{25,25}$. This is the same massless spectrum as found by considering the low energy field theory.

It is interesting to consider the charges of massive states under the $U(1) \times U(1)$ symmetry. Again the Kaluza–Klein gauge symmetry is from X^{25}-translations, so the corresponding charge is the compact momentum p_{25}. For the antisymmetric tensor gauge symmetry, consider the zero-momentum vertex operator, which measures the charge of any state to which it couples. It is proportional to

$$\partial X^\mu \bar{\partial} X^{25} - \partial X^{25} \bar{\partial} X^\mu = \bar{\partial}(X^{25}\partial X^\mu) - \partial(X^{25}\bar{\partial}X^\mu) . \tag{8.3.4}$$

This is a total derivative, but it does not integrate to zero in a winding state because X^{25} is not single-valued. So the $B_{\mu,25}$ charge is the winding number. This is our first of several examples of 'stringy' physics in this compactification: in field theory there were no states carrying this charge.

Let us check the gauge couplings explicitly from the string three-point coupling. The gauge boson vertex operators are

$$\frac{2^{1/2}g_{c,25}}{\alpha'} :(\partial X^\mu \bar\partial X^{25} \pm \partial X^{25}\bar\partial X^\mu)e^{ik\cdot X}: \ . \tag{8.3.5}$$

We have defined the 25-dimensional string coupling $g_{c,25} = g_c(2\pi R)^{-1/2}$. The factor $(2\pi R)^{-1/2}$ is from normalizing the zero-mode wavefunction. The vertex operator for a tachyon with general compact momentum and winding is

$$g_{c,25}\ :e^{ik_L\cdot X_L(z)+ik_R\cdot X_R(\bar z)}: \ . \tag{8.3.6}$$

The three-point amplitude for one gauge boson and two tachyons, with the tachyons having nonzero compact momenta and winding, is similar to eq. (6.6.14),

$$\begin{aligned} &-2^{-1/2}\pi i g_{c,25}(2\pi)^{25}\delta^{25}(\textstyle\sum_i k_i)k_{23}^\mu(k_{L23}^{25} \pm k_{R23}^{25}) \\ &\to -2^{3/2}\pi i g_{c,25}(2\pi)^{25}\delta^{25}(\textstyle\sum_i k_i)k_2^\mu(k_{L2}^{25} \pm k_{R2}^{25}) \ . \end{aligned} \tag{8.3.7}$$

In the second line we have taken the gauge boson momentum $k_1 \to 0$, which defines the gauge coupling. The two gauge bosons thus couple respectively to $k_{L2}^{25} \pm k_{R2}^{25}$, the compact momentum and winding, as expected. By considering also the amplitude with a graviton replacing the gauge boson, one can recover the same relations (8.1.11), (8.1.12) between the various couplings as found from the effective action.

Enhanced gauge symmetries

Everything so far is almost exactly as it would be in field theory with one compact dimension. The 26-dimensional graviton gives rise to a 25-dimensional graviton plus vector plus modulus, and so on. The one stringy effect is that winding states carry the $B_{\mu,25}$ charge.

Further and more striking effects arise at special compactification radii. Our discussion of the massless spectrum omitted states that are massless only for special values of the radius R. The richest case is $R = \alpha'^{1/2}$, where $k_{L,R}^{25} = (n \pm w)\alpha'^{-1/2}$ and the condition for a massless state is

$$(n+w)^2 + 4N = (n-w)^2 + 4\tilde N \ = \ 4 \ . \tag{8.3.8}$$

In addition to the generic solution $n = w = 0$, $N = \tilde N = 1$, there are now also

$$n = w = \pm 1,\ N = 0,\ \tilde N = 1\ , \quad n = -w = \pm 1,\ N = 1,\ \tilde N = 0\ , \tag{8.3.9}$$

and

$$n = \pm 2,\ w = N = \tilde N = 0\ , \quad w = \pm 2,\ n = N = \tilde N = 0\ . \tag{8.3.10}$$

The states (8.3.9) include four new gauge bosons, with vertex operators

$$:\bar\partial X^\mu e^{ik\cdot X}\exp[\pm 2i\alpha'^{-1/2}X_L^{25}]:\ , \quad :\partial X^\mu e^{ik\cdot X}\exp[\pm 2i\alpha'^{-1/2}X_R^{25}]:\ . \tag{8.3.11}$$

The precise definition of the exponential operators is as discussed in the previous section. These states have internal momentum and winding number, so they carry the Kaluza–Klein and antisymmetric tensor gauge charges. The only consistent theory of charged massless vectors is non-Abelian gauge theory, so the new gauge bosons must combine with the old into a non-Abelian theory. It is now useful to work with the basis $\partial X^{25}\bar\partial X^\mu$ and $\partial X^\mu \bar\partial X^{25}$ for the earlier vectors. The first of these couples to k_L^{25}, under which the first pair of states in (8.3.11) carries charges ± 1 and the second pair is neutral. The second similarly couples to k_R^{25} and so to the other pair. This identifies the gauge group as $SU(2) \times SU(2)$, with the three vectors involving $\bar\partial X^\mu$ forming one $SU(2)$ and the three involving ∂X^μ forming the other.

To exhibit the $SU(2) \times SU(2)$ define the three $(1,0)$ currents

$$j^1(z) = :\cos[2\alpha'^{-1/2}X_L^{25}(z)]:\ , \tag{8.3.12a}$$

$$j^2(z) = :\sin[2\alpha'^{-1/2}X_L^{25}(z)]:\ , \tag{8.3.12b}$$

$$j^3(z) = i\partial X_L^{25}(z)/\alpha'^{1/2}\ . \tag{8.3.12c}$$

These have been normalized to have the OPE

$$j^i(z)j^j(0) \sim \frac{\delta^{ij}}{2z^2} + i\frac{\epsilon^{ijk}}{z}j^k(0)\ . \tag{8.3.13}$$

There is of course also a similar OPE of $(0,1)$ currents $\bar j^i$. The single pole term implies that the corresponding charges form an $SU(2)$ algebra. In fact, since the currents are holomorphic there is an infinite-dimensional algebra formed by the Laurent coefficients

$$j^i(z) = \sum_{m=-\infty}^{\infty}\frac{j_m^i}{z^{m+1}}\ , \tag{8.3.14a}$$

$$[j_m^i, j_n^j] = \frac{m}{2}\delta_{m,-n}\delta^{ij} + i\epsilon^{ijk}j_{m+n}^k\ , \tag{8.3.14b}$$

known variously as a *current algebra*, *affine Lie algebra*, or *Kac–Moody algebra*. We will encounter such algebras extensively beginning in chapter 11. It will be shown that the coefficient of the z^{-2} term is quantized, with the value in (8.3.13) being the minimum, so this is known as a *level one* $SU(2)$ current algebra.

The existence of this $SU(2)\times SU(2)$ symmetry is the first indication that string theory sees spacetime geometry differently from what we are used

to in field theory, where only a $U(1) \times U(1)$ symmetry is evident at any radius. The j^3 and $j^{1,2}$ currents appear to have very different origins, as oscillator excitations and as winding–momentum excitations respectively, and their expressions in terms of X^{25} are very different, but in their action on the string spectrum they are related to one another by symmetry. The extra vectors are massless because the energies of internal momentum and winding are offset by the negative zero-point energy. The reader might therefore suspect that these could arise only in theories with tachyons, but we will see otherwise in volume two.

Scales and couplings

At the $SU(2) \times SU(2)$ radius, the relation (8.1.11) becomes[2]

$$g_{25}^2 = 2\kappa_{25}^2/\alpha' \ . \tag{8.3.15}$$

This relation continues to hold if we compactify more dimensions, each side scaling the same way, so in particular in four dimensions

$$g_4^2 = 2\kappa_4^2/\alpha' \ . \tag{8.3.16}$$

In nature, the non-Abelian gauge couplings g_4^2 are within an order of magnitude of unity, so this implies that the string length $\alpha'^{1/2}$ is not too far from the gravitational length κ_4.

We can also think about this as follows. The four-dimensional gauge coupling is dimensionless, but the four-dimensional gravitational coupling is dimensional. Define an effective dimensionless gravitational coupling depending on the energy scale E,

$$\hat{g}_{G,4}^2(E) = \kappa_4^2 E^2 \ , \tag{8.3.17}$$

which increases quadratically with energy. Of course, the 'dimensionless' gauge couplings run with energy, but this is slow compared to the dimensional scaling of the gravitational coupling. Then by eq. (8.3.16), the string mass scale is where the gauge and gravitational couplings are roughly equal, $g_{G,4}^2(E) \approx g_4^2$.

It is also interesting to consider the possibility that in compactifying down to four dimensions, some of the compactified dimensions are larger than others. This can best be analyzed by starting at low energy and working upwards. At low energy there is a dimensionless gauge coupling g_4^2. At an energy ρ_5^{-1}, one compact dimension becomes visible and the physics

[2] To be precise, this was the coupling of an $(n, w) = (1, 0)$ state. The non-Abelian gauge bosons have $(|n|, |w|) = (1, 1)$ and so their coupling, which is the conventional particle physics definition of the $SU(2)$ coupling, is $g_{SU(2),25}^2 = 4\kappa_{25}^2/\alpha'$. We will see in chapter 18 that this holds for all level one current algebras.

becomes five-dimensional (for simplicity we consider only one dimension at this scale; with more dimensions the conclusion is strengthened). The effective five-dimensional coupling is

$$g_5^2 = 2\pi\rho_5 g_4^2 \ . \tag{8.3.18}$$

This follows from the effective action, as in eq. (8.1.12). The coupling g_5^2 has units of length. That is, five-dimensional Yang–Mills theory, like four-dimensional gravity, is a nonrenormalizable theory. Correspondingly, the effective dimensionless coupling

$$\hat{g}_5^2 = g_5^2 E = 2\pi\rho_5 E g_4^2 \tag{8.3.19}$$

grows linearly with energy. However, g_4^2 is not much less than one in nature, so the coupling quickly becomes strong at high energy; presumably string theory is then needed to make the short-distance behavior sensible. So the compactification scale must be close to the string scale, which as we have already learned is close to the gravitational scale. This analysis is modified somewhat in open string theory, and when the recent understanding of strongly coupled string theory is taken into account; it is possible, but unlikely, that string or Kaluza–Klein excitations lie far below the Planck scale.

Higgs mechanism

It is instructive to consider what happens as we move R away from the $SU(2) \times SU(2)$ radius. The extra gauge bosons now acquire masses

$$m = \frac{|R^2 - \alpha'|}{R\alpha'} \approx \frac{2}{\alpha'} \left| R - \alpha'^{1/2} \right| \ , \tag{8.3.20}$$

the approximation being for radii close to the $SU(2) \times SU(2)$ radius $\alpha'^{1/2}$. Close to this radius the mass is much less than the string scale and we should be able to understand the physics in terms of low energy field theory.

There is only one way to give such a mass to a gauge boson, spontaneous symmetry breaking, and that is just what is happening here. At $R = \alpha'^{1/2}$ there are ten massless scalars: the dilaton, the modulus $G_{25,25}$, the four states in (8.3.9) where a 25-oscillator is excited, and the four states (8.3.10). The last nine are created by the vertex operators

$$:j^i(z)\bar{j}^j(\bar{z})e^{ik\cdot X(z,\bar{z})}: \ . \tag{8.3.21}$$

The index i is a vector under the left-moving $SU(2)$ and the index j is a vector under the right-moving $SU(2)$. That is, these transform as **(3,3)** under $SU(2) \times SU(2)$. In particular, the modulus for the radius is $j^3\bar{j}^3$. Moving away from the $SU(2) \times SU(2)$ radius gives an expectation value

to this field and so breaks the gauge symmetry down to the $U(1) \times U(1)$ that leaves the 3-axes invariant. Indeed, this is the gauge group at generic radii. Near the $SU(2)$ radius, the mass (8.3.20) is linear in the magnitude $|R - \alpha'^{1/2}|$ of the modulus, as usual in spontaneous breaking.

Let us label the spacetime fields corresponding to the nine massless scalars as M_{ij}. A change in the modulus is $M_{33} \neq 0$, but it is instructive to look for more general backgrounds near the $SU(2) \times SU(2)$ point. The interactions of these massless fields will be described by some spacetime action, which will include a potential $U(M)$. These fields are massless at the symmetric point so there is no term of order M^2 in U, but an $(SU(2) \times SU(2))$-invariant cubic term is possible,

$$U(M) \propto \epsilon^{ijk}\epsilon^{i'j'k'} M_{ii'} M_{jj'} M_{kk'} = \det M \ . \tag{8.3.22}$$

It is not hard to verify from the string three-point amplitude that this term does appear. Being cubic in bosons it is unbounded below, but like the tachyon this is an irrelevant artifact of bosonic string theory.

It is interesting to look for static classical solutions, ignoring the fact that they are unstable to small variations. In order to have a static background solution, the graviton and dilaton field equations require the potential to vanish and the M_{ij} field equations require the potential to be stationary,

$$U(M) = \frac{\partial U(M)}{\partial M_{ij}} = 0 \ . \tag{8.3.23}$$

Using an $SU(2) \times SU(2)$ rotation to diagonalize M, the conditions (8.3.23) become

$$M_{11}M_{22}M_{33} = M_{11}M_{22} = M_{11}M_{33} = M_{22}M_{33} = 0 \ . \tag{8.3.24}$$

Thus, only one of the diagonal components can be nonvanishing. By an $SU(2) \times SU(2)$ rotation we can take this to be M_{33}. So we do not construct any physically new string backgrounds in this way.

A continuous family of static background solutions is known as a *flat direction*. Here there are nine massless scalars, which by an $SU(2) \times SU(2)$ rotation are reduced to the three diagonal fields, but there is only one flat direction (gauge-equivalent directions not being counted). Of course, we have only computed the cubic term in the potential, and in a general situation this could vanish and higher powers still be nonzero. To establish an exactly flat direction one needs some special argument. Here the special argument is that we can explicitly construct the free CFT for any value of R. Notice that if we treat $j^1\tilde{j}^1$ or $j^2\tilde{j}^2$ as the modulus the world-sheet action looks very complicated, with cosines and sines of X^{25}, but the world-sheet theory is solvable because it is equivalent to the free theory obtained by varying R.

T-duality

From the mass formula

$$m^2 = \frac{n^2}{R^2} + \frac{w^2R^2}{\alpha'^2} + \frac{2}{\alpha'}(N + \tilde{N} - 2) \ , \tag{8.3.25}$$

we see that as $R \to \infty$ winding states become infinitely massive, while the compact momenta go over to a continuous spectrum. This is as would be expected with a noncompact dimension. Looking at the opposite limit $R \to 0$, however, one gets a surprise. The states with compact momentum become infinitely massive, but the spectrum of winding states now approaches a continuum — it does not cost much energy to wrap a string around a small circle. Thus as the radius goes to zero the spectrum again seems to approach that of a *noncompact* dimension. This is quite different from the behavior in field theory, where there is compact momentum n but no winding number w, and no states become light as $R \to 0$.

In fact, the $R \to 0$ and $R \to \infty$ limits are physically identical. The spectrum (8.3.25) is invariant under

$$R \to R' = \frac{\alpha'}{R} \ , \quad n \leftrightarrow w \ . \tag{8.3.26}$$

This equivalence extends to the interactions as well. Notice that reversing n and w is the same as

$$p_L^{25} \to p_L^{25} \ , \quad p_R^{25} \to -p_R^{25} \ . \tag{8.3.27}$$

Consider the theory at radius R. Recalling the separation $X^{25}(z, \bar{z}) = X_L^{25}(z) + X_R^{25}(\bar{z})$, define

$$X'^{25}(z, \bar{z}) = X_L^{25}(z) - X_R^{25}(\bar{z}) \ . \tag{8.3.28}$$

The field X'^{25} has the same OPEs and energy momentum tensor as X^{25}, the minus signs always entering in pairs. The only change in the CFT from using X'^{25} in place of X^{25} is that it includes the sign change (8.3.27), which changes the spectrum from that of the theory with radius R to that of the theory with radius R'. That is, these are the same theory, one written in terms of X^{25} and one in terms of X'^{25}. It should be noted that the relation between $X^{25}(z, \bar{z})$ and $X'^{25}(z, \bar{z})$ is nonlocal on the world-sheet, because the separate left- and right-moving fields are nonlocal operators.

This equivalence is known as *T-duality*. That the $R \to 0$ limit is physically the same as the $R \to \infty$ limit is completely unlike the behavior of point particles, and is one more sign that strings see geometry very differently at short distances. The space of inequivalent theories is the half-line $R \geq \alpha'^{1/2}$. We could take instead the range $0 \leq R \leq \alpha'^{1/2}$ but it is more natural to think in terms of the larger of the two equivalent radii:

momentum continua are more familiar than winding number continua, and in particular questions of locality are clearer in the larger-R picture. Thus, there is no radius smaller than the self-dual radius

$$R_{\text{self-dual}} = R_{SU(2)\times SU(2)} = \alpha'^{1/2} . \tag{8.3.29}$$

This indication of a minimum distance scale of the order of the string length will recur elsewhere in string perturbation theory, but we will see that nonperturbatively there is structure at shorter distance.

It is important in many applications that T-duality acts nontrivially on the string dilaton Φ. Consider the scattering amplitude for gravitons without winding or compact momentum. These states are invariant under T-duality, and so must be the amplitude as well. The latter can be read from the low energy action (8.1.9), and so the 25-dimensional coupling κ_{25} must be invariant under duality. Since the 26-dimensional coupling is related to this by $\kappa = (2\pi\rho)^{1/2}\kappa_{25}$, duality must act as

$$\rho' = \frac{\alpha'}{\rho} , \quad \kappa' = \frac{\alpha'^{1/2}}{\rho}\kappa . \tag{8.3.30}$$

Since $\kappa \propto e^{\Phi}$, this implies

$$e^{\Phi'} = \frac{\alpha'^{1/2}}{\rho}e^{\Phi} . \tag{8.3.31}$$

T-duality is a symmetry, relating different states (backgrounds) in a single theory, and in fact it is a gauge symmetry. We saw that the modulus δR is the 33-component of a (3,3) field. A rotation by π around the 1-axis in *one* of the $SU(2)$s reverses the sign of this modulus, so decreasing R from the $SU(2) \times SU(2)$ radius is gauge-equivalent to increasing it. Thus the $\mathbf{Z}_2$ symmetry of duality is a small part of the $SU(2) \times SU(2)$ gauge symmetry. This implies further that duality is a symmetry not only of string perturbation theory but of the *exact* theory. If we have a massless gauge boson to leading approximation, even small explicit violations of the symmetry lead to inconsistency (spontaneous violation is not an issue, as T-duality is already spontaneously broken away from the self-dual radius).

This last argument illustrates an important idea, the complementary use of stringy and spacetime arguments. The existence of the enhanced gauge symmetry is a purely stringy phenomenon. We do not have a nonperturbative understanding of string theory, but we do know a great deal about low energy field theory, even nonperturbatively, and we can use this.

8.4 Compactification of several dimensions

It is useful to generalize the analysis to k periodic dimensions,

$$X^m \cong X^m + 2\pi R , \quad 26-k \le m \le 25 . \tag{8.4.1}$$

Let $d = 26-k$ be the number of noncompact dimensions. Spacetime is now $M^d \times T^k$. The coordinate periodicity (8.4.1) is being held fixed, but the actual geometry of the k-torus depends on internal metric G_{mn}. With more than one compact dimension, the antisymmetric tensor also has scalar components B_{mn}. The total number of scalars from both sources is k^2. There are also Kaluza–Klein gauge bosons A_μ^m and antisymmetric tensor gauge bosons $B_{m\mu}$. The low energy action, obtained as in section 8.1, is

$$\begin{aligned} S = \frac{(2\pi R)^k}{2\kappa_0^2}\int d^d x\,(-G_d)^{1/2}e^{-2\Phi_d}\Big[& R_d + 4\partial_\mu\Phi_d\partial^\mu\Phi_d \\ & -\frac{1}{4}G^{mn}G^{pq}(\partial_\mu G_{mp}\partial^\mu G_{nq} + \partial_\mu B_{mp}\partial^\mu B_{nq}) \\ & -\frac{1}{4}G_{mn}F^m_{\mu\nu}F^{n\mu\nu} - \frac{1}{4}G^{mn}H_{m\mu\nu}H_n^{\mu\nu} - \frac{1}{12}H_{\mu\nu\lambda}H^{\mu\nu\lambda}\Big] , \end{aligned} \tag{8.4.2}$$

where $\Phi_d = \Phi - \frac{1}{4}\ln\det G_{mn}$.

The string spectrum

The main new issue is the antisymmetric tensor background B_{mn}. The contribution of this term to the world-sheet Lagrangian density is proportional to

$$B_{mn}\partial_a(\epsilon^{ab}X^m\partial_b X^n) , \tag{8.4.3}$$

which is a total derivative for constant B_{mn}. It thus has no effect locally and the world-sheet theory is still a CFT, a good string background. It does, however, change the spectrum. We will see this in two ways, canonically and via the path integral.

In the canonical approach, focus on the zero-mode contribution to the world-sheet action. Inserting

$$X^m(\sigma^1,\sigma^2) = x^m(\sigma^2) + w^m R\sigma^1 \tag{8.4.4}$$

into the work' sheet action gives

$$L = \frac{1}{2\alpha'}G_{mn}(\dot{x}^m\dot{x}^n + w^m w^n R^2) + \frac{i}{\alpha'}B_{mn}\dot{x}^m w^n R . \tag{8.4.5}$$

A dot indicates a derivative with respect to world-sheet time σ^2. The canonical momenta are

$$p_m = -\frac{\partial L}{\partial v^m} = \frac{1}{\alpha'}(G_{mn}v^n - B_{mn}w^n R) , \tag{8.4.6}$$

where $v^m = i\dot{x}^m$. Some unfamiliar signs and is appear because we are using Euclidean world-sheet time. Continuing to Minkowski time, v^m becomes the velocity $\partial_0 x^m$. The periodicity of the wavefunction implies the quantization of the canonical momenta, $p_m = n_m/R$, and so

$$v_m = \alpha' \frac{n_m}{R} + B_{mn} w^n R \,. \tag{8.4.7}$$

The zero-mode contribution to the world-sheet Hamiltonian is

$$\frac{1}{2\alpha'} G_{mn}(v^m v^n + w^m w^n R^2) \,, \tag{8.4.8}$$

and the closed string mass is

$$m^2 = \frac{1}{2\alpha'^2} G_{mn}(v_L^m v_L^n + v_R^m v_R^n) + \frac{2}{\alpha'}(N + \tilde{N} - 2) \,, \tag{8.4.9a}$$

$$v_{L,R}^m = v^m \pm w^m R \,. \tag{8.4.9b}$$

The B_{mn} background thus shifts the masses of the winding states through the dependence of v^m on B_{mn}. The $L_0 - \tilde{L}_0$ constraint is

$$\begin{aligned} 0 &= G_{mn}(v_L^m v_L^n - v_R^m v_R^n) + 4\alpha'(N - \tilde{N}) \\ &= 4\alpha'(n_m w^m + N - \tilde{N}) \,. \end{aligned} \tag{8.4.10}$$

Alternatively, consider the torus path integral for the partition function. The antisymmetric tensor term is locally a total derivative and so depends only on the topology of the configuration. Consider the configuration

$$X^m = (w_1^m \sigma^1 + w_2^m \sigma^2) R \,. \tag{8.4.11}$$

The spatial direction on the torus winds w_1^m times around each periodic spacetime dimension x^m, and the timelike dimension w_2^m times. The B_{mn} world-sheet action is

$$2\pi i b_{mn} w_1^m w_2^n \,, \tag{8.4.12}$$

where $b_{mn} = B_{mn} R^2/\alpha'$. One can check that the sum over path integral sectors with phase (8.4.12) is related by Poisson resummation to the partition function for the shifted spectrum (8.4.9).

In the description above the metric G_{mn} appears in the action for the world-sheet fields. For vertex operator calculations it is more convenient to keep the usual action. Write the metric in terms of a tetrad e_m^r,

$$G_{mn} = e_m^r e_n^r \,, \tag{8.4.13}$$

with tangent space indices $r, s, \ldots$. The coordinates $X^r = e_m^r X^m$ then have the standard OPEs. The vertex operator momenta are

$$k_{rL} = e_r^m \frac{v_{mL}}{\alpha'} \,, \quad k_{rR} = e_r^m \frac{v_{mR}}{\alpha'} \,, \tag{8.4.14}$$

with the inverse tetrad e_r^m. The mass-shell conditions become

$$m^2 = \frac{1}{2}(k_{rL} k_{rL} + k_{rR} k_{rR}) + \frac{2}{\alpha'}(N + \tilde{N} - 2) \,, \tag{8.4.15a}$$

$$0 = \alpha'(k_{rL} k_{rL} - k_{rR} k_{rR}) + 4(N - \tilde{N}) \,. \tag{8.4.15b}$$

We use the coordinates X^r implicitly in any discussion of vertex operators, such as the one that now follows.

Narain compactification

There is an elegant description of the general toroidal compactification. Consider the winding state vertex operators $e^{ik_L \cdot X_L + ik_R \cdot X_R}$. For any given compactification, the spectrum of momenta (k_{rL}, k_{rR}) forms a lattice in a $2k$-dimensional momentum space $\mathbf{R}^{2k}$. That is, the momentum spectrum consists of all integer linear combinations of $2k$ linearly independent basis vectors. Let us work with the dimensionless momenta $l = k(\alpha'/2)^{1/2}$ and call the corresponding lattice Γ. The OPE of two vertex operators is

$$\begin{aligned} &:e^{ik_L \cdot X_L(z) + ik_R \cdot X_R(\bar{z})}: \, :e^{ik'_L \cdot X_L(0) + ik'_R \cdot X_R(0)}: \\ &\quad \sim z^{l_L \cdot l'_L} \bar{z}^{l_R \cdot l'_R} \, :e^{i(k_L + k'_L) \cdot X_L(0) + i(k_R + k'_R) \cdot X_R(0)}: \,. \end{aligned} \tag{8.4.16}$$

As one vertex operator circles the other, the product picks up the phase $\exp[2\pi i(l_L \cdot l'_L - l_R \cdot l'_R)]$. Single-valuedness of the operator product thus requires

$$l_L \cdot l'_L - l_R \cdot l'_R \equiv l \circ l' \in \mathbf{Z} \,, \tag{8.4.17}$$

for all l and l' in Γ. We have defined the product $\circ$, which has signature (k,k) in $\mathbf{R}^{2k}$. The dual lattice Γ^* is defined to be the set of points in $\mathbf{R}^{2k}$ that have integer $\circ$ product with all points in Γ. The single-valuedness condition (8.4.17) is then the statement that

$$\Gamma \subset \Gamma^* \,. \tag{8.4.18}$$

Modular invariance also constrains Γ. In section 7.2 it was shown that invariance under $\tau \to \tau + 1$ requires that $L_0 - \tilde{L}_0$ be an integer for all states in the string CFT. Then

$$l \circ l \in 2\mathbf{Z} \quad \text{for all } l \in \Gamma \,. \tag{8.4.19}$$

The condition (8.4.19) actually implies condition (8.4.17), given closure of the OPE

$$2l \circ l' = (l + l') \circ (l + l') - l \circ l - l' \circ l' \in 2\mathbf{Z} \,. \tag{8.4.20}$$

Modular invariance under $\tau \to -1/\tau$ requires a more detailed calculation. The partition function for the compact dimensions is

$$Z_\Gamma(\tau) = |\eta(\tau)|^{-2k} \sum_{l \in \Gamma} \exp(\pi i \tau l_L^2 - \pi i \bar{\tau} l_R^2) \,. \tag{8.4.21}$$

To determine the transformation property of Z_Γ use again Poisson resummation. Write

$$\sum_{l'\in\Gamma}\delta(l-l') = V_\Gamma^{-1}\sum_{l''\in\Gamma^*}\exp(2\pi i l''\circ l)\,. \tag{8.4.22}$$

Here, V_Γ is the volume of a unit cell of the lattice Γ. Eq. (8.4.22) is familiar from Fourier series: the phase averages to zero except when $l\in\Gamma$, and the normalization is fixed by integrating over a unit cell. Using the representation (8.4.22), the sum (8.4.21) can be written

$$\begin{aligned} Z_\Gamma(\tau) &= V_\Gamma^{-1}|\eta(\tau)|^{-2k}\sum_{l''\in\Gamma^*}\int d^{2k}l\,\exp(2\pi i l''\circ l+\pi i\tau l_L^2-\pi i\bar\tau l_R^2)\\ &= V_\Gamma^{-1}(\tau\bar\tau)^{-k/2}|\eta(\tau)|^{-2k}\sum_{l''\in\Gamma^*}\exp(-\pi i l_L''^2/\tau+\pi i l_R''^2/\bar\tau)\\ &= V_\Gamma^{-1}Z_{\Gamma^*}(-1/\tau)\,, \end{aligned} \tag{8.4.23}$$

where in the last line we have used the modular transformation (7.2.44) of the eta function. Then the condition

$$\Gamma=\Gamma^* \tag{8.4.24}$$

is sufficient for modular invariance, because $V_\Gamma = V_{\Gamma^*}^{-1}$ and so is equal to unity if (8.4.24) holds. A little thought shows that this is also necessary if modular invariance is to hold for all τ.

The consistency conditions can be summarized as the requirement that Γ be an even self-dual lattice of signature (k,k), *even* referring to the property (8.4.19) and *self-dual* to the property (8.4.24).

All such lattices have been classified. Note that the consistency conditions (8.4.19) and (8.4.24) depend on the momenta l only through the $\circ$ product, which is invariant under Lorentz boosts of the $2k$-dimensional space, $O(k,k,\mathbf{R})$ transformations. If Γ is an even self-dual lattice then so is

$$\Gamma' = \Lambda\Gamma\,, \tag{8.4.25}$$

which is obtained by rotating all the points of the lattice by a common $O(k,k,\mathbf{R})$ transformation Λ. Now, it is important that $O(k,k,\mathbf{R})$ is *not* a symmetry of the theory. The mass-shell condition and operator products involve the separate dot products $l_L\cdot l_L'$ and $l_R\cdot l_R'$ and so are invariant only under the $O(k,\mathbf{R})\times O(k,\mathbf{R})$ that acts on l_L and l_R separately. Thus, most $O(k,k,\mathbf{R})$ transformations produce inequivalent theories.

Consider for example $k=1$, where

$$l_{L,R} = \frac{n}{r}\pm\frac{mr}{2} \tag{8.4.26}$$

in terms of the dimensionless radius $r = R(2/\alpha')^{1/2}$. This is indeed an even self-dual lattice. The boost

$$l_L' = l_L\cosh\lambda + l_R\sinh\lambda\,,\quad l_R' = l_L\sinh\lambda + l_R\cosh\lambda \tag{8.4.27}$$

changes this to the lattice for $r' = re^{-\lambda}$.

All even self-dual lattices of given Lorentzian signature can be obtained from any single lattice by $O(k,k,\mathbf{R})$ transformations. Start with a given solution, for example all compact dimensions orthogonal and at the $SU(2)$ radius with $B_{mn}=0$. The corresponding momentum lattice Γ_0 is known as kP_2 and gives the gauge group $SU(2)^{2k}$. As discussed above, two $O(k,k,\mathbf{R})$ transformations Λ and $\Lambda'\Lambda$ give equivalent string theories if $\Lambda'\in O(k,\mathbf{R})\times O(k,\mathbf{R})$, so the space of inequivalent theories obtained in this way is

$$\frac{O(k,k,\mathbf{R})}{O(k,\mathbf{R})\times O(k,\mathbf{R})}\,, \tag{8.4.28}$$

up to some discrete identifications to be discussed shortly. This is equivalent to the earlier description in terms of the backgrounds of G_{mn} and B_{mn}. In particular, the number of parameters in the coset (8.4.28) is

$$\frac{2k(2k-1)}{2} - k(k-1) = k^2\,, \tag{8.4.29}$$

the same as the number of components in G_{mn} and B_{mn}.

The T-duality symmetry is much enlarged from the case of one compactified dimension. In the abstract description, there is some discrete subgroup of $O(k,k,\mathbf{R})$ that takes the initial lattice Γ_0 into itself (though it permutes the individual points of the lattice). This subgroup is conventionally but somewhat inaccurately denoted $O(k,k,\mathbf{Z})$. If Λ'' is in this group, then clearly $\Lambda\Gamma_0$ and $\Lambda\Lambda''\Gamma_0$ are the same lattice. In all we have the equivalence

$$\Lambda\Gamma_0 \cong \Lambda'\Lambda\Lambda''\Gamma_0\,, \tag{8.4.30a}$$

$$\Lambda'\in O(k,\mathbf{R})\times O(k,\mathbf{R})\,,\quad \Lambda''\in O(k,k,\mathbf{Z})\,. \tag{8.4.30b}$$

The space of inequivalent lattices, and inequivalent backgrounds, is then

$$\frac{O(k,k,\mathbf{R})}{O(k,\mathbf{R})\times O(k,\mathbf{R})\times O(k,k,\mathbf{Z})}\,. \tag{8.4.31}$$

One should remember that the continuous denominator group acts by left-multiplication and the discrete group by right-multiplication, as in eq. (8.4.30a).

In terms of the backgrounds, the T-duality group $O(k,k,\mathbf{Z})$ contains transformations of several types. One is $R\to\alpha'/R$ dualities on the individual axes. Another is large spacetime coordinate transformations respecting

the periodicity,

$$x'^m = L^m{}_n x^n \tag{8.4.32}$$

with $L^m{}_n$ integers and $\det L = 1$ for invertibility. This is the group $SL(k,\mathbf{Z})$. Finally there are shifts of the antisymmetric tensor background,

$$b_{mn} \to b_{mn} + N_{mn} \tag{8.4.33}$$

with N_{mn} integers. From either the canonical result (8.4.7) or the path integral phase (8.4.12) it follows that this leaves the spectrum invariant. Together these generate the whole T-duality group.

Identifying first by the left-multiplication produces the space (8.4.28), and one can think of the discrete right-multiplication as acting on this covering space. For generic Λ, there are no $\Lambda_{1,2}$ in the respective denominator groups such that $\Lambda_1\Lambda\Lambda_2 = \Lambda$, so that there are no T-dualities that leave this point on the covering space fixed. At special Λ there are solutions to $\Lambda_1\Lambda\Lambda_2 = \Lambda$, and these will be fixed points of the T-duality element Λ_2. In the one-dimensional case we had such a fixed point, which was also a point of enhanced gauge symmetry. In the higher-dimensional case, besides products of $SU(2)$s, larger gauge groups arise at various points. We will defer further discussion of this to chapter 11.

When there is a moduli space parameterized by scalar fields ϕ^i, the kinetic term

$$-\frac{1}{2} g_{ij}(\phi)\partial_\mu\phi^i\partial^\mu\phi^j \tag{8.4.34}$$

defines a natural *metric on moduli space*. To be precise, an unambiguous definition requires that one first perform a spacetime Weyl transformation as in eq. (3.7.25) so as to make the coefficient of the gravitational action moduli-independent and eliminate mixing between the moduli and the spacetime metric. In the low energy action (8.4.2), the resulting scalar kinetic terms are proportional to

$$\frac{16}{d-2}\partial_\mu\Phi\partial^\mu\Phi + G^{mn}G^{pq}(\partial_\mu G_{mp}\partial^\mu G_{nq} + \partial_\mu B_{mp}\partial^\mu B_{nq}) \,. \tag{8.4.35}$$

The coset space (8.4.31) has a unique $O(k,k,\mathbf{R})$-invariant metric, which is just the second term in eq. (8.4.35). This $O(k,k,\mathbf{R})$ is not a symmetry of the full theory; only its discrete T-duality subgroup $O(k,k,\mathbf{Z})$ is. The difference between the two is not visible at low energy — it comes from the quantization of the string zero modes and so affects the massive spectrum. Thus $O(k,k,\mathbf{R})$ is an accidental symmetry of the low energy theory. In this example the moduli space is a product of the dilaton moduli space and the moduli space of compactifications.

An example

The case of two compact dimensions is a good example. The four moduli $G_{24,24}$, $G_{24,25}$, $G_{25,25}$, and $B_{24,25}$ are usefully joined into two complex fields $\tau = \tau_1 + i\tau_2$ and $\rho = \rho_1 + i\rho_2$. Define

$$\begin{aligned}\rho &= \frac{R^2}{\alpha'}\left(B_{24,25} + i\det{}^{1/2}G_{mn}\right)\\ &= b_{24,25} + i\frac{V}{4\pi^2\alpha'} \,,\end{aligned} \tag{8.4.36}$$

where V is the volume of the compact two-torus. Parameterize the compact metric as

$$ds^2 = \frac{\alpha'\rho_2}{R^2\tau_2}\left|dX^{24} + \tau dX^{25}\right|^2 \,. \tag{8.4.37}$$

The coordinate transformations (8.4.32) act on the spacetime two-torus just as the modular group acts on the world-sheet two-torus, generating a $PSL(2,\mathbf{Z})$ on τ and leaving ρ invariant. The antisymmetric tensor shift (8.4.33) is $\rho \to \rho + N_{24,25}$ with τ invariant. A simultaneous T-duality on $X^{24,25}$ acts as $\rho \to -1/\rho$ with τ invariant. Together the last two transformations generate a second $PSL(2,\mathbf{Z})$, acting on ρ. In addition, duality on X^{24} alone takes $(\tau,\rho) \to (\rho,\tau)$ and spacetime parity $X^{24} \to -X^{24}$ takes $(\tau,\rho) \to (-\bar\tau,-\bar\rho)$. The full T-duality group is then

$$PSL(2,\mathbf{Z}) \times PSL(2,\mathbf{Z}) \times \mathbf{Z}_2^2 \,. \tag{8.4.38}$$

A further equivalence comes from world-sheet parity, which takes $(\tau,\rho) \to (\tau,-\bar\rho)$. It reverses the sign of the $\circ$ product.

The moduli space is therefore two copies of the moduli space of the torus, together with some discrete identifications. The kinetic terms are proportional to

$$\frac{\partial_\mu\tau\partial^\mu\bar\tau}{\tau_2^2} + \frac{\partial_\mu\rho\partial^\mu\bar\rho}{\rho_2^2} \,. \tag{8.4.39}$$

The modular-invariant combination $d^2\tau/\tau_2^2$ was encountered in the discussion of the genus-zero amplitude. The continuous group $PSL(2,\mathbf{R})$ carries any τ in the upper half-plane into any other, and a $U(1)$ subgroup leaves each τ invariant. Thus the upper half-plane can be regarded as $PSL(2,\mathbf{R})/U(1)$, and the full moduli space is

$$\frac{PSL(2,\mathbf{R}) \times PSL(2,\mathbf{R})}{U(1) \times U(1) \times PSL(2,\mathbf{Z}) \times PSL(2,\mathbf{Z}) \times \mathbf{Z}_2^2} \,. \tag{8.4.40}$$

This is the same as the earlier (8.4.31). In particular, $PSL(2,\mathbf{R}) \times PSL(2,\mathbf{R})$ is locally the same as $O(2,2,\mathbf{R})$, in parallel to the fact that $SU(2) \times SU(2)$ is locally the same as $O(4,\mathbf{R})$.

8.5 Orbifolds

Rather than the periodic identification of X^{25}, consider instead identifying points under the reflection

$$X^{25} \cong -X^{25} . \tag{8.5.1}$$

Now a fundamental region is the half-line $X^{25} \geq 0$, and the hyperplane $X^{25} = 0$ becomes a boundary. This hyperplane is distinguished by the fact that its points are fixed under the reflection (8.5.1). More generally we could identify spacetime under the simultaneous reflection of k coordinates

$$X^m \to -X^m , \quad 26-k \leq m \leq 25 . \tag{8.5.2}$$

Again there is a space of fixed points, at $X^{26-k} = \ldots = X^{25} = 0$. For $k \geq 2$ this space is a singularity. In particular, for $k = 2$ it is a conic singularity with deficit angle π.

One can form a compact space by combining the reflection (8.5.1) with the periodic identification $X^{25} \cong X^{25} + 2\pi R$; we will call these r and t respectively. One can think of first identifying by t to form the circle S^1, and then identifying points on the circle under r. The resulting compact space is a line segment, with $0 \leq X^{25} \leq \pi R$ as a fundamental region. There are fixed points at both ends of the segment, the point πR being fixed by tr. In all, one is identifying the real line under the group

$$t^m : \quad X^{25} \cong X^{25} + 2\pi R m , \tag{8.5.3a}$$

$$t^m r : \quad X^{25} \cong 2\pi R m - X^{25} , \tag{8.5.3b}$$

for any integer m.

Similarly one could identify the k-torus under the reflection (8.5.2). In that case there are 2^k distinct fixed points, with each of the X^m being either 0 or πR. The singular spaces obtained in this way are known as *orbifolds*. The space obtained by the identification (8.5.2) in noncompact space is $\mathbf{R}^k/\mathbf{Z}_2$, while the compact space obtained by the same identification on the k-torus is $T^k/\mathbf{Z}_2$.

It is not obvious that string theory on such singular spaces should make sense, but we will see that it does. It is not so different from toroidal compactification, but it breaks more symmetry (and, later on, supersymmetry) and so will be of interest in constructing string models.

The identification (8.5.1) or (8.5.2) has two effects. The first is that the string wavefunction must be invariant under the reflection, equal at identified points. Second, there is a new sector in the closed string spectrum, in which

$$X^{25}(\sigma^1 + 2\pi) = -X^{25}(\sigma^1) , \tag{8.5.4}$$

since these are the same point in the identified spacetime. Strings in this sector are known as *twisted states*. This is directly parallel to periodic identification. There, invariance of the wavefunction made the momentum discrete, while strings that closed only up to the periodic identification (winding strings) appeared. One can again make a pictorial argument, analogous to figure 8.1, to show that twisted strings can be produced from untwisted ones. We will also see below that modular invariance requires them in the spectrum.

We will focus on the compact one-dimensional orbifold $S^1/\mathbf{Z}_2$. Examining first the untwisted sector, the spectrum of the periodic theory is reduced by the projection onto invariant states. The effect of r on a general state is

$$|N, \tilde{N}; k^\mu, n, w\rangle \to (-1)^{N^{25}+\tilde{N}^{25}} |N, \tilde{N}; k^\mu, -n, -w\rangle , \tag{8.5.5}$$

in particular reversing the compact winding and momentum. We must form linear combinations that are invariant under this. The states that are massless at generic R have $n = w = 0$, so the projection simply requires that the number of 25-excitations be even. The spacetime graviton, antisymmetric tensor, dilaton, and tachyon survive this projection. The modulus $\alpha_{-1}^{25}\tilde{\alpha}_{-1}^{25}|0; k^\mu, 0, 0\rangle$ also survives — this could have been anticipated, since the radius R can take any value. However, the Kaluza–Klein gauge bosons are no longer in the spectrum. The fate of the extra massless states at the self-dual point is quite interesting, and we will discuss it in detail later.

In the sector twisted by r, X^{25} is antiperiodic and so has a half-integral mode expansion

$$X^{25}(z, \bar{z}) = i\left(\frac{\alpha'}{2}\right)^{1/2} \sum_{m=-\infty}^{\infty} \frac{1}{m+1/2} \left(\frac{\alpha_{m+1/2}^{25}}{z^{m+1/2}} + \frac{\tilde{\alpha}_{m+1/2}^{25}}{\bar{z}^{m+1/2}} \right) . \tag{8.5.6}$$

The antiperiodicity forbids any center-of-mass coordinate or momentum, so the string cannot move away from the $X^{25} = 0$ fixed point: to the extent that the oscillations are small, one has $X^{25}(z, \bar{z}) \approx 0$. There are also twisted strings localized at the other fixed point,

$$X^{25}(\sigma^1 + 2\pi) = 2\pi R - X^{25}(\sigma^1) . \tag{8.5.7}$$

The fields change by tr as one circles the loop. The mode expansion is as above with an additional constant term πR. All other fixed points $n\pi R$ are images of one of these under t.

According to the zero-point mnemonic from section 2.9, the zero-point energy of X^{25} is shifted from $-\frac{1}{24}$ for a periodic boson to $\frac{1}{48}$ for an antiperiodic boson, for a net $+\frac{1}{16}$. The mass-shell conditions in the twisted

sectors are then

$$m^2 = \frac{4}{\alpha'}\left(N - \frac{15}{16}\right), \quad N = \tilde{N}\,. \tag{8.5.8}$$

Also, the 25-oscillators make half-integral contributions to the level N. The r-projection again requires the total number of 25-excitations to be even, which can also be deduced from the level-matching condition $N = \tilde{N}$. In the twisted sectors the ground states are $|T_{1,2}\rangle$, the subscript referring to the two fixed points. They are tachyonic, as are the first excited states $\alpha^{25}_{-1/2}\tilde{\alpha}^{25}_{-1/2}|T_{1,2}\rangle$. There are no massless twisted states. The extension to the more general case (8.5.2) is straightforward. The quantization is perfectly consistent in spite of the singularities in spacetime.

The one complication with orbifolds is that the twisted state vertex operators are not as simple as those for winding states; there is no such explicit formula as the exponential of a free field. For two twisted strings and any number of untwisted ones, the tree-level amplitude can be written in operator form $\langle T|\mathscr{V}_U\mathscr{V}_U\ldots|T\rangle$, which uses only untwisted vertex operators. For four or more twisted vertex operators (the number must be even for the path integral to make sense) two methods have been used. One is the stress tensor method, which we used in section 7.2 to get the partition function on the torus. The other is to go to the covering space on which X^{25} is single-valued (this is a genus g surface for $2g+2$ twisted vertex operators on the sphere) and evaluate the path integral.

There is a simple result for the tree-level couplings of untwisted states. On the sphere, with all external states untwisted, the twists and projections do not enter into the calculation at all and the amplitudes are the same as in the untwisted theory. For example, the low energy effective action for the massless untwisted fields is the same as we found for toroidal compactification, with the omission of the vector fields. This property is often useful in analyzing the physics of orbifold theories; it is called the *inheritance principle.*

Consider now the partition function of the X^{25} CFT. In the untwisted sector, one is to project the spectrum onto $r = +1$ states,

$$(q\bar{q})^{-1/24}\mathrm{Tr}_U\left(\frac{1+r}{2}q^{L_0}\bar{q}^{\tilde{L}_0}\right). \tag{8.5.9}$$

The first term in the projection operator just gives half the partition function (8.2.9) for toroidal compactification. In the term with r in the trace, the diagonal elements of r must have $n = w = 0$, as one sees from eq. (8.5.5). The $(-1)^{N^{25}+\tilde{N}^{25}}$ changes the oscillator sum to $1-q+q^2-q^3+\ldots = (1+q)^{-1}$. The partition function for the untwisted states is then

$$\frac{1}{2}Z_{\mathrm{tor}}(R,\tau) + \frac{1}{2}(q\bar{q})^{-1/24}\prod_{m=1}^{\infty}|1+q^m|^{-2}\,. \tag{8.5.10}$$

The contribution of the twisted sectors is from the sum over half-integer moded oscillations,

$$(q\bar{q})^{1/48}\mathrm{Tr}_T\left(\frac{1+r}{2}q^{L_0}\bar{q}^{\tilde{L}_0}\right)$$
$$= (q\bar{q})^{1/48}\left[\prod_{m=1}^{\infty}\left|1-q^{m-1/2}\right|^{-2} + \prod_{m=1}^{\infty}\left|1+q^{m-1/2}\right|^{-2}\right], \tag{8.5.11}$$

the factor of 2 from the number of twisted sectors offsetting the $\frac{1}{2}$ in the projection operator.

In all, the partition function can be written as

$$Z_{\mathrm{orb}}(R,\tau) = \frac{1}{2}Z_{\mathrm{tor}}(R,\tau) + \left|\frac{\eta(\tau)}{\vartheta_{10}(0,\tau)}\right| + \left|\frac{\eta(\tau)}{\vartheta_{01}(0,\tau)}\right| + \left|\frac{\eta(\tau)}{\vartheta_{00}(0,\tau)}\right|. \tag{8.5.12}$$

The first term is already known to be modular-invariant, and it follows from the modular transformation properties given in section 7.2 that the sum of the final three is invariant as well. In terms of the path integral on the torus, $Z_{\mathrm{tor}}(R,\tau)$ is from fields that are periodic up to translations. The term with ϑ_{ab} is from the path integral with

$$X^{25}(\sigma^1+2\pi,\sigma^2) = (-1)^{a+1}X^{25}(\sigma^1,\sigma^2)\,, \tag{8.5.13a}$$
$$X^{25}(\sigma^1+2\pi\tau_1,\sigma^2+2\pi\tau_2) = (-1)^{b+1}X^{25}(\sigma^1,\sigma^2)\,. \tag{8.5.13b}$$

For example, the ϑ_{10} term is from the untwisted sector with r in the trace, that is, from fields that are periodic in the spatial direction and antiperiodic in the time direction. The ϑ_{01} term, from the twisted sector with 1 in the trace, has the opposite boundary conditions. These are interchanged by $\tau \to -1/\tau$. Thus, the twisted sectors are necessary for modular invariance.

Twisting

Toroidal and orbifold compactifications are two examples of a useful construction, known as *twisting,* for constructing new string theories from old. In a given CFT whose symmetries include some discrete group H, we can form a new CFT in two steps. First add *twisted sectors,* in which the closed string world-sheet fields are periodic only up to some transformation $h \in H$,

$$\phi(\sigma^1+2\pi) = h\cdot\phi(\sigma^1)\,, \tag{8.5.14}$$

where ϕ denotes a generic world-sheet field. Then, restrict the spectrum to H-invariant states. Although the fields (8.5.14) are no longer periodic, the vertex operators are periodic because of the projection. So the OPE is still well defined, and it closes because the product of H-invariant operators is H-invariant.

In the partition function, we sum over twists h_1 in the spatial direction and include a projection operator onto invariant states,

$$P_H = \frac{1}{\text{order}(H)} \sum_{h_2 \in H} \hat{h}_2 \,. \tag{8.5.15}$$

The operator $\hat{h}_2$ in the trace makes the fields in the time direction aperiodic by h_2, so in the end we are summing over twists in both the spatial and time directions on the torus,

$$Z = \frac{1}{\text{order}(H)} \sum_{h_1,h_2 \in H} Z_{h_1,h_2} \,. \tag{8.5.16}$$

The sum over boundary conditions is modular-invariant after a shift of summation variables. The modular transformation $\tau \to -1/\tau$ takes (h_1, h_2) to (h_2, h_1^{-1}), and $\tau \to \tau + 1$ takes (h_1, h_2) to $(h_1, h_1 h_2)$. Thus, if we project onto H-invariant states by summing over h_2, we must also sum over the twisted sectors. In more general situations, in particular right–left asymmetric theories, there will be an issue of phases possibly spoiling the naive modular invariance of the sum over boundary conditions, but we will deal with that in chapters 10, 11, and 16.

In the case of toroidal compactification, the original CFT is the noncompact theory and H consists of $t^m = \exp(2\pi i R m p)$ for all integer m. This happens to have infinite order, but we can regulate the sum by putting the system in a large box. For the orbifold, we can think of starting from the toroidal theory and twisting by r, or from the noncompact theory and twisting by r and t.

When H is non-Abelian, the path integral boundary conditions are inconsistent for h_1 and h_2 noncommuting and the path integral vanishes. This also follows from the canonical expression (8.5.15). If ϕ has spatial twist h_1, then

$$\phi'(\sigma^1) = h_2 \cdot \phi(\sigma^1) \tag{8.5.17}$$

has a different spatial twist

$$\phi'(\sigma^1 + 2\pi) = h_1' \cdot \phi'(\sigma^1) \,, \tag{8.5.18}$$

where

$$h_1' = h_2 h_1 h_2^{-1} \tag{8.5.19}$$

and so the diagonal matrix elements of $\hat{h}_2$ are zero. We also see that the sectors twisted by h_1 and h_1' are not independent — they are related by the projection onto h_2-invariant states. Thus the independent twisted sectors are in one-to-one correspondence with the *conjugacy classes* of H.

Twisting can also be thought of as *gauging* the discrete group H. It may not be obvious that this makes sense, as the discreteness forces the gauge

parameter to be constant. In particular, there can be no associated gauge field. The point is that on a world-sheet with nontrivial closed paths, only gauge-invariant quantities need to be periodic, so the fields can be aperiodic by an H transformation. Gauging a discrete symmetry adds in these path integral sectors and correspondingly projects onto invariant states. This is just the twisting construction. It is also possible to gauge a continuous world-sheet symmetry to produce a new CFT. This is more involved and will be discussed in chapter 15.

The inheritance principle holds for general twisted theories: the path integral with untwisted vertex operators on the sphere is unaffected by the twist.

$c = 1$ CFTs

We have found two families of compact $c = 1$ CFTs. These are toroidal compactification, restricting the radius to $R \geq \alpha'^{1/2}$ so as to count only distinct theories, and orbifold compactification with the same range of R. These two families are connected. Starting from the $SU(2) \times SU(2)$ radius $R = \alpha'^{1/2}$, twist by

$$r' : \quad X^{25} \to X^{25} + \pi\alpha'^{1/2} \,. \tag{8.5.20}$$

This produces the toroidal theory at $R = \alpha'^{1/2}/2$, which by T-duality is the same as at $R = 2\alpha'^{1/2}$. In terms of the $SU(2)$ currents (8.3.12), r' flips the signs of $j^{1,2}$ and $\tilde{\jmath}^{1,2}$ and so is a rotation by π around the 3-axis in each $SU(2)$. The orbifold r flips the signs of $j^{2,3}$ and $\tilde{\jmath}^{2,3}$ and is a rotation by π around the 1-axis in each $SU(2)$. But of course these rotations are equivalent under the $SU(2) \times SU(2)$, so the resulting theories are the same, the $R = 2\alpha'^{1/2}$ torus and the $R = \alpha'^{1/2}$ orbifold. The corresponding partition functions are equal,

$$Z_{\text{orb}}(\alpha'^{1/2}, \tau) = Z_{\text{tor}}(2\alpha'^{1/2}, \tau) \,, \tag{8.5.21}$$

as can be verified by theta function identities.

This equivalence holds only at these radii. For example, at generic R the toroidal theory has a $U(1) \times U(1)$ gauge symmetry while the orbifold has no gauge symmetry. Thus, the two moduli spaces are connected as shown in figure 8.2. This structure, the moduli space of vacua having distinct branches meeting at special points, will be a very common and important feature in *supersymmetric* theories. On each branch there is in general a distinct low energy effective field theory, here differing in the low energy gauge symmetry. In the neighborhood of the special point there are more light fields, different subsets of which become massive as we move away along the various branches.

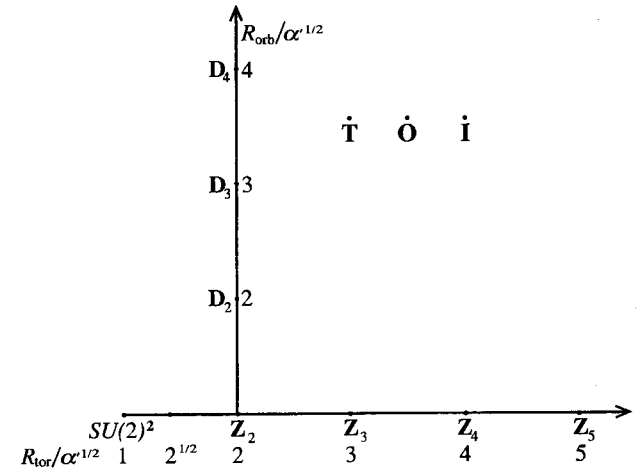


Fig. 8.2. Moduli space of $c = 1$ CFTs. The toroidal branch is horizontal and the orbifold branch is vertical. The special points obtained by twisting by a discrete subgroup of $SU(2)$ are indicated, including the three isolated points obtained from the tetrahedral, octahedral, and dodecahedral groups. Other special radii, such as the torus at $R = (2\alpha')^{1/2}$, will arise later.

It is instructive to analyze the low energy physics near the meeting of the toroidal and orbifold branches. Describe this point in terms of twisting the $SU(2)$ theory by r', eq. (8.5.20). The massless scalars surviving the twist are $j^3\tilde{j}^3$ and $j^i\tilde{j}^j$ for $i, j \in \{1, 2\}$. Since these are all untwisted, the potential is the same as before twisting, namely $\det M$. Again the scalar field M_{ij} can be diagonalized, by a $U(1) \times U(1)$ rotation, and the potential is flat when exactly one diagonal element is nonzero. Now, however, the modulus M_{33} cannot be rotated into M_{11} or M_{22} so these directions are physical distinct. The modulus M_{33} leaves the $U(1) \times U(1)$ unbroken and corresponds to moving along the toroidal branch. The modulus M_{11} or M_{22} breaks the gauge symmetry and corresponds to moving along the orbifold branch. Since this modulus is charged, a reversal of its sign is a gauge transformation and so the line of physically distinct orbifolds ends at the point where the branches meet. We cannot simultaneously vary both moduli, because the linear combination is not a flat direction of the potential. That string theory on an orbifold can be equivalent to string theory on a circle is quite remarkable, one more indication that the string does not see spacetime geometry in the same way that we do.

Moving along the toroidal branch, there are extra massless states at all multiples of the $SU(2)$ radius, $R = k\alpha'^{1/2}$. These are the states with

$(n, w) = (\pm 2k, 0)$. We can think of these points as obtained, in the T-dual description $R = \alpha'^{1/2}/k$, by twisting the $SU(2) \times SU(2)$ theory by the $\mathbf{Z}_k$ symmetry

$$X^{25} \to X^{25} + \frac{2\pi\alpha'^{1/2}}{k} . \tag{8.5.22}$$

This multiplies $j^1 + ij^2$ and $\tilde{j}^1 + i\tilde{j}^2$ by $\exp(2\pi i/k)$ and so the surviving massless scalars are

$$j^3\tilde{j}^3 , \quad j^1\tilde{j}^1 + j^2\tilde{j}^2 , \quad j^1\tilde{j}^2 - j^2\tilde{j}^1 . \tag{8.5.23}$$

The first of these changes the radius of the torus and so is a flat direction as always. However, there are no other branches emerging from these points, because the conditions (8.3.23) on the potential for a flat direction are not met in the directions $M_{11} = M_{22}$ or $M_{12} = -M_{21}$.

The shift (8.5.22) generates a $\mathbf{Z}_k$ subgroup of $SU(2) \times SU(2)$. Let us consider twists by other subgroups. The dihedral group $\mathbf{D}_k$ generated by the shift (8.5.22) and the reflection $X^{25} \to -X^{25}$ produces the orbifold at radius $k\alpha'^{1/2}$. The other three discrete subgroups are the tetrahedral, octahedral, and icosahedral groups. These remove *all* moduli from the CFT, so the twisted CFTs are isolated points, not members of continuous families. Unlike the $\mathbf{Z}_k$ and $\mathbf{D}_k$ twists, which can be defined at all radii, these discrete groups involve elements of $SU(2)$ that exist only at the critical radius and so the radius cannot be varied after twisting. String theory with such a compact space would have only one scalar, the dilaton.

These are all of the known $c = 1$ CFTs, and so give all known 25-dimensional bosonic string backgrounds.

8.6 Open strings

The new feature in toroidal compactification of the open string is the possibility of a nontrivial *Wilson line*, a flat background for the gauge field. Studying first a $U(1)$ gauge theory with some charged fields, consider the constant background

$$A_{25}(x^M) = -\frac{\theta}{2\pi R} = -i\Lambda^{-1}\frac{\partial\Lambda}{\partial x^{25}} , \quad \Lambda(x^{25}) = \exp\left(-\frac{i\theta x^{25}}{2\pi R}\right) , \tag{8.6.1}$$

where θ is a constant. Locally this is pure gauge. The field strength vanishes and the field equations are trivially satisfied. However, the gauge parameter Λ does not satisfy the spacetime periodicity, and so the background has a physical effect. The gauge-invariant quantity that measures this is the Wilson line

$$W_q = \exp\left(iq \oint dx^{25} A_{25}\right) = \exp(-iq\theta) . \tag{8.6.2}$$

Consider first a point particle of charge q, with gauge-fixed world-line action

$$S = \int d\tau \left(\frac{1}{2} \dot{X}^M \dot{X}_M - \frac{m^2}{2} - iqA_M \dot{X}^M \right) . \tag{8.6.3}$$

The gauge action is simply $-iq \int dx^M A_M$, so paths that wind around the compact dimension pick up a phase just equal to the Wilson line W_q. The canonical momentum is

$$p_{25} = -\frac{\partial L}{\partial v^{25}} = v^{25} - \frac{q\theta}{2\pi R} , \tag{8.6.4}$$

where $v^{25} = i\dot{X}^{25}$ as in eq. (8.4.6). The wavefunctions must be periodic in the compact space, so $p_{25} = l/R$ for integer l and

$$v_{25} = \frac{2\pi l + q\theta}{2\pi R} . \tag{8.6.5}$$

The Hamiltonian, which annihilates physical states, is

$$H = \frac{1}{2}(p_\mu p^\mu + v_{25}^2 + m^2) , \tag{8.6.6}$$

so the masses $-p_\mu p^\mu$ are shifted due to the dependence of v_{25} on θ. The same spectrum is readily obtained in a field rather than particle description. Note that v_{25} is just the gauge-invariant momentum $-i\partial_{25} - qA_{25}$.

One can also make the gauge transformation Λ^{-1} to set A_{25} to zero. Charged fields in this gauge are no longer periodic, picking up a phase $\exp(iq\theta)$ under $x^{25} \to x^{25} + 2\pi R$. Physical (gauge-invariant) objects are still periodic. The canonical momentum is now shifted, and we have

$$v_{25} = p_{25} = \frac{2\pi l + q\theta}{2\pi R} \tag{8.6.7}$$

with the same result as before for the gauge-invariant momentum v_{25}.

Let us turn now to the string, and include $U(n)$ Chan–Paton factors. A general constant A_{25} can be diagonalized by a gauge transformation,

$$A_{25} = -\frac{1}{2\pi R} \mathrm{diag}\,(\theta_1, \theta_2, \ldots, \theta_n) . \tag{8.6.8}$$

The gauge field is in the diagonal subgroup of $U(n)$, namely $U(1)^n$. The gauge field couples to the Chan–Paton factor of a general state as $[A_M, \lambda]$, so a string in Chan–Paton state $|ij\rangle$ has charge $+1$ under $U(1)_i$, -1 under $U(1)_j$, and is neutral under the others. Thus it has

$$v_{25} = \frac{2\pi l - \theta_j + \theta_i}{2\pi R} . \tag{8.6.9}$$

The open string spectrum is then

$$m^2 = \frac{(2\pi l - \theta_j + \theta_i)^2}{4\pi^2 R^2} + \frac{1}{\alpha'}(N-1) . \tag{8.6.10}$$

Consider in particular the gauge bosons, where $l = 0$ and $N = 1$ and so

$$m^2 = \frac{(\theta_j - \theta_i)^2}{4\pi^2 R^2} . \tag{8.6.11}$$

In generic backgrounds, all the θs are distinct and the only massless vectors are the diagonal ones, $i = j$. The unbroken gauge group in this case is $U(1)^n$. If r of the θs are equal, the corresponding $r \times r$ matrix of vectors is massless, carrying the gauge symmetry $U(r)$. With the n θs equal in sets of r_i, the gauge symmetry is

$$U(r_1) \times \ldots \times U(r_s) , \quad \sum_{i=1}^{s} r_i = n . \tag{8.6.12}$$

As always, there is a low-energy interpretation. The gauge field A_{25} is a 25-dimensional scalar in the adjoint representation of $U(n)$, and its vacuum expectation value breaks the symmetry spontaneously to $U(r_1) \times \ldots \times U(r_s)$.

The reader can fill in the details of the effective low-energy action. It is important that with k compact dimensions, the potential contains a term

$$\mathrm{Tr}\left([A_m, A_n]^2 \right) \tag{8.6.13}$$

from the field strength in the 26-dimensional Yang–Mills action. This forces the gauge fields in different directions to commute in a static solution, so that they are simultaneously diagonalizable. There are then kn moduli from the gauge field and k^2 from the metric and antisymmetric tensor.

T-duality

Now consider the $R \to 0$ limit of the open string spectrum. Open strings with Neumann boundary conditions have no quantum number comparable to w; they can always be unwound from the periodic dimension. So when $R \to 0$ the states with nonzero momentum go to infinite mass, but there is no new continuum of states. The behavior is as in field theory: the resulting states move in 25 spacetime dimensions.

A seeming paradox arises when one remembers that theories with open strings always have closed strings as well, so that in the $R \to 0$ limit the closed strings move in 26 spacetime dimensions but the open strings only in 25. One can reason out what is happening as follows. The interior of the open string is made of the same 'stuff' as the closed string, and so should still be vibrating in 26 dimensions. What distinguishes the open string is its endpoints, so it is these that must be restricted to a 25-dimensional hyperplane.

Indeed, recall that one describes the T-dual theory by using the new

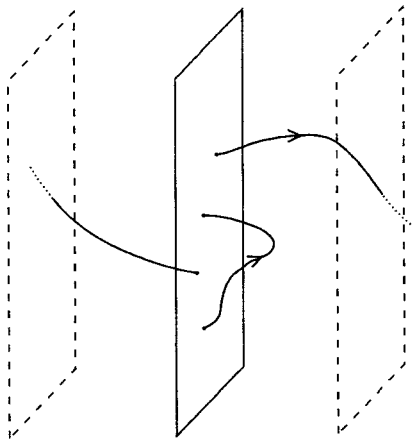


Fig. 8.3. Open strings with endpoints attached to a hyperplane. The dashed planes are periodically identified. Two strings are shown, with winding numbers zero and one.

embedding coordinate

$$X'^{25}(z,\bar{z}) = X_L^{25}(z) - X_R^{25}(\bar{z}) \,. \tag{8.6.14}$$

Then

$$\partial_n X^{25} = -i\partial_t X'^{25}, \tag{8.6.15}$$

with n the normal and t the tangent at the boundary. The Neumann condition on the original coordinate is a Dirichlet condition on the dual coordinate: the X^{25} coordinate of each string endpoint is fixed.

Let us first consider compactification without Wilson lines. Then all endpoints are actually constrained to lie on *the same* hyperplane. To see this, integrate

$$\begin{aligned} X'^{25}(\pi) - X'^{25}(0) &= \int_0^\pi d\sigma^1 \, \partial_1 X'^{25} = -i \int_0^\pi d\sigma^1 \, \partial_2 X^{25} \\ &= -2\pi\alpha' v^{25} = -\frac{2\pi\alpha' l}{R} = -2\pi l R' \,. \end{aligned} \tag{8.6.16}$$

The total change in X'^{25} between the two ends is an integral multiple of the periodicity $2\pi R'$ of the dual dimension, so the ends lie on the same hyperplane in the periodic T-dual space. For two *different* open strings, consider the connected world-sheet that results from graviton exchange between them. One can carry out the same argument (8.6.16) on a path connecting any two endpoints, so all endpoints of all strings lie on the

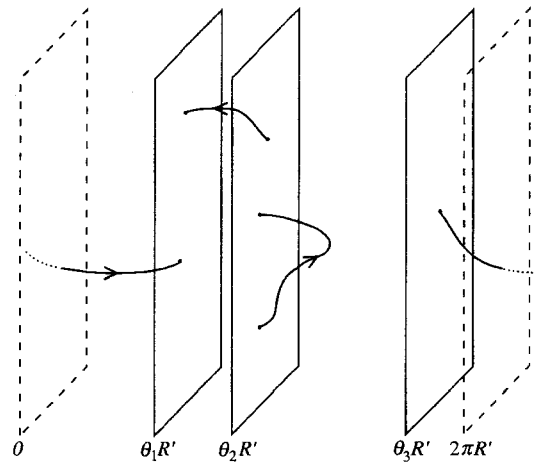


Fig. 8.4. $n = 3$ hyperplanes at different positions, with various strings attached. Ignoring the tachyon, the lightest strings with both ends on the same hyperplane are massless, while strings stretched between noncoincident hyperplanes are massive due to their tension. When two hyperplanes become coincident, the lightest strings stretched between them become massless.

same hyperplane, as in figure 8.3. The ends are still free to move in the other 24 spatial dimensions.

Now we consider the effect of Wilson lines. Due to the shift (8.6.9) in v_{25}, the difference (8.6.16) is replaced by

$$\Delta X'^{25} = X'^{25}(\pi) - X'^{25}(0) = -(2\pi l - \theta_j + \theta_i)R' \,. \tag{8.6.17}$$

In other words, up to an arbitrary additive normalization, the endpoint in state i is at

$$X'^{25} = \theta_i R' = -2\pi\alpha' A_{25,ii}. \tag{8.6.18}$$

Thus there are in general n hyperplanes at different positions as depicted in figure 8.4.

The mode expansion in the sector with Chan–Paton state $|ij\rangle$, for an open string winding l times around the compact dimension, is

$$\begin{aligned} X'^{25}(z,\bar{z}) &= \theta_i R' - \frac{iR'}{2\pi}(2\pi l - \theta_j + \theta_i)\ln(z/\bar{z}) + i\left(\frac{\alpha'}{2}\right)^{1/2} \sum_{\substack{m=-\infty \\ m\neq 0}}^{\infty} \frac{\alpha_m^{25}}{m}(z^{-m} - \bar{z}^{-m}) \\ &= \theta_i R' + \frac{\sigma^1}{\pi}\Delta X'^{25} + (2\alpha')^{1/2} \sum_{\substack{m=-\infty \\ m\neq 0}}^{\infty} \frac{\alpha_m^{25}}{m} \exp(-m\sigma^2)\sin m\sigma^1. \end{aligned} \tag{8.6.19}$$

The spectrum (8.6.10) becomes

$$m^2 = \left(\frac{\Delta X'^{25}}{2\pi\alpha'}\right)^2 + \frac{1}{\alpha'}(N-1)\,, \tag{8.6.20}$$

where $\Delta X'^{25}$, given by eq. (8.6.17), is the minimum length for a string in the given sector.

Now consider how this picture generalizes if several coordinates are periodic, say X^m for $p+1 \le m \le 25$, and we rewrite the periodic dimensions in terms of the dual coordinate. The Neumann conditions on X^m again become Dirichlet conditions for the dual coordinates X'^m, so the open string endpoints are confined to n parallel $(p+1)$-dimensional subspaces. The Wilson lines in the various directions become independent coordinates for each subspace.

The dual picture is rather odd in that translation invariance is broken by the open string boundary conditions. This simply reflects the fact that in the original open string theory winding number is not conserved (a closed winding string can break open and unwind) and winding number is T-dual to momentum. It is important to remember that T-duality is just a different description of the same theory, and further that when the compactification radii in the original theory are small the T-dual picture is the natural one to use. Figure 8.4 then shows how to interpret the resulting physics.

8.7 D-branes

We have just learned that when an open string theory is compactified on a small torus, the physics is described by compactification on a large torus but with the open string endpoints restricted to lie on subspaces. In fact, these subspaces are new dynamical objects in their own right.

Consider the spectrum of massless open strings for the generic configuration in which all θ_i are distinct, taking for simplicity the case in which only one coordinate is dualized. Ignoring the tachyon, the mass (8.6.20) vanishes only for $N=1$ and $\Delta X'^{25}=0$. That is, both endpoints are on the same hyperplane, with zero winding. We have therefore the massless states

$$\alpha^\mu_{-1}|k;ii\rangle\,, \quad \mathcal{V} = i\partial_t X^\mu\,, \tag{8.7.1a}$$

$$\alpha^{25}_{-1}|k;ii\rangle\,, \quad \mathcal{V} = i\partial_t X^{25} = \partial_n X'^{25}\,. \tag{8.7.1b}$$

These are of course the same massless states as in the original theory, T-duality simply giving a different picture of the same spectrum. For the generic Wilson line, the only massless open string states in the original theory would have been the n massless $U(1)$ vectors. In eq. (8.7.1) we have separated these according to whether the polarization is tangent or perpendicular to the hyperplane. The 25 states with tangent polarization comprise a gauge field living on the hyperplane. The state with perpendicular polarization was the gauge field in the compact direction of the original theory, and it has a simple and important interpretation in the T-dual theory: it is a collective coordinate for the shape of the hyperplane. This is already evident from the fact that a constant gauge background (8.6.18) corresponds to a uniform translation of the hyperplane. An x^μ-dependent background would by the same token go over to an x^μ-dependent translation, a curved hyperplane, and the quanta of the field A^{25}_{ii} correspond to small oscillations of the hyperplane.

This is the same phenomenon as with spacetime itself. We start with strings in a flat background and discover that a massless closed string state corresponds to fluctuations of the geometry. Here we found first a flat hyperplane, and then discovered that a certain open string state corresponds to fluctuations of its shape. We should not be surprised that the hyperplane has become dynamical. String theory contains gravity. A gravitational wave passing through the hyperplane would warp spacetime itself, so the hyperplane could hardly remain rigid.

Thus the hyperplane is indeed a dynamical object, a Dirichlet membrane, or *D-brane* for short. The p-dimensional D-brane, from dualizing $25-p$ dimensions, is a Dp-brane. In this terminology, the original $U(n)$ open string theory contains n D25-branes. A D25-brane fills space, so the string endpoint can be anywhere: it just corresponds to an ordinary Chan–Paton factor.

Since T-duality interchanges Neumann and Dirichlet boundary conditions, a further T-duality in a direction tangent to a Dp-brane reduces it to a D$(p-1)$-brane, while a T-duality in an orthogonal direction turns it into a D$(p+1)$-brane. The case of a nontrivial angle will arise shortly.

We could have started by investigating Dirichlet boundary conditions in their own right. The route we have taken, discovering them via T-duality, is a good way to develop many of their properties. It also shows that one can reach a state containing D-branes from an 'ordinary' state by a continuous process. That is, take the $R \to 0$ limit of the original theory, leaving n D-branes in a noncompact space. In the superstring, we will use this to argue that D-branes are an essential part of the nonperturbative definition of the theory. For the bosonic string we will not try to make this argument because there is no strong argument that the theory has a nonperturbative definition. It is possible that bosonic string theory exists only as a mutilation of superstring theory.

It is interesting to look at the $U(n)$ symmetry breaking in the T-dual picture. When no D-branes coincide there is just one massless vector on each, giving gauge group $U(1)^n$ in all. If r D-branes coincide, there are new

massless states because strings that are stretched between these branes can have vanishing length: the $\Delta X'^{25}$ in the mass formula (8.6.20) vanishes for $n = 0$ if both i and j are in this set of r. Thus, there are r^2 vectors, forming the adjoint of a $U(r)$ gauge group. This is just as found in the dual Wilson line picture. It is striking, however, that there will also be r^2 massless scalars from the components normal to the D-brane. We will discuss the significance of this below.

The D-brane action

The massless fields on the world-volume of a Dp-brane are a $U(1)$ vector plus $25-p$ world-brane scalars describing the fluctuations. The low energy effective action for this system is as always a very useful thing to consider. We now take the dual radius R' to infinity, focusing on one D-brane in the resulting 26-dimensional spacetime. The world-brane fields are in interaction with the massless closed string fields, whose action has already been discussed in section 3.7. Introduce coordinates ξ^a, $a = 0,\dots p$ on the brane. The fields on the brane are the embedding $X^\mu(\xi)$ and the gauge field $A_a(\xi)$. We claim that the action is

$$S_p = -T_p \int d^{p+1}\xi\, e^{-\Phi}[-\det(G_{ab} + B_{ab} + 2\pi\alpha' F_{ab})]^{1/2} , \qquad (8.7.2)$$

with T_p a constant to be determined later. Here

$$G_{ab}(\xi) = \frac{\partial X^\mu}{\partial \xi^a}\frac{\partial X^\nu}{\partial \xi^b} G_{\mu\nu}(X(\xi)) , \quad B_{ab}(\xi) = \frac{\partial X^\mu}{\partial \xi^a}\frac{\partial X^\nu}{\partial \xi^b} B_{\mu\nu}(X(\xi)) \qquad (8.7.3)$$

are the induced metric and antisymmetric tensor on the brane.

All features of the action (8.7.2) can be understood from general reasoning. Considering first just the spacetime metric and the embedding, the simplest and lowest-derivative coordinate-invariant action is the integral of $(-\det G_{ab})^{1/2}$, the world-volume. Note that this term has an implicit dependence on $X^\mu(\xi)$ through the induced fields (8.7.3). Expanded around a flat D-brane it gives the action for the fluctuations, just as the Nambu–Goto action describes the fluctuations of a string.

The dilaton dependence $e^{-\Phi} \propto g_c^{-1}$ arises because this is an open string tree-level action. The self-interactions of the open string fields, and their couplings to the closed string fields, arise first from the disk.

The dependence on F_{ab} can be understood by using T-duality. Consider a D-brane that is extended in the X^1- and X^2-directions, with the other dimensions unspecified, and let there be a constant gauge field F_{12} on it. Go to the gauge $A_2 = X^1 F_{12}$. Now take the T-dual along the 2-direction. The Neumann condition in this direction is replaced by a Dirichlet condition, so the D-brane loses a dimension. However, the T-duality relation (8.6.18)

between the potential and coordinate implies that the D-brane is tilted in the (1–2)-plane,

$$X'^2 = -2\pi\alpha' X^1 F_{12} . \qquad (8.7.4)$$

The tilt gives a geometric factor in the action,

$$\int dX^1 \left[1 + (\partial_1 X'^2)^2\right]^{1/2} = \int dX^1 \left[1 + (2\pi\alpha' F_{12})^2\right]^{1/2} . \qquad (8.7.5)$$

For any D-brane, by boosting to align it with the coordinate axes and then rotating to bring F_{ab} to block-diagonal form, one can reduce the action to a product of factors (8.7.5) in each plane, equivalent to the F_{ab} term in the determinant (8.7.2). This determinant form for the gauge field action is known as the Born–Infeld action. It was originally proposed, unsuccessfully, as a solution to the short-distance divergences of quantum electrodynamics.

Finally, the dependence on B_{ab} is given by the following argument. The closed string field $B_{\mu\nu}$ and the open string field A^μ appear in the *string* world-sheet action as[3]

$$\frac{i}{4\pi\alpha'}\int_M d^2\sigma\, g^{1/2}\epsilon^{ab}\partial_a X^\mu \partial_b X^\nu B_{\mu\nu} + i\int_{\partial M} dX^\mu A_\mu . \qquad (8.7.6)$$

Associated with each of these fields is a spacetime gauge invariance, which must be preserved for the consistency of the spacetime theory. The ordinary gauge transformation

$$\delta A_\mu = \partial_\mu \lambda \qquad (8.7.7)$$

is an invariance of the action (8.7.6), the boundary term changing by the integral of a total derivative. The antisymmetric tensor variation

$$\delta B_{\mu\nu} = \partial_\mu \zeta_\nu - \partial_\nu \zeta_\mu \qquad (8.7.8)$$

similarly changes the bulk action by a total derivative, but on a bounded world-sheet this gives rise to a surface term. This can only be canceled if the open string field A_μ also transforms under the tensor gauge symmetry,

$$\delta A_\mu = -\zeta_\mu/2\pi\alpha' . \qquad (8.7.9)$$

Now, however, only the combination

$$B_{\mu\nu} + 2\pi\alpha' F_{\mu\nu} \equiv 2\pi\alpha' \mathscr{F}_{\mu\nu} \qquad (8.7.10)$$

[3] For the reader familiar with differential forms, this is

$$\frac{i}{2\pi\alpha'}\int_M B + i\int_{\partial M} A .$$

The reader can translate the subsequent equations similarly.

is invariant under both symmetries, and it is this combination that must appear in the action. Thus the form (8.7.2) of the action is fully determined.

As a check, it is natural that the combination $G_{\mu\nu} + B_{\mu\nu}$ should appear, since this same combination appears in the world-sheet action in conformal gauge. Incidentally, the action (8.7.2) can also be determined by taking the low energy limit of various open and open plus closed string amplitudes, but this is rather more laborious.

For n separated D-branes, the action is n copies of the action for a single D-brane. We have seen, however, that when the D-branes are coincident there are n^2 rather than n massless vectors and scalars on the brane, and we would like to write down the effective action governing these. The fields $X^\mu(\xi)$ and $A_a(\xi)$ will now be $n \times n$ matrices. For the gauge field the meaning is obvious — it becomes a non-Abelian $U(n)$ gauge field. For the collective coordinate X^μ, however, the meaning is mysterious: the collective coordinates for the embedding of n D-branes in spacetime are now enlarged to $n \times n$ matrices. This 'noncommutative geometry' has proven to play an key role in the dynamics of D-branes, and there are conjectures that it is an important hint about the nature of spacetime.

We can gain more insight by considering the effective action. There is now a nonderivative term in the action, a potential for the collective coordinates, which can be deduced by T-duality from a constant A_m field. For such a vector potential, the field strength is just $[A_m, A_n]$, which becomes $(2\pi\alpha')^{-2}[X_m, X_n]$ in the T-dual picture. Expanding in the field strength, the leading term in the action is roughly

$$V \propto \mathrm{Tr}\,([X_m, X_n][X^m, X^n]) \ . \tag{8.7.11}$$

This potential has the property that at the point $X^m = 0$ its second derivative vanishes, so that all kn^2 scalars are massless; again $k = 25-p$ is the number of dualized dimensions. However, the space of flat dimensions is smaller. The potential is the sum of squares and so vanishes only if all of the $[X^m, X^n]$ vanish. We can then use the $U(n)$ gauge symmetry to go to a gauge in which the X^m are all diagonal. There are thus n flat directions, just the number of diagonal elements. These n diagonal elements can then be interpreted as the collective coordinates of the n Dp-branes. Thus the potential (8.7.11) correctly interpolates the physics of separated and coincident D-branes.

When the X^m do commute, the action should reduce to that of n separate D-branes, so we must have

$$S_p = -T_p \int d^{p+1}\xi\, \mathrm{Tr}\Big\{e^{-\Phi}[-\det(G_{ab} + B_{ab} + 2\pi\alpha' F_{ab})]^{1/2} + O([X^m, X^n]^2)\Big\} \ . \tag{8.7.12}$$

The determinant is on the world-volume indices ab, while the trace is over

the n Chan–Paton indices. The trace is the appropriate $U(n)$ invariant which for diagonal matrices reduces to a sum over the separate D-branes. The full dependence on the commutator is more complicated than the simple potential (8.7.11), with additional couplings to other fields and corrections at higher orders in the commutator. However, the key property, the form of the flat directions, is unaffected. The full dependence can be determined by starting from the Born–Infeld action for the fully Neumann case and performing the T-duality. Incidentally, higher derivative terms involving commutators of field strengths cannot be determined by T-duality alone.

The D-brane tension

It is instructive to compute the constant T_p, and for the superstring the precise value will be significant. Before we carry out the explicit calculation, it is interesting to note that one gets a recursion relation for T_p from T-duality. Note that in a constant dilaton background, the tension of a Dp-brane is $T_p e^{-\Phi}$: this is minus the action per unit volume for a static Dp-brane. Consider now a Dp-brane in which the p directions tangent to the D-brane have been periodically identified. That is, the Dp-brane is *wrapped* on a p-torus in spacetime. Its mass is its tension times the volume of the torus,

$$T_p e^{-\Phi} \prod_{i=1}^{p} (2\pi R_i) \ . \tag{8.7.13}$$

Now take the T-dual on one of the periodic dimensions X^p. This does not change the mass, being just a new description of the same state. In terms of the dilaton (8.3.31) of the T-dual theory, the mass (8.7.13) is

$$2\pi\alpha'^{1/2} T_p e^{-\Phi'} \prod_{i=1}^{p-1} (2\pi R_i) \ . \tag{8.7.14}$$

In the dual theory, however, it is a D$(p-1)$-brane wrapped on a $(p-1)$-torus, so its mass is

$$T_{p-1} e^{-\Phi'} \prod_{i=1}^{p-1} (2\pi R_i) \ . \tag{8.7.15}$$

Equating the masses (8.7.14) and (8.7.15) gives

$$T_p = T_{p-1}/2\pi\alpha'^{1/2} \ , \tag{8.7.16}$$

which determines the T_p up to one overall normalization.

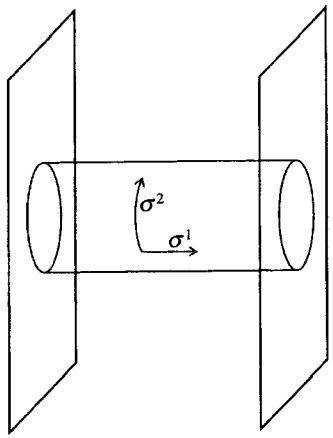


Fig. 8.5. Exchange of a closed string between two D-branes. Equivalently, a vacuum loop of an open string with one end on each D-brane.

To determine the actual value of the D-brane tension we need to calculate a string amplitude. For example, we could deduce it from the gravitational coupling to the D-brane, given by the disk with a graviton vertex operator. However, this involves the yet-unknown ratio g_c/g_o^2, the closed string coupling arising from the vertex operator and the open string coupling from the disk. Instead we can obtain the absolute normalization as follows. Consider two parallel Dp-branes at positions $X_1^m = 0$ and $X_2^m = y^m$. These two objects can feel each other's presence by exchanging closed strings as shown in figure 8.5. This string amplitude is an annulus, with no vertex operators, and so can be calculated by the same method as in the previous chapter. The poles from graviton and dilaton exchange then give the coupling T_p of closed string states to the D-brane.

Recall that in section 7.4 we calculated the annulus vacuum amplitude in terms of a loop of open strings, but then found closed string poles from the $t \to 0$ limit of the modular integration. Here it is just those poles that we are interested in, but once again the simplest way to calculate the amplitude is to think of it as an open string loop. In fact, the earlier result (7.4.1) carries over with just a few changes. The number of momentum integrations is reduced from 26 to $p+1$ and similarly V_{26} becomes V_{p+1}; the weight h_i acquires an additional term $y^2/4\pi^2\alpha'$ from the tension of the stretched open string; the Chan–Paton weight n^2 becomes 2 (since the string can attach with either orientation). Thus we

have

$$\begin{aligned}\mathcal{A} &= iV_{p+1}\int_0^\infty \frac{dt}{t}(8\pi^2\alpha' t)^{-(p+1)/2}\exp(-ty^2/2\pi\alpha')\eta(it)^{-24} \\ &= \frac{iV_{p+1}}{(8\pi^2\alpha')^{(p+1)/2}}\int_0^\infty dt\, t^{(21-p)/2}\exp(-ty^2/2\pi\alpha') \\ &\qquad \times[\exp(2\pi/t)+24+\ldots]\,, \end{aligned} \tag{8.7.17}$$

where the asymptotics have been obtained just as in section 7.4. The first term is from tachyon exchange and so is again an uninteresting bosonic artifact. The integral over the second term gives

$$\begin{aligned}\mathcal{A} &= iV_{p+1}\frac{24}{2^{12}}(4\pi^2\alpha')^{11-p}\pi^{(p-23)/2}\Gamma\left(\frac{23-p}{2}\right)|y|^{p-23} \\ &= iV_{p+1}\frac{24\pi}{2^{10}}(4\pi^2\alpha')^{11-p}G_{25-p}(y)\,, \end{aligned} \tag{8.7.18}$$

where $G_d(y)$ is the massless scalar Green's function in d dimensions, the inverse of $-\nabla^2$.

We now need to compare this with a field theory calculation of the same amplitude. Since the antisymmetric tensor does not couple to the D-branes, the relevant terms from the spacetime action (3.7.25) are

$$S = \frac{1}{2\kappa^2}\int d^{26}X\,(-\tilde{G})^{1/2}\left(\tilde{R} - \frac{1}{6}\nabla_\mu\tilde{\Phi}\tilde{\nabla}^\mu\tilde{\Phi}\right)\,. \tag{8.7.19}$$

Recall that the tilde denotes the Einstein metric; this is the convenient form to use because its action decouples from that of the dilaton. The tilded dilaton has been shifted so that its expectation value is zero. In terms of the same variables, the relevant terms from the D-brane action (8.7.2) are

$$S_p = -\tau_p\int d^{p+1}\xi\,\exp\left(\frac{p-11}{12}\tilde{\Phi}\right)(-\det\tilde{G}_{ab})^{1/2}\,. \tag{8.7.20}$$

We have defined $\tau_p = T_p e^{-\Phi_0}$; this is the actual physical tension of the Dp-brane when the background value of the dilaton is Φ_0.

The field theory graph analogous to figure 8.5 is the exchange of a single graviton or dilaton between the D-branes. To obtain the propagator we expand the spacetime action to second order in $h_{\mu\nu} = G_{\mu\nu} - \eta_{\mu\nu}$ and to second order in $\tilde{\Phi}$. Also, we need to choose a gauge for the gravitational calculation. The simplest gauge for perturbative calculation is

$$F_\nu \equiv \partial^{\hat{\mu}}h_{\mu\nu} - \frac{1}{2}\partial_\nu h^{\hat{\mu}}{}_{\mu} = 0\,, \tag{8.7.21}$$

where a hat indicates that an index has been raised with the flat space metric $\eta^{\mu\nu}$. Expanding the action to second order and adding a gauge-

fixing term $-F_\nu F^{\hat\nu}/4\kappa^2$, the spacetime action becomes

$$S = -\frac{1}{8\kappa^2}\int d^{26}X\left(\partial_\mu h_{\nu\lambda}\partial^{\hat\mu}h^{\hat\nu\hat\lambda} - \frac{1}{2}\partial_\mu h^{\hat\nu}{}_{\hat\nu}\partial^{\hat\mu}h^{\hat\lambda}{}_{\hat\lambda} + \frac{2}{3}\partial_\mu\tilde\Phi\partial^{\hat\mu}\tilde\Phi\right) . \quad (8.7.22)$$

Inverting the kinetic term, one obtains the momentum space propagators

$$\langle\tilde\Phi\tilde\Phi\rangle = -\frac{(D-2)i\kappa^2}{4k^2} , \quad (8.7.23a)$$

$$\langle h_{\mu\nu}h_{\sigma\rho}\rangle = -\frac{2i\kappa^2}{k^2}\left(\eta_{\mu\sigma}\eta_{\nu\rho} + \eta_{\mu\rho}\eta_{\nu\sigma} - \frac{2}{D-2}\eta_{\mu\nu}\eta_{\sigma\rho}\right) . \quad (8.7.23b)$$

For later reference we have given these for general D. The D-brane action expanded around a flat configuration is

$$S_p = -\tau_p\int d^{p+1}\xi\left(\frac{p-11}{12}\tilde\Phi - \frac{1}{2}h_{aa}\right) . \quad (8.7.24)$$

Note that $h_{\mu\nu}$ here is traced only over the directions tangent to the D-brane; we have taken ξ to be rectilinear coordinates with metric δ_{ab}.

We can now read off the Feynman graph,

$$\begin{aligned}\mathscr{A} &= \frac{i\kappa^2\tau_p^2}{k^2}\left\{6\left[\frac{p-11}{12}\right]^2 + \frac{1}{2}\left[2(p+1) - \frac{1}{12}(p+1)^2\right]\right\} \\ &= \frac{6i\kappa^2\tau_p^2}{k^2} . \end{aligned} \quad (8.7.25)$$

Comparison with the string result (8.7.18) yields

$$\tau_p^2 = \frac{\pi}{256\kappa^2}(4\pi^2\alpha')^{11-p} . \quad (8.7.26)$$

It is a useful check that, once the dust has cleared, the recursion relation from T-duality is satisfied.

As one application, consider a state with n 25-branes, which is the same as an ordinary (fully Neumann) n-valued Chan–Paton factor. Expanding the 25-brane action (8.7.12) to second order in the gauge field gives

$$\frac{\tau_{25}}{4}(2\pi\alpha')^2\mathrm{Tr}(F_{\mu\nu}F^{\mu\nu}) . \quad (8.7.27)$$

This relates the open string gauge coupling to the gravitational coupling. Using eqs. (6.5.14), (6.5.16), (6.6.15), and (6.6.18), we can write it as a relation between the vertex operator normalizations g_c and g_o,

$$\frac{g_o^2}{g_c} = \frac{4\pi\alpha' g_o'^2}{\kappa} = 2^{18}\pi^{25/2}\alpha'^6 . \quad (8.7.28)$$

This has the right form, the square of the open string coupling being proportional to the closed string coupling, but now with the numerical coefficient included. We have essentially obtained this via unitarity, factoring on the closed string poles of the annulus diagram. We could also have done this more directly, as in exercise 7.9, without the excursion through D-branes.

8.8 T-duality of unoriented theories

The $R \to 0$ limit in an unoriented string theory also leads to interesting new physics. Consider closed strings first. To form the unoriented theory we impose $\Omega = +1$ on the states. Following the discussion in section 8.5, we can think about this as *gauging* Ω. In particular, the transition functions used to build the world-sheet may now include orientation reversal, since it is a gauge symmetry, and this produces unoriented world-sheets.

The T-dual theory is obtained by using the coordinate $X'^m(z,\bar z) = X_L^m(z) - X_R^m(\bar z)$ in place of $X^m(z,\bar z) = X_L^m(z) + X_R^m(\bar z)$. In the original description we are gauging world-sheet parity Ω, which acts as

$$\Omega : \quad X_L^M(z) \leftrightarrow X_R^M(\bar z) . \quad (8.8.1)$$

In terms of the T-dual coordinates, this is

$$\Omega : \quad X'^m(z,\bar z) \leftrightarrow -X'^m(\bar z, z) , \quad (8.8.2a)$$

$$X'^\mu(z,\bar z) \leftrightarrow X'^\mu(\bar z, z) , \quad (8.8.2b)$$

where again m indexes the coordinates on which the T-dual has been taken and μ those on which it has not. In the dual picture, the symmetry (8.8.2) is thus the *product* of a world-sheet parity transformation and a spacetime reflection. We have seen that gauging world-sheet parity alone produces the unoriented theory, while gauging the reflection alone produces an orbifold. The result of the combined projection is known as an *orientifold.*

Orientifolds are rather similar to orbifolds. Separate the string wavefunction into its internal part and its dependence on the center of mass x^m, and take the internal wavefunction to be an eigenstate of Ω. The projection onto $\Omega = +1$ then determines the string wavefunction at $-x^m$ to be the same as that at x^m, up to a sign. For example, the various components of the metric and antisymmetric tensor satisfy

$$G_{\mu\nu}(x') = G_{\mu\nu}(x) , \quad B_{\mu\nu}(x') = -B_{\mu\nu}(x) , \quad (8.8.3a)$$

$$G_{\mu n}(x') = -G_{\mu n}(x) , \quad B_{\mu n}(x') = B_{\mu n}(x) , \quad (8.8.3b)$$

$$G_{mn}(x') = G_{mn}(x) , \quad B_{mn}(x') = -B_{mn}(x) , \quad (8.8.3c)$$

where $(x^\mu, x^m)' = (x^\mu, -x^m)$. There is a minus sign for each m,n and an additional minus sign for each B_{MN}; the orbifold is the same except that the latter sign would be absent. In other words, the T-dual spacetime is the torus T^k identified under a $\mathbf{Z}_2$ reflection in the k compact directions, just as in the orbifold construction. In the case of a single periodic dimension,

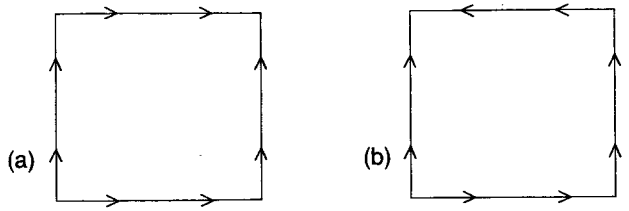


Fig. 8.6. Identifications defining (a) the torus and (b) the Klein bottle.

for example, the dual spacetime is the line segment $0 \leq x^{25} \leq \pi R'$, with orientifold fixed planes at the ends.

Notice also that away from the orientifold fixed planes, the local physics is that of the *oriented* string theory. Unlike the original unoriented theory, where the projection removes half the states locally, here it relates the string wavefunction at any point to its value at the image point, as in eq. (8.8.3).

One difference between the orbifold and orientifold constructions is that in the latter there is no direct analog of the twisted states, because the Klein bottle does not have the modular transformation $\tau \to -1/\tau$. Consider figure 8.6, which shows the torus and Klein bottle. On the torus, the projection operator inserts a twist in the timelike direction of figure 8.6(a); rotating the figure by 90°, this becomes a twist in the spacelike direction, implying twisted states in the spectrum. However, if the Klein bottle of figure 8.6(b) is rotated by 90°, the directions of time on the two opposite edges do not match and there is no interpretation in terms of intermediate states in this channel. It should be noted then that orientifold planes cannot be dynamical. Unlike the case of D-branes, there are no string modes tied to the orientifold plane to represent fluctuations in its shape. Our heuristic argument that a gravitational wave forces a D-brane to oscillate does not apply to the orientifold plane. Essentially, the identifications (8.8.3) become boundary conditions at the fixed plane, such that the incident and reflected waves cancel. For the D-brane, the reflected wave is higher order in the string coupling.

We draw oriented strings with arrows to indicate the orientation. In the unoriented theory, we can either omit the arrow or take a linear combination of the two orientations. The latter picture is more consistent with the idea of gauging the discrete symmetry, and is more general: in an orientifold the parity operation is accompanied by a spacetime transformation, so we cannot just forget the arrow.

Although we have introduced the orientifold construction via T-duality, one can consider more general orientifolds that are not simply T-duals

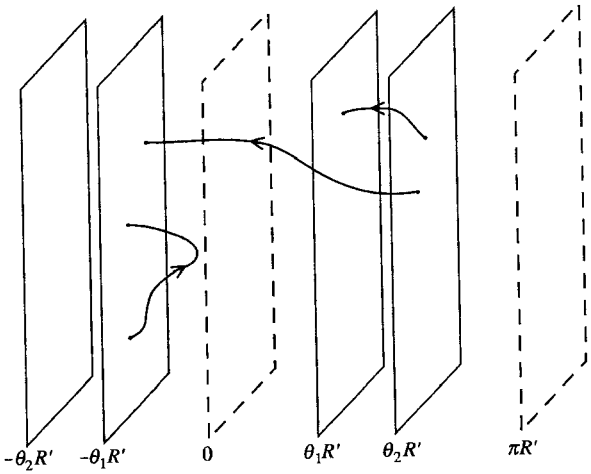


Fig. 8.7. Orientifold planes at 0 and $\pi R'$, D-branes at $\theta_1 R'$ and $\theta_2 R'$, and D-brane images at $-\theta_1 R'$ and $-\theta_2 R'$. The twist operator Ω acts on any string by a combination of a spacetime reflection and reversing the orientation arrow, so that the string running from 2 to the image of 1 becomes a string running from 1 to the image of 2.

of toroidal compactifications, gauging a group that combines world-sheet parity with various spacetime symmetries.

Open strings

In the case of open strings, the situation is similar. Let us focus for convenience on a single compact dimension. Again there is one orientifold fixed plane at 0 and another at $\pi R'$. Introducing $SO(n)$ Chan–Paton factors (the symplectic case works in the same way), a general Wilson line can be brought to diagonal form, and for n even the eigenvalues are paired:

$$W = \mathrm{diag}\left(e^{i\theta_1}, e^{-i\theta_1}, e^{i\theta_2}, e^{-i\theta_2}, \cdots, e^{i\theta_{n/2}}, e^{-i\theta_{n/2}}\right) . \tag{8.8.4}$$

Thus in the dual picture there are $\frac{1}{2}n$ D-branes on the line segment $0 \leq X'^{25} \leq \pi R'$, and $\frac{1}{2}n$ at their image points under the orientifold identification. Strings can stretch between D-branes and their images as shown in figure 8.7. The generic gauge group is $U(1)^{n/2}$. As in the oriented case, if r D-branes are coincident there is a $U(r)$ gauge group. However, if now the r D-branes in addition lie at one of the fixed planes, then strings stretching between one of these branes and one of the image branes also become massless and we have the right spectrum of additional states

to fill out $SO(2r)$. The maximal $SO(n)$ is restored if all of the branes are coincident at a single orientifold plane. Note that this maximally symmetric case is asymmetric between the two fixed planes.

If n is odd, the final eigenvalue of W is ± 1, so that in the T-dual picture it is fixed to lie on one or the other orientifold plane. Having no image, it is actually a half D-brane, as measured for example by its mass. Also, if we consider for $n = 2$ the Wilson line diag$(1, -1)$ (which is in $O(2)$ but not $SO(2)$), then rather than one D-brane and its image there are two half D-branes, one at each fixed plane.

The orientifold plane, like the D-brane, couples to the dilaton and metric. The amplitude is just as in the previous section, with the Klein bottle and Möbius strip replacing the annulus. In fact, we have already done the relevant calculation in chapter 7. There we found that the total dilaton coupling cancels for $SO(2^{13})$. In the T-dual picture, the total dilaton coupling of the orientifold planes cancels that of 2^{12} D-branes (the images don't count!). If we T-dualize on $k = 25 - p$ dimensions there are 2^k fixed planes, where the coordinates are all combinations of 0 and $\pi R'_m$. Thus the effective action for a single fixed plane is

$$2^{12-k} T_p \int d^{p+1}\xi \, e^{-\Phi} (-\det G_{ab})^{1/2} , \tag{8.8.5}$$

the integral running over the fixed plane.

Although the orientifold has no direct analog of twisted sectors, it has been argued that there is a sense in which the twisted states are *open strings*. We prefer not to emphasize this, as the analogy is not complete: gauging Ω does not of itself introduce world-sheet boundaries. However, there is a definite physical sense in which it is natural to add open strings to an orientifold theory, to cancel the dilaton tadpole. This cancellation will play an important role in the superstring.

Exercises

8.1 (a) Derive the equal time commutator (8.2.21).
(b) Work out the details showing that vertex operators of the form (8.2.22) commute at equal time and finite separation.

8.2 For a single periodic dimension, calculate the scattering amplitude for four closed string tachyons with winding numbers $+1$, -1, 0, and 0, and vanishing compact momenta.

8.3 We have shown T-duality for spacetime-independent backgrounds. We would expect that a background with a varying metric $G_{25,25}(x^\mu)$ would be T-dual to one with the inverse metric $G_{25,25}(x^\mu)^{-1}$. In principle we can show this by expanding around the static configuration, but it is useful to show it directly as follows.

(a) Write the world-sheet action for a string moving in a general massless background, with only the requirement that the background be invariant under translations of X^{25}. Let the coordinate X^{25} have periodicity R.
(b) Show that replacing $\partial_a X^{25}$ with $\partial_a X^{25} + A_a$ makes the X^{25} translations a *local* symmetry on the world-sheet.
(c) Add a term $i\phi F_{ab}$ to the world-sheet action, with $\phi(\sigma)$ a new field and F_{ab} the field strength for A_a. The path integral over ϕ forces the field strength to vanish, so the gauge field is trivial: this effectively undoes step (b). Argue that in the gauge $\partial_a A^a = 0$ the resulting theory is just a rewriting of the original one.
(d) Show that you can reach the gauge $X^{25} = 0$, and then carry out the Gaussian path integration over A_a (treat it classically: ignore the determinant). Show that the resulting action has the expected form for the T-dual theory, if ϕ is identified with X'^{25} times an appropriate constant. To get the transformation (8.3.31) of the dilaton one would need to carry the A_a path integration to one loop. Also, to get the coordinate periodicity of X'^{25} requires a somewhat careful argument. (References: Buscher (1987, 1988), Rocek & Verlinde (1992).)

8.4 Verify the equality of the string partition function for several compact dimensions as calculated directly and via the path integral; see the comment below eq. (8.4.12).

8.5 (a) Show that the spectrum (8.4.26) for a single compact dimension defines an even self-dual lattice.
(b) Do the same for the spectrum (8.4.7).

8.6 In the example of two periodic dimensions, section 8.4, show that the general kinetic term (8.4.35) reduces to the form (8.4.39). Verify the assertions about the action of the various dualities on τ and ρ in the special case that τ and ρ are both imaginary.

8.7 (a) Calculate the scattering amplitude for ripples on a Dp-brane. That is, consider the tree-level amplitude with four collective coordinate vertex operators. Note that the polarizations are orthogonal to all momenta, a major simplification.
(b) Examine the Regge and hard scattering limits.
(c) Take the low momentum limit and compare with what is obtained from the action (8.7.2).

8.8 Repeat the previous problem for two coincident Dp-branes. The collective coordinates now include Chan–Paton factors. Note in particular the change in part (c).

8.9 (a) Calculate the expectation value of n exponential vertex operators in the interior of the disk, with Neumann boundary conditions on $p + 1$ coordinates and Dirichlet conditions on $25 - p$.

(b) Extend the calculation to include operators $\partial_a X^M$ in the interior.

8.10 From the result of the previous exercise, calculate the amplitude for a Dp-brane to emit a graviton. Compare with the Dp-brane effective action and the result (8.7.28) for the couplings.

8.11 (a) For the boundary conditions of exercise 8.9, calculate the disk amplitude with two closed string tachyons: this corresponds to scattering a closed string from a D-brane.
(b) Examine the Regge and hard scattering limits.
(c) Show that the amplitude has poles from intermediate open strings. Give a physical interpretation.

8.12 (a) Do the calculation of the previous problem for the messier but more interesting case of massless closed strings.
(b) Show that there are poles in the graviton amplitude due to the open string collective coordinate, and compare with a calculation using the action (8.7.2).
(c) Do the same for the antisymmetric tensor and the open string photon.

9
Higher order amplitudes

In this chapter we discuss higher order amplitudes in string theory. The main goal of the first half of the chapter is an understanding of the finiteness and unitarity of string perturbation theory. After discussing the tree-level amplitude in some detail, we consider general amplitudes, using the idea of factorization to relate higher order amplitudes to lower order ones. In the second half, we briefly discuss some ideas that have been put forward to understand all-order and even nonperturbative string theory. The more powerful methods based on supersymmetry will not be developed until volume two. The present chapter is outside the main line of development in this book and can be skipped on a first reading.

9.1 General tree-level amplitudes

Chapter 6 was primarily concerned with the calculation of some specific tree-level amplitudes. It is useful to begin the present chapter with an analysis of the general tree amplitude, as it illustrates some patterns that will arise in the higher order amplitudes. We focus on the closed oriented string, and on the three- and four-point amplitudes, which illustrate the important features. The main issues are convergence and 'unitarity.' The last term is used somewhat loosely to cover both unitarity of the S-matrix, $SS^\dagger = 1$, and positivity of the Hilbert space. Of course we already know from chapter 4 that the BRST cohomology has a positive inner product, so what we need to show is that amplitudes respect the BRST invariance: that only BRST-invariant states are produced, and that the amplitudes for BRST-equivalent states are equal. We have given a general argument for the latter in section 5.4, but it is useful to illustrate it explicitly.

Before going on to a general analysis let us record some results that are useful in calculation. Suppose we have d infinite flat dimensions and the rest compact (or replaced by a compact unitary CFT), and the tachyon

vertex operator is normalized

$$g_{c,d} :e^{ik\cdot X}: \ . \tag{9.1.1}$$

For example, with toroidal compactification of one dimension we would have $g_{c,25} = g_c(2\pi R)^{-1/2}$ as in the previous chapter. Then for the matter part of a general OCQ vertex operator, the correct normalization can be fixed by the most singular term in the OPE

$$\mathscr{V}_j(k;z,\bar{z})\overline{\mathscr{V}_{j'}(k;0,0)} = \frac{g_{c,d}^2}{z^2\bar{z}^2} + \dots \ . \tag{9.1.2}$$

A bar denotes the Euclidean adjoint (6.7.28). The singularity is $|z|^{-4}$ because the vertex operator is of weight (1, 1). A repetition of the tachyon pole calculation (6.6.7) then determines the normalization

$$e^{-2\lambda}\left\langle \tilde{c}(\bar{z}_1)c(z_1)\,\tilde{c}(\bar{z}_2)c(z_2)\,\tilde{c}(\bar{z}_3)c(z_3)\ :e^{ik\cdot X}(z,\bar{z}):\right\rangle$$
$$= \frac{8\pi i}{\alpha' g_{c,d}^2}|z_{12}z_{13}z_{23}|^2(2\pi)^d\delta^d(\textstyle\sum_i k_i) \ . \tag{9.1.3}$$

So all normalization constants in the S-matrix are known in terms of $g_{c,d} = \kappa_d/2\pi$.

Three-point amplitudes

Let us first express the tree amplitude in several useful forms. The gauge-fixed expression (5.3.9) for the string S-matrix reduces, for the general tree-level, three-point amplitude, to a path integral with three fixed vertex operators:

$$S_{S_2}(1;2;3) = e^{-2\lambda}\left\langle \hat{\mathscr{V}}'_1(\infty,\infty)\hat{\mathscr{V}}_2(1,1)\hat{\mathscr{V}}_3(0,0)\right\rangle_{S_2} \ , \tag{9.1.4}$$

where each numeral 'n' is an abbreviation for string momentum and state, (k_n, j_n). A $\hat{}$ is used to designate a vertex operator in the 'fixed' picture — that is, not including the b-ghost insertions for the integral over position. As we have discussed in section 5.4, this is the form that comes naturally out of the state–operator mapping, and in the OCQ case takes the form

$$\hat{\mathscr{V}}_j = \tilde{c}c\mathscr{V}_j \ . \tag{9.1.5}$$

Using the results from section 6.7 for translating between path integrals on the sphere and matrix elements, this amplitude can also be represented as a matrix element or as an operator product coefficient:

$$S_{S_2}(1;2;3) = e^{-2\lambda}\langle\!\langle \hat{\mathscr{V}}_1|\hat{\mathscr{V}}_2(1,1)|\hat{\mathscr{V}}_3\rangle = e^{-2\lambda}c_{123} \ . \tag{9.1.6}$$

This relates the three-point amplitude in spacetime to a matrix element on the world-sheet, and further to the operator product coefficients which encode the local properties of the CFT.

The first issue is the decoupling of BRST-null states, so that BRST-equivalent states have equal amplitudes. Suppose that the state $|\hat{\mathscr{V}}_3\rangle$ in the three-point amplitude is null,

$$|\hat{\mathscr{V}}_3\rangle = Q_B|\chi\rangle \ . \tag{9.1.7}$$

Then

$$e^{2\lambda}S_{S_2}(1;2;3) = \langle\!\langle \hat{\mathscr{V}}_1|\,[\hat{\mathscr{V}}_2(1,1), Q_B]\,|\chi\rangle + \langle\!\langle \hat{\mathscr{V}}_1|Q_B\hat{\mathscr{V}}_2(1,1)|\chi\rangle$$
$$= 0 \ , \tag{9.1.8}$$

the terms vanishing respectively by the BRST invariance of states 2 and 1. Similarly, if $\hat{\mathscr{V}}_2 = \{Q_B, \chi\}$ or $|\hat{\mathscr{V}}_1\rangle = Q_B|\chi\rangle$ the amplitude also vanishes. We have used the Hilbert space formalism here, but it is worthwhile to phrase this same argument in path integral language. Suppose $\mathscr{V}_3$ is null,

$$\hat{\mathscr{V}}_3 = Q_B \cdot \mathscr{X} \ . \tag{9.1.9}$$

Expand the contour integral on the sphere away from $\mathscr{X}$ until it becomes a small loop around $\hat{\mathscr{V}}_1$ plus a small loop around $\hat{\mathscr{V}}_2$. These then vanish by BRST invariance, $Q_B \cdot \hat{\mathscr{V}}_1 = Q_B \cdot \hat{\mathscr{V}}_2 = 0$.

To this order then, amplitudes are well defined on the physical Hilbert space, the BRST cohomology. The decoupling of the null graviton, associated with spacetime coordinate invariance, is a special case of the decoupling of BRST-null states. It is then natural to wonder whether spacetime coordinate invariance is a piece of some much larger symmetry in string theory. We will discuss some ideas in this direction in section 9.6.

We now consider the unitarity of the S-matrix,

$$\sum_n S_{mn}S_{pn}^* \stackrel{?}{=} \delta_{mp} \ , \tag{9.1.10}$$

which is equivalent to the conservation of probability. It is useful to break up the S-matrix into a term in which nothing happens and a scattering term,

$$S_{mn} = \delta_{mn} + iT_{mn} \ . \tag{9.1.11}$$

For the scattering matrix T, unitarity implies

$$T_{mp} - T_{pm}^* = i\sum_n T_{mn}T_{pn}^* \ . \tag{9.1.12}$$

The T-matrix by definition involves a total of at least three particles, so for three-particle amplitudes on the sphere the right-hand side of eq. (9.1.12) vanishes. Unitarity then requires

$$T_{S_2}(1;2;3) = T_{S_2}(-1;-2;-3)^* \ , \tag{9.1.13}$$

where a minus sign is shorthand for replacing an incoming vertex operator by an outgoing vertex operator for the same state, or vice versa. The

condition (9.1.13) is satisfied if the incoming and outgoing vertex operators are related

$$\hat{\mathscr{V}}_{-i} = -\overline{\hat{\mathscr{V}}_i} \,. \tag{9.1.14}$$

This is a natural definition: conjugation takes $e^{ik\cdot X}$ into $e^{-ik\cdot X}$, turning a vertex operator that injects momentum k^μ into the world-sheet into one that absorbs it, and similarly for other quantum numbers. The minus sign is included to undo the reversal of the order of $\tilde{c}c$ by the adjoint, so that $\hat{\mathscr{V}}_{-i}$ is obtained simply by conjugating all explicit factors of i in the vertex operators. Complex conjugation of the path integral then gives

$$\left\langle \hat{\mathscr{V}}'_1(\infty,\infty)\hat{\mathscr{V}}_2(1,1)\hat{\mathscr{V}}_3(0,0)\right\rangle_{S_2} = -\left\langle \hat{\mathscr{V}}'_{-1}(\infty,\infty)\hat{\mathscr{V}}_{-2}(1,1)\hat{\mathscr{V}}_{-3}(0,0)\right\rangle^*_{S_2} \,, \tag{9.1.15}$$

implying the unitarity relation (9.1.13). Naively, the path integral is manifestly real and so the relation (9.1.15) should hold without the minus sign. However, we must remember that the path integral has to be defined by specific analytic continuations, and the explicit result (6.2.6) acquired an overall factor of i in this way.

Four-point amplitudes and world-sheet duality

The connected four-point amplitude is

$$S_{S_2}(1;2;3;4) = e^{-2\lambda}\int_{\mathrm{C}} d^2z \left\langle \hat{\mathscr{V}}'_1(\infty,\infty)\,\hat{\mathscr{V}}_2(1,1)\,\hat{\mathscr{V}}_3(0,0)\,\mathscr{V}_4(z,\bar{z})\right\rangle_{S_2} . \tag{9.1.16}$$

Writing the expectation value as a matrix element gives

$$\langle\!\langle \hat{\mathscr{V}}_1|\mathrm{T}[\,\hat{\mathscr{V}}_2(1,1)\,\mathscr{V}_4(z,\bar{z})\,]|\hat{\mathscr{V}}_3\rangle \,. \tag{9.1.17}$$

The ghost expectation value is just a constant for OCQ vertex operators, but to make the connection with the more general case it is useful to move the ghosts around a bit. Using a contour argument one can write the ghost insertion for the d^2z integration as

$$\mathscr{V}_4(z,\bar{z}) = (z\bar{z})^{-1}\{b_0,[\tilde{b}_0,\hat{\mathscr{V}}_4(z,\bar{z})]\} \,. \tag{9.1.18}$$

Using this and the condition (4.3.29) that b_0 and $\tilde{b}_0$ annihilate the physical states, the matrix element becomes

$$\begin{aligned}&\theta(1-|z|)\langle\!\langle \hat{\mathscr{V}}_1|\hat{\mathscr{V}}_2(1,1)b_0\tilde{b}_0 z^{L_0-1}\bar{z}^{\tilde{L}_0-1}\hat{\mathscr{V}}_4(1,1)|\hat{\mathscr{V}}_3\rangle \\ &\quad+\theta(|z|-1)\langle\!\langle \hat{\mathscr{V}}_1|\hat{\mathscr{V}}_4(1,1)b_0\tilde{b}_0 z^{-L_0-1}\bar{z}^{-\tilde{L}_0-1}\hat{\mathscr{V}}_2(1,1)|\hat{\mathscr{V}}_3\rangle \,.\end{aligned} \tag{9.1.19}$$

Let us also write the matrix element after insertion of a complete set of

intermediate states,

$$\begin{aligned}&\theta(1-|z|)\sum_{i,i'} z^{\alpha'(k_i^2+m_i^2)/4-1}\bar{z}^{\alpha'(k_i^2+\tilde{m}_i^2)/4-1} \\ &\qquad\times\langle\!\langle \hat{\mathscr{V}}_1|\hat{\mathscr{V}}_2(1,1)b_0\tilde{b}_0|i\rangle\mathscr{G}^{ii'}\langle\!\langle i'|\hat{\mathscr{V}}_4(1,1)|\hat{\mathscr{V}}_3\rangle \\ &+\theta(|z|-1)\sum_{i,i'} z^{-\alpha'(k_i^2+m_i^2)/4-1}\bar{z}^{-\alpha'(k_i^2+\tilde{m}_i^2)/4-1} \\ &\qquad\times\langle\!\langle \hat{\mathscr{V}}_1|\hat{\mathscr{V}}_4(1,1)b_0\tilde{b}_0|i\rangle\mathscr{G}^{ii'}\langle\!\langle i'|\hat{\mathscr{V}}_2(1,1)|\hat{\mathscr{V}}_3\rangle \,.\end{aligned} \tag{9.1.20}$$

In this last form the behavior of the integral at $z\to 0$ is evident, and it is the same as for the tachyon amplitudes studied in chapter 6. If $k_i^2+m_i^2$ and $k_i^2+\tilde{m}_i^2$ are positive for all i, that is,

$$k_i^2 = (k_1+k_2)^2 > \frac{4}{\alpha'} \,, \tag{9.1.21}$$

then the integral converges. Elsewhere it is defined as in eq. (6.6.7),

$$\int_{|z|<1} d^2z\, z^{\alpha'(k^2+m^2)/4-1}\bar{z}^{\alpha'(k^2+\tilde{m}^2)/4-1} = \frac{8\pi}{\alpha'}\frac{\delta_{m^2,\tilde{m}^2}}{k^2+m^2-i\epsilon} \,. \tag{9.1.22}$$

One can think of this as analytic continuation from large positive values of m^2, keeping m^2 slightly below the real axis. Similarly, the integral as $z\to\infty$ converges for $(k_1+k_4)^2>4/\alpha'$ and is defined elsewhere by analytic continuation. The form (9.1.20) is not useful for studying the $z\to 1$ limit. Instead use the symmetry of the amplitude to interchange vertex operators 2 and 3. The integral for $z\to 1$ is then defined by analytic continuation from the convergent region $(k_1+k_3)^2>4/\alpha'$. The need to decompose the range of integration and to make possibly different analytic continuations for the different regions poses no difficulty.

After the expansion (9.1.20) is integrated as in (9.1.22), the amplitude is expressed as a series of poles in s and u. We know from chapter 6 that the amplitude also has poles in t, but these are not explicit: they come from the divergences of the infinite sums. Similarly the open string amplitude $I(s,t)$ has poles in s and t, but can be written as an infinite sum of terms having poles only in s, or only in t. This property is known as *duality*, or *world-sheet duality* to distinguish it from T- and other dualities. We can think of this as arising because the world-sheet can be cut open on different curves and so written as sums over intermediate states in different ways. Another example is the sum over open string states giving closed string poles on the cylinder.

Unitarity of the four-point amplitude

Now let us check that four-point amplitudes are equal for equivalent states. Let $|\hat{\mathscr{V}}_3\rangle = Q_{\mathrm{B}}|\chi\rangle$ be null. Using the form (9.1.19), we can commute

Q_B to the left until it annihilates $\langle\langle \hat{\mathscr{V}}_1|$. The BRST operator commutes with L_0, $\tilde{L}_0$ and with the vertex operators, but not with b_0 and $\tilde{b}_0$. The nonzero commutators involve

$$\left[b_0\tilde{b}_0 z^{\pm L_0-1}\bar{z}^{\pm\tilde{L}_0-1}, Q_B\right] = \pm b_0\partial_{\bar{z}}\left(z^{\pm L_0-1}\bar{z}^{\pm\tilde{L}_0}\right) \mp \tilde{b}_0\partial_z\left(z^{\pm L_0}\bar{z}^{\pm\tilde{L}_0-1}\right) . \tag{9.1.23}$$

These are total derivatives, illustrating the general principle (5.4.6) that the couplings of null states are proportional to total derivatives on moduli space. We have already seen an example of this for the massless vertex operators in section 6.6. As in that example, the integral vanishes by the canceled propagator argument. The integral at $z \to 0$ is defined by analytic continuation from large positive $(k_1+k_2)^2$. The surface term vanishes identically there, and so by analyticity must vanish everywhere. The same argument applies to the other limits of moduli space.

For the four-particle amplitude on the sphere, the unitarity condition is

$$\begin{aligned} &T_{S_2}(1;2;3;4) - T_{S_2}(-1;-2;-3;-4)^* \\ &\quad = i\int \frac{d^{d-1}\mathbf{k}_5}{2E_5(2\pi)^{d-1}} \sum_{\substack{j_5 \\ k_5^0=\pm\omega_5}} T_{S_2}(1;2;5)T_{S_2}(-5;3;4) \\ &\qquad\qquad + 2 \text{ permutations} \\ &\quad = -ie^{-4\lambda}\int \frac{d^{d-1}\mathbf{k}_5}{2E_5(2\pi)^{d-1}} \sum_{\substack{j_5 \\ k_5^0=\pm\omega_5}} \langle\langle \hat{\mathscr{V}}_1|\hat{\mathscr{V}}_2(1,1)|\hat{\mathscr{V}}_5\rangle\langle\langle \hat{\mathscr{V}}_{-5}|\hat{\mathscr{V}}_3(1,1)|\hat{\mathscr{V}}_4\rangle \\ &\qquad\qquad + 2 \text{ permutations} . \end{aligned} \tag{9.1.24}$$

The intermediate particle is on-shell, the sum j_5 is over a complete set of physical states, and the right-hand side is summed over all ways to break 1234 into pairs: 12–34, 13–24, and 14–23.

Eq. (9.1.24) is what we wish to show. Using the reality condition (9.1.13) in the matrix element (9.1.20), the left-hand side of eq. (9.1.24) vanishes in the region where the z integral converges. The nonzero part comes from the $i\epsilon$ introduced in the analytic continuation,

$$\frac{1}{k^2+m^2-i\epsilon} - \frac{1}{k^2+m^2+i\epsilon} = 2\pi i\delta(k^2+m^2) . \tag{9.1.25}$$

Thus,

$$\begin{aligned} &T_{S_2}(1;2;3;4) - T_{S_2}(-1;-2;-3;-4)^* \\ &\quad = \frac{16\pi^2 i}{\alpha' e^{2\lambda}} \sum_{i,i'} \delta_{m_i^2,\tilde{m}_i^2}\delta(k_i^2+m_i^2)\langle\langle \hat{\mathscr{V}}_1|\hat{\mathscr{V}}_2(1,1)b_0\tilde{b}_0|i\rangle\mathscr{G}^{ii'}\langle\langle i'|\hat{\mathscr{V}}_3(1,1)|\hat{\mathscr{V}}_4\rangle \\ &\qquad\qquad + 2 \text{ permutations} . \end{aligned} \tag{9.1.26}$$

The resemblance to the desired unitarity relation (9.1.24) is obvious, and a little bookkeeping and definition-chasing established the result. The index i in the sum (9.1.26), running over a complete set of string states, labels both internal state and momentum,

$$i \leftrightarrow (j,k) . \tag{9.1.27}$$

The delta functions and $b_0\tilde{b}_0$ factor in eq. (9.1.26) project onto the space $\hat{\mathscr{H}}$ annihilated by L_0, $\tilde{L}_0$, b_0, and $\tilde{b}_0$, which was introduced in the discussion of the BRST cohomology. Thus,

$$\begin{aligned} &\sum_{i,i'} \delta_{m_i^2,\tilde{m}_i^2}\delta(k_i^2+m_i^2)b_0\tilde{b}_0|i\rangle\mathscr{G}^{ii'}\langle\langle i'| \\ &\quad = \delta_{L_0,\tilde{L}_0}\delta(4L_0/\alpha')b_0\tilde{b}_0 \\ &\quad = -\frac{i\alpha'}{16\pi^2 e^{2\lambda}}\int \frac{d^{d-1}\mathbf{k}}{2E(2\pi)^{d-1}} \sum_{\substack{j\in\hat{\mathscr{H}} \\ k^0=\pm\omega_k}} |\hat{\mathscr{V}}_j(k)\rangle\langle\langle \overline{\hat{\mathscr{V}}_j(k)}| . \end{aligned} \tag{9.1.28}$$

The normalization of the basis $|\mathscr{V}_j(k)\rangle$ follows from the OPE (9.1.2) and the expectation value (9.1.3), which imply

$$\langle\langle \overline{\hat{\mathscr{V}}_j(k)}|\tilde{c}_0 c_0|\hat{\mathscr{V}}_{j'}(k')\rangle = \frac{8\pi i e^{2\lambda}}{\alpha'}\delta_{jj'}(2\pi)^d\delta^d(k-k') . \tag{9.1.29}$$

Using the relation (9.1.28), the form and normalization of eq. (9.1.26) agree with the unitarity condition (9.1.24), leaving one important loose end. The sum in the unitarity relation runs only over the physical Hilbert space, the BRST cohomology, while the final sum in (9.1.28) runs over a complete basis for $\hat{\mathscr{H}}$. We now show that these sums are equal: Consider the following basis for $\hat{\mathscr{H}}$. First, take a complete set of 'unphysical states' $|a\rangle_U$. That is, no linear combination of these is annihilated by Q_B. Second, take a complete set of BRST-invariant states $|b\rangle_P$ which are independent in the sense that no linear combination of them is null. Finally, take the complete set of BRST-null states $|a\rangle_N = Q_B|a\rangle_U$ to complete the basis. Define the operators $\mathscr{P}$ and $\mathscr{U}$:

$$\mathscr{P}|b\rangle_P = |b\rangle_P , \quad \mathscr{P}|a\rangle_U = 0 , \quad \mathscr{P}|a\rangle_N = 0 ; \tag{9.1.30a}$$

$$\mathscr{U}|b\rangle_P = 0 , \quad \mathscr{U}|a\rangle_U = 0 , \quad \mathscr{U}|a\rangle_{NS} = |a\rangle_U . \tag{9.1.30b}$$

Then

$$1 = \mathscr{P} + Q_B\mathscr{U} + \mathscr{U}Q_B . \tag{9.1.31}$$

Insert this decomposition into the factorization relation. The first term projects onto one state from each cohomology class, the physical spectrum appearing in the unitarity sum. The second and third terms each contain Q_B, and so vanish by the earlier argument that null states decouple. Thus unitarity is established.

It is worth noting that the argument just given is very similar to the argument that amplitudes in gauge theory are simultaneously Lorentz-

invariant and unitary. The numerator of the covariant photon propagator

$$-i\frac{\eta^{\mu\nu}}{k^2}\,, \tag{9.1.32}$$

has a negative eigenvalue, corresponding to a negative norm intermediate state. Choose any null vector n^μ such that $n\cdot k = 1$, and decompose

$$\eta^{\mu\nu} = (\eta^{\mu\nu} - k^\mu n^\nu - n^\mu k^\nu) + k^\mu n^\nu + n^\mu k^\nu\,. \tag{9.1.33}$$

At $k^2 = 0$, the first term has $d-2$ positive eigenvalues and no negative eigenvalues. If the photon is coupled to conserved currents at both ends, the last two terms vanish, so the unitarity relation is satisfied with just the $d-2$ positive norm states expected for a photon in d dimensions. The decomposition (9.1.31) is the generalization of eq. (9.1.33), with $k^\mu \to |a\rangle_{\rm N}$ and $n^\mu \to |a\rangle_{\rm U}$.

To conclude, let us review how various general principles are illustrated by this example. BRST null states decouple, but only after integration on moduli space. The potential surface terms from the latter vanish by the canceled propagator argument. The ghost manipulations leading to eq. (9.1.19) have a natural interpretation in terms of the Riemann surface picture from chapter 5. Consider the sphere with vertex operators at $0, 1, z, \infty$. To be specific take $|z| > 1$ as in the second term of eq. (9.1.19). We can think of this sphere as being formed from two disks, with coordinates z_1 and z_2. Put vertex operators at $z_1 = 0, 1$ and at $z_2 = 0, 1$, and join the disks by the identification $z_1 = z/z_2$. For the modulus z, we have

$$\left.\frac{\partial z_1}{\partial z}\right|_{z_2} = \frac{z_1}{z}\,. \tag{9.1.34}$$

The ghost insertion (5.4.15) for dz is then b_0/z, and similarly for $d\bar z$, as illustrated by eq. (9.1.19).

9.2 Higher genus Riemann surfaces

There are many different representations of higher genus Riemann surfaces. The one that will be most useful for our purposes is based on the idea of building higher genus surfaces by sewing together lower genus surfaces. This will be described in the next section, but first for completeness we describe some other common constructions. The treatment is necessarily rather brief.

Schottky groups

Imagine removing two disks from a sphere and joining the edges of the resulting holes. This produces a handle, so the surface obtained in this way is a torus. Repeating g times, with a total of $2g$ holes, produces a genus g surface.

We can make this more precise as follows. Regarding the sphere as the complex z-plane plus point at infinity, consider the mapping

$$\gamma:\quad \frac{z'-U}{z'-V} = K\frac{z-U}{z-V}\,, \tag{9.2.1}$$

which depends on three complex parameters U, V, and K. This is a Möbius transformation, and in particular is holomorphic and one-to-one. The points U and V are fixed under γ. Taking $|K| > 1$, the surface expands away from U and contracts toward V. A fundamental region for γ is the annular region

$$|K|^{-1/2} < \left|\frac{z-U}{z-V}\right| < |K|^{1/2}\,. \tag{9.2.2}$$

Identifying points under γ glues together the inner and outer boundaries of the annulus (9.2.2), producing a torus. The group generated by g such elements (9.2.1) produces the genus g surface, with restrictions on the parameters so that the sewn circles do not overlap.

Every closed oriented Riemann surface can be obtained in this way. The range of the parameters is not simple, in part because some of the modular transformations act in a complicated way. Riemann surfaces with boundary and unoriented Riemann surfaces can be constructed by dividing by one or two reflections of a closed oriented surface, as we have illustrated in the lower genus examples.

Let us see that the Schottky construction produces the expected number of moduli. Each transformation (9.2.1) has three complex parameters for a total of $3g$. One can think of the six real parameters for each handle as being two for the position of each hole, one for the product of the radii of the holes (only the product, because if one hole is made smaller and the other larger we get the same surface), and one for the relative angle at which the holes are sewn. There is redundancy, because an overall Möbius transformation of the sphere gives an equivalent surface. Subtracting the six real parameters for this group leaves $6g-6$, the correct number of moduli for genus $g \geq 2$ surfaces. For $g = 1$, two Möbius transformations leave the torus invariant, so two real parameters remain.

The Schottky parameterization shows that the moduli space has natural complex coordinates. The path integral has certain holomorphicity properties with respect to these coordinates, though we will not emphasize this. Having constructed the general surface as a coset of the sphere, it is evident that Green's functions can be constructed by use of the method of images, summing over all images of a source. Functional determinants and other objects of interest can be expressed similarly.

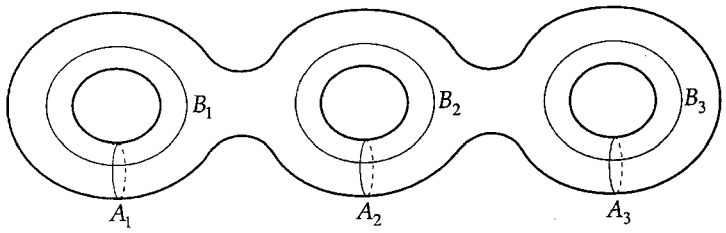


Fig. 9.1. Genus-three surface with A- and B-cycles marked.

Fuchsian groups

For this construction, consider first the world-sheet metric. It is a theorem that by a Weyl transformation one can always make the world-sheet curvature R a constant. At genus zero this gives the round sphere, at genus one the flat torus, and at higher genus a surface of constant negative curvature. The simply connected covering space for any of these will similarly have constant curvature. There are precisely three such simply connected spaces — the round sphere again, the flat plane (the covering space of the torus) and the upper half-plane with metric

$$ds^2 = \frac{dz d\bar{z}}{(\mathrm{Im}\, z)^2} \,. \tag{9.2.3}$$

The last is the covering space for all the genus $g \geq 2$ surfaces.

Correspondingly, all genus $g \geq 2$ surfaces can be obtained from the upper half-plane by dividing by the action of some discrete group. The elements of this group must be isometries of the metric (9.2.3); these isometries are just the Möbius transformations with real coefficients, $PSL(2, \mathbf{R})$. Figure 9.1 shows the genus-three surface with some closed curves drawn in, an *A-cycle* and *B-cycle* for each handle. These are nontrivial curves, in that they cannot be contracted to points. If we traverse one of these curves, on the covering space we come back not to the original point but to one of those identified with it. Thus to each of the cycles is associated an element of $PSL(2, \mathbf{R})$, also denoted A_i or B_i.

The A- and B-cycles are complete, in that every closed path is equivalent (homotopic) to a succession of A- and B-cycles. In fact they are slightly overcomplete: the path

$$(A_1 B_1^{-1} A_1^{-1} B_1)(A_2 B_2^{-1} A_2^{-1} B_2) \ldots (A_g B_g^{-1} A_g^{-1} B_g) \tag{9.2.4}$$

can be contracted to a point, and the corresponding product must be the identity. Counting moduli, there are three real parameters in each $PSL(2, \mathbf{R})$ element, giving $6g$, minus three from the constraint (9.2.4), minus three more because overall conjugation by $PSL(2, \mathbf{R})$ gives an equivalent surface, for $6g - 6$ in all.

Although the Schottky and Fuchsian parameterizations are similar, they have rather different properties. In particular, the complex structure is much less evident in the latter, and the extension to the superstring correspondingly less simple.

The period matrix

The modulus τ of the torus has a natural $g \times g$ matrix generalization at genus g. To see this, we need first the *Abelian differentials* or holomorphic one-forms. These are world-sheet fields ω_a with the property that in complex coordinates

$$\partial_{\bar{z}} \omega_z = 0 \,, \quad \omega_{\bar{z}} = 0 \,. \tag{9.2.5}$$

There are precisely g such fields, and g more satisfying the conjugate relation. This follows from the Riemann–Roch theorem (5.3.22). The Abelian differentials (9.2.5) and their conjugates are just the kernel of the operator P_0^T, for which the Riemann–Roch theorem is

$$\dim \ker P_0^T - \dim \ker P_0 = -\chi = 2g - 2 \,. \tag{9.2.6}$$

The operator P_0 acts on scalar fields c and $\tilde{c}$ with $h = 0$, and so reduces to the ordinary gradient. This always has two zero modes ($c =$ constant or $\tilde{c} =$ constant), and so $\dim \ker P_0^T = 2g$.

The line integral of a one-form is a coordinate-invariant quantity. A natural basis for the Abelian differentials is then ω_{zi}, $i = 1, \ldots, g$, defined by

$$\oint_{A_i} dz\, \omega_{zj} = \delta_{ij} \,. \tag{9.2.7}$$

Then the *period matrix*

$$\tau_{ij} = \oint_{B_i} dz\, \omega_{zj} \tag{9.2.8}$$

characterizes the complex structure of the surface. At genus one, where ω_z is constant, this is the usual τ.

It can be shown that τ_{ij} is symmetric, so it has $\frac{1}{2}g(g+1)$ complex parameters. For $g = 1, 2, 3$ this is equal to the number of complex moduli, but for $g \geq 4$ it is greater, so not all period matrices correspond to Riemann surfaces. The theta function (7.2.36) has a natural genus-g generalization, where τ becomes the period matrix and v, n, a, b become g-component vectors.

Hyperelliptic surfaces

Here is another description which is sometimes useful. Consider the space $S_2 \times S_2$, with coordinates z and y. Consider the two-dimensional submanifold given by the equation

$$y^2 = \prod_{i=1}^{2k}(z - Z_i) \tag{9.2.9}$$

for complex parameters Z_i. Projecting down to the z-coordinate, there are two sheets joined across k branch cuts. This produces a surface with $g = k - 1$ handles. The branch points are singular when projected down to z, but the full surface (9.2.9) is nonsingular there: y is locally a good coordinate.

The surface depends on the $2k$ complex Z_i, but three of these are again redundant under a Möbius transformation, leaving $2k-3 = 2g-1$ complex parameters. For $g = 1, 2$ this is the same as the number of moduli, but for $g \geq 3$ the number of moduli is greater and only special surfaces have the form (9.2.9). These are known as *hyperelliptic surfaces.*

9.3 Sewing and cutting world-sheets

A useful concept for analyzing string amplitudes is the relation between higher and lower genus amplitudes. A key ingredient is the *plumbing fixture*, for building up higher genus surfaces from lower ones. Consider two world-sheets, M_1 with coordinate z_1 in some patch, and M_2 with coordinate z_2 in some patch. We can join these to form a new world-sheet M as follows. Let q be a complex parameter. Cut out from M_1 the disk $|z_1| < (1-\epsilon)|q|^{1/2}$ for some small ϵ, and from M_2 the disk $|z_2| < (1-\epsilon)|q|^{1/2}$. Now form a single surface by identifying pairs of points such that

$$z_1 z_2 = q \ . \tag{9.3.1}$$

This is shown in figure 9.2. The annular region

$$(1-\epsilon)^{-1}|q|^{1/2} > |z_1| > (1-\epsilon)|q|^{1/2} \tag{9.3.2}$$

is sewn as indicated. One notation for this construction is

$$M = M_1 \infty M_2(z_1, z_2, q) \ . \tag{9.3.3}$$

The notation emphasizes that the resulting surface depends on the choice of local coordinates. With different local coordinates z_1' and z_2', one would be cutting out different circles and the joined surface will in general be different. The two patches might also be in different parts of a single

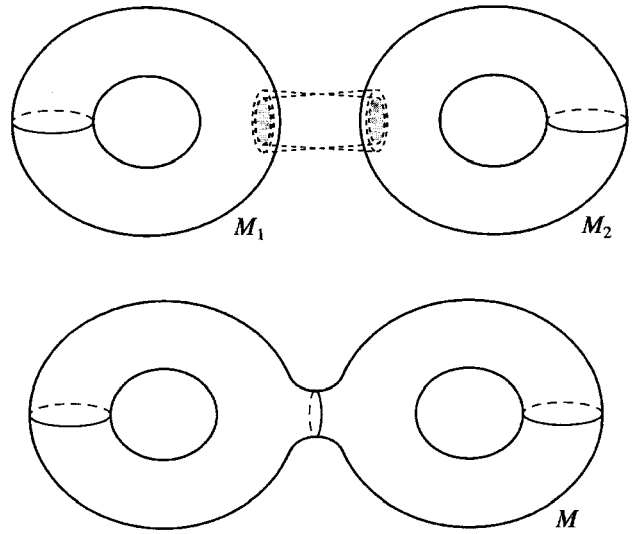


Fig. 9.2. Joining two world-sheets by removing a disk (shaded) from each and gluing the edges as indicated by the dashed lines.

surface M_0. The plumbing fixture construction then adds a handle to the surface. This is denoted

$$M = M_0 8(z_1, z_2, q) \ . \tag{9.3.4}$$

The symbols '∞' and '8' are supposed to depict the two sewing processes. The parameter q is not actually needed — taking $z_1' = z_1 q^{-1/2}$, $z_2' = z_2 q^{-1/2}$, and $q' = 1$ gives an equivalent surface. It is useful to introduce q in order to discuss the behavior as a function of q with fixed coordinate patches.

We can also reverse this construction. Given any Riemann surface M, and any non-self-intersecting curve C on it, one can always find holomorphic coordinates $z_1 = z_2^{-1}$ in a neighborhood of C such that C is given by $|z_1| = |z_2| = |q|^{1/2}$. Cutting on C (undoing the identification $z_1 z_2 = q$) and gluing in the disks $|z_1| < |q|^{1/2}$ and $|z_2| < |q|^{1/2}$ produces M_1 and M_2 (or M_0) such that M would be reconstructed by the sewing operation — either (9.3.3) or (9.3.4) depending on whether the cut along C makes the surface disconnected.

For the genus-two example in figure 9.2, consider the case that $q \ll 1$. Figure 9.3(a) shows M in this case. We have smoothly interpolated the metric between $dz_1 d\bar{z}_1$ and $dz_2 d\bar{z}_2$ in the overlap region. The surface

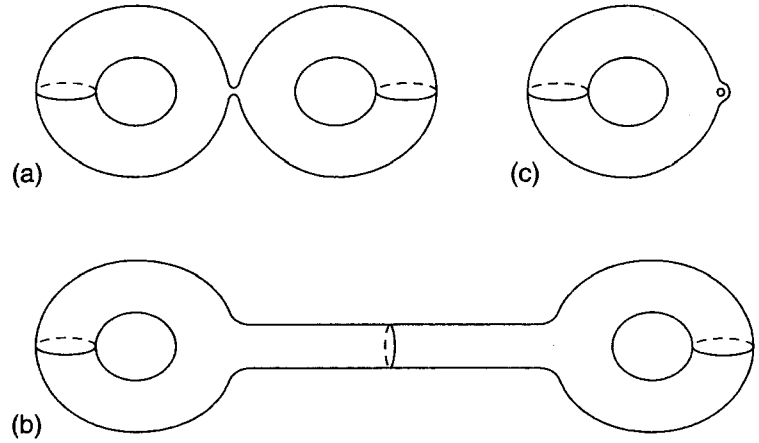


Fig. 9.3. Three representations of the genus two surface with degenerating cycle.

pinches down to a radius of order $|q|^{1/2}$. By a Weyl transformation, one can find a metric in which the radius stays of order 1, as shown in figure 9.3(b). Since a Weyl transformation acts isotropically, the effect is that the *length* of the handle becomes large, of order $-\ln|q|$. The handle region $|z_1| < 1$, $|z_2| < 1$ is equivalent to the cylinder

$$\ln|q| < \mathrm{Im}\, w < 0\,, \quad w \cong w + 2\pi\,, \tag{9.3.5}$$

where $z_1 = \exp(-iw)$ and $z_2 = q\exp(iw)$. Figure 9.3(c) shows one more representation of the same surface. The annular region $O(|q|) < |z| < O(1)$ surrounding the small handle is conformally equivalent to the long cylinder of figure 9.3(b). In terms of figure 9.3(a), one can think of this picture as a Weyl transformation where all of M_2 has been reduced.

In all the pictures, it is clear that the surface becomes special in the $q \to 0$ limit. The handle is said to *pinch* or *degenerate*. This is an example of a *boundary of moduli space*. The boundary of moduli space is important, because any divergences must come from there. In fact, we will see that every boundary is of this form, where one or more handles pinches. For the genus-two surface, there is one other way for a single handle to degenerate, if one of the A- or B-cycles pinches. This would correspond to building the genus-two surface from a single torus via the '8' operation (rather than from two tori via the '∞' operation) and taking $q \to 0$.

It is a useful fact that by repeated use of the cutting operation, every Riemann surface can be reduced to spheres with three punctures. A puncture is a marked point, with a specified complex coordinate near

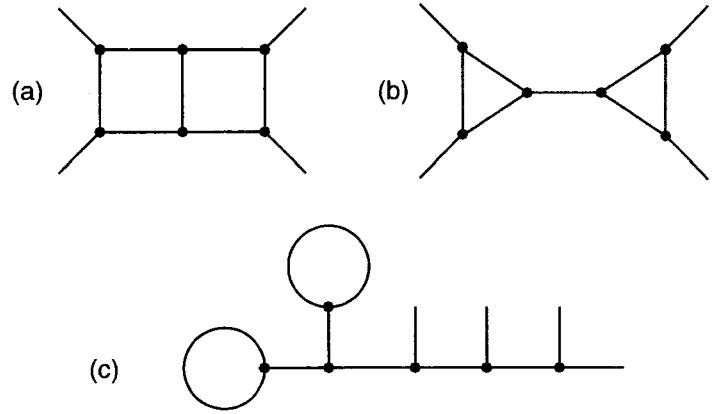


Fig. 9.4. Some sewing constructions of the genus-two surface with four operators. Each vertex is a sphere with three punctures, each internal line represents the plumbing fixture construction, and each external line represents a vertex operator.

the point; this is the data needed to sew in a plumbing fixture. In other words, every Riemann surface can be constructed from spheres with three punctures via sewing. For example, figure 9.4 shows three of the many ways to construct the genus-two surface with four vertex operators. There is one complex modulus q for each sewing operation, giving seven in all for each of the constructions in figure 9.4. These correspond to the $3g - 3 = 3$ complex moduli for the genus-two surface plus the positions of the four operators.

A graphical decomposition

The graphs in figure 9.4 are reminiscent of Feynman diagrams. We now describe a Feynman-diagram-like decomposition of moduli space. The vertices are rather more complicated than in figure 9.4, with any number of external lines and with internal corrections from string loops. The point of the construction is to make explicit the nature of the boundary of moduli space.

The statement is that the integral over moduli space is given by a sum of Feynman graphs. Each internal propagator in a graph represents the sewing operation, with the sewing parameter q integrated over some standard region such as $|q| < 1$. The n-point vertex represents a Riemann surface with n punctures. We will see that the surfaces in the vertices must be integrated over some *interior* region of moduli space, and also summed over genus. The sewing operation refers to a specific complex

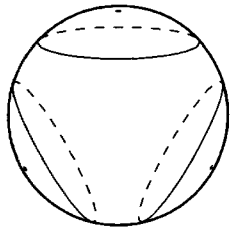


Fig. 9.5. Trilinear string vertex, with three marked points and three sewing circles on the sphere.

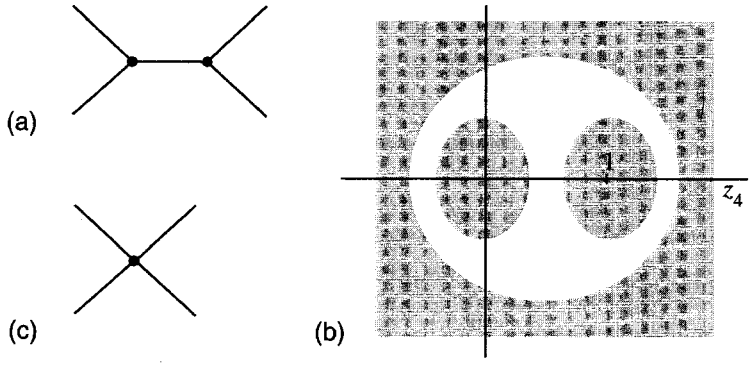


Fig. 9.6. (a) Constructing the sphere with four vertex operators. (b) The region of the z-plane covered by this graph. (c) The additional vertex needed to give the correct amplitude. The four-punctured sphere is integrated over the unshaded region.

coordinate for each point to be sewn, so the definition of the vertex must include such a choice at each point. The set of genus-g Riemann surfaces appearing in the n-point vertex is denoted $\mathscr{V}_{g,n}$. External lines represent vertex operators.

Let us illustrate this construction. Start with a three-point vertex $\mathscr{V}_{0,3}$, shown in figure 9.5 as a sphere with three marked points and three *sewing circles*. Sewing circles are unit circles in the complex coordinates associated with the respective points. Build a sphere with four points by taking two of these vertices, removing the interior of one sewing circle from each, and sewing them to the ends of a cylinder; this is equivalent to the plumbing fixture construction. This corresponds to the Feynman graph of figure 9.6(a); the length of the cylinder (with the circumference scaled to 2π) is $-\ln|q|$. Integrating over $|q| < 1$ and summing over the other two channels gives only part of the integration region for the sphere with four points, as shown in figure 9.6(b). The shaded regions are included but the unshaded interior region is missed.

Let us make this more explicit. For the trilinear vertex, put the points on the sphere at the standard $0, 1, \infty$ and choose coordinate frames $z^{(0)}$, $z^{(1)}$, and $z^{(\infty)}$ at each, so the three sewing circles are $|z^{(0,1,\infty)}| = 1$. The choice should be invariant (up to phase) under the six Möbius transformations that permute the vertex operators, so that the vertex is symmetric. There are many possible choices, for example

$$z^{(0)} = \frac{az}{z-2}\,, \quad z^{(1)} = \frac{a(z-1)}{z+1}\,, \quad z^{(\infty)} = \frac{a}{2z-1}\,, \tag{9.3.6}$$

for any constant a. For $a \geq 3$, the sewing circles do not overlap, guaranteeing that the plumbing fixture produces a good surface for all $q \leq 1$. As the reader can verify, sewing the $z^{(\infty)}$ and $z'^{(0)}$ patches from two spheres with parameter q produces a sphere with vertex operators at 0, 1, $\frac{1}{2}(1+a^2/q)$, and $\frac{1}{2}(1-a^2/q)$. This is equivalent under a Möbius transformation to vertex operators at $0, 1, \infty$, and

$$z_4 = (q-a^2)^2/4a^2q\,. \tag{9.3.7}$$

As q runs over the unit disk, z_4 covers the outer shaded region of figure 9.6(b). Permuting the external vertex operators gives the other two regions. We could try to take a smaller, so the shaded regions grow, but they will eventually overlap and double-count some surfaces while other surfaces will still not be produced. They will never precisely fill the plane, so we need to add in an explicit four-point vertex $\mathscr{V}_{0,4}$ which is given by the sphere with four marked points, integrated over the unshaded part of the z-plane. This does not define $\mathscr{V}_{0,4}$ completely because there is also a choice of coordinate at each puncture for each value of z_4. It is always possible to make this choice in such a way as to connect smoothly at the boundary between the shaded and unshaded regions. Note that the Feynman graph of figure 9.6(a) includes all the asymptotic regions $z_4 \to 0, 1, \infty$, while the four-point vertex includes only nondegenerate surfaces. The behavior in the limits is then given explicitly by the $q \to 0$ limit of the plumbing fixture: this is why this construction is useful.

As another example, sewing two of the circles of a single vertex to the ends of a cylinder produces a torus with one vertex operator. This covers the limit $\mathrm{Im}\,\tau \to \infty$ of the moduli space of the torus, but an interior region is missed and must be added by hand as a vertex region $\mathscr{V}_{1,1}$.

This construction can be extended inductively. The sewing operation always increases the parameter

$$r = 3g + n - 2\,. \tag{9.3.8}$$

The vertex $\mathscr{V}_{0,3}$, the original sphere with three punctures, has $r = 1$. The 8 operation gives $r = r_0 + 1$, while the ∞ operation gives $r = r_1 + r_2$. Thus, $\mathscr{V}_{0,4}$ and $\mathscr{V}_{1,1}$ both have $r = 2$. It has been shown that one can construct $\mathscr{V}_{g,n}$ inductively in r, and so cover moduli space as claimed.

A more explicit construction is as follows (the proofs can be found in the references). On any Riemann surface there is a unique *minimal area metric*, defined as the metric of least area subject to the constraint that every nontrivial closed curve (those which cannot be continuously contracted to a point) has length at least 2π. Every point on the surface lies on at least one *saturating geodesic*, a nontrivial closed curve of length exactly 2π. Saturating geodesics that can be continuously deformed into one another (that is, are homotopically equivalent) cannot intersect. Each set of homotopically equivalent saturating geodesics forms a band (topologically an annulus) on the surface. The different bands cover the surface completely, and may overlap. The height h of such a band is defined as the minimum distance between its two bounding geodesics. The idea of a minimal area metric can be extended to surfaces with punctures, with the specification that a curve that circles a puncture is regarded as nontrivial. The minimal area metric near a puncture is of the form $dzd\bar{z}/z\bar{z}$, a semi-infinite cylinder of radius 2π. There are thus two kinds of bands: the finite ones in the interior of the surface, and the semi-infinite ones centered on the punctures, terminating in the interior of the surface.

We can now describe $\mathscr{V}_{g,n}$ in this construction. Take all Riemann surfaces with n punctures, with the restriction that the height of any internal band, in the minimal area metric, is less than 2π. Cut off the semi-infinite cylinders, leaving a stub of length π, and glue in the usual unit disk with vertex operator in the end of the stub. With these vertices, the Feynman graph sum covers moduli space exactly once. It can be shown that the sewing of these minimal area surfaces still produces a minimal area surface. Bands of height less than 2π are explicitly included in the vertices. Bands of height greater than 2π are produced by the sewing operation on the stubs, the resulting height being 2π from the stubs plus $-\ln|q|$ from the plumbing fixture; this runs from 2π to ∞ when q is integrated. The minimum length and maximum height restrictions guarantee that degenerating surfaces do not appear in the vertices, so they arise only from one or more qs going to zero, as claimed.

9.4 Sewing and cutting CFTs

We have described higher genus Riemann surfaces as built up by sewing together lower genus surfaces. Using the sewing properties of path integrals, the CFT on the higher genus surface can similarly be related to

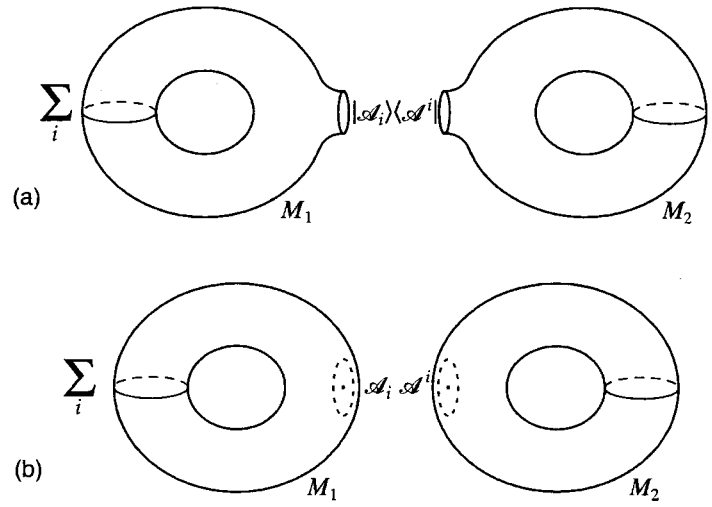


Fig. 9.7. (a) The path integral on M is cut open by inserting a complete set of boundary states. (b) Each state is replaced by a disk with local operator.

lower order amplitudes. The relation is shown pictorially in figure 9.7. For clarity we first discuss the case $|q| = 1$. The path integral on M is cut open on the circle $|z_1| = |z_2| = 1$, inserting a complete set of boundary states. Each boundary state is then replaced by a disk with the corresponding local operator, leaving the uncut world-sheets M_1 and M_2. The precise form of the result is

$$\langle \ldots_1 \ldots_2 \rangle_M = \sum_{ij} \left\langle \ldots_1 \mathscr{A}_i^{(z_1)} \right\rangle_{M_1} \mathsf{G}^{ij} \left\langle \mathscr{A}_j^{(z_2)} \ldots_2 \right\rangle_{M_2}. \tag{9.4.1}$$

We are allowing arbitrary additional insertions $\ldots_1$ on M_1 and $\ldots_2$ on M_2; these must of course lie outside the circles $|z_1| = 1$ and $|z_2| = 1$ that are being cut. We have not specified any particular orthonormality for the basis; the appropriate inverse metric G^{ij} is to be determined. The insertions $\mathscr{A}_i$ and $\mathscr{A}_j$ that replace the handle are implicitly at the origin in the z_1 and z_2 frames respectively.

Now, let us consider the case in which M_2 is a sphere, with the sewing coordinate z_2 being u and with $\ldots_2$ being some local operator $\mathscr{A}_k$ at $z = 0$. We have the identifications $z_1 u = 1$ from the plumbing fixture and $zu = 1$ from the sphere, so we can just extend $z_1 = z$ through the whole sewn region. The sewn surface M is the same as M_1 with an additional

operator $\mathscr{A}_k$ at $z_1 = 0$. Then eq. (9.4.1) becomes

$$\left\langle \ldots_1 \mathscr{A}_k^{(z_1)} \right\rangle_{M_1} = \sum_{ij} \left\langle \ldots_1 \mathscr{A}_i^{(z_1)} \right\rangle_{M_1} \mathsf{G}^{ij} \left\langle \mathscr{A}_j^{(u)} \mathscr{A}_k^{(z)} \right\rangle_{S_2} . \tag{9.4.2}$$

The two-point function $\langle \mathscr{A}_j^{(u)} \mathscr{A}_k^{(z)} \rangle_{S_2}$ is just the inner product $\langle\!\langle j|k \rangle = \mathscr{G}_{jk}$. Eq. (9.4.2) then implies

$$\mathsf{G}^{ij} \mathscr{G}_{jk} = \delta^i_k \tag{9.4.3}$$

or

$$\mathsf{G}^{ij} = \mathscr{G}^{ij} . \tag{9.4.4}$$

That is, the inner product defined by the two-point function on the sphere is the correct one for sewing. We have thus derived the relation (9.4.1) for the case in which M_2 is a sphere with an arbitrary local operator. By the state–operator isomorphism, any boundary state can be obtained in this way. Since the path integral is local, the result depends only on the region of the sewing and not on the rest of the surface, so it holds for arbitrary M_2 and $\ldots_2$.

Sewing with parameter q is the same as sewing with $q = 1$ and local coordinates z_1 and $z_2' = z_2/q$. We thus obtain eq. (9.4.1) with $\mathscr{A}_j$ in the primed frame, and so obtain an extra factor from the conformal transformation back to z_2,

$$\mathscr{A}_j^{(z_2')} = q^{h_j} \bar{q}^{\tilde{h}_j} \mathscr{A}_j^{(z_2)} . \tag{9.4.5}$$

The result is

$$\langle \ldots_1 \ldots_2 \rangle_M = \sum_{ij} q^{h_j} \bar{q}^{\tilde{h}_j} \left\langle \ldots_1 \mathscr{A}_i^{(z_1)} \right\rangle_{M_1} \mathscr{G}^{ij} \left\langle \mathscr{A}_j^{(z_2)} \ldots_2 \right\rangle_{M_2} , \tag{9.4.6}$$

where we have taken a basis diagonal in the weights. For the '8' operation the path integrals on M and M_0 are related in the same way:

$$\langle \ldots \rangle_M = \sum_{ij} q^{h_j} \bar{q}^{\tilde{h}_j} \mathscr{G}^{ij} \left\langle \ldots \mathscr{A}_i^{(z_1)} \mathscr{A}_j^{(z_2)} \right\rangle_{M_0} . \tag{9.4.7}$$

It is interesting to consider again the boundary of moduli space, given by the $q \to 0$ limit of one or more handles. We see that the operators of lowest weight will control the behavior. The results (9.4.6) and (9.4.7) resemble the OPE, and in fact (9.4.6) reduces to the OPE when M_2 is a sphere and '$\ldots_2$' is a pair of vertex operators. This analogy is especially clear from figure 9.4(c), representing the limit as the insertion of a small handle. The sewing relation shows that not only a product of operators, but also a small handle, or any other disturbance that we might sew into a small circle on the world-sheet, is equivalent to a sum of local operators. The result (9.4.6) is an elegant generalization of the OPE. From the point of view of M_1, there is a tiny copy of M_2 glued in near $z_1 = 0$. This can be expanded in local operators $\mathscr{A}_i$, with coefficient functions $\langle \mathscr{A}^i \ldots_2 \rangle_{M_2}$. From the point of view of M_2, there is a tiny copy of M_1 glued in near $z_2 = 0$, which is expanded in local operators $\mathscr{A}^i$, with coefficient functions $\langle \ldots_1 \mathscr{A}_i \rangle_{M_1}$.

We have argued that every Riemann surface can be reduced to thrice-punctured spheres. The sewing construction thus relates all amplitudes to the three-point amplitudes on the sphere, which are essentially the operator product coefficients: *these are the data that define the CFT.* We will encounter many equivalences between CFTs — we have already seen T-duality, and bosonization (chapter 10) will be another notable example. The sewing construction provides the criteria for two theories to be the same: a one-to-one correspondence between their states (or local operators) and identical OPEs.[1] In fact, as we will show in chapter 15, the operator products of the primary fields alone determine those of all the others.

Let us explore this point of view further. One can cut a surface along many different sets of curves, so there must be consistency conditions to guarantee that the sewing relation always gives the same result for the expectation values on M. Consider first an infinitesimal change in the cutting curve C. A general infinitesimal change leaving M invariant can be obtained by an infinitesimal change in the coordinates used for the sewing,

$$z_1' = z_1 + \sum_{n=-\infty}^{\infty} \epsilon_n z_1^{n+1} , \tag{9.4.8a}$$

$$z_2' = z_2 - \sum_{n=-\infty}^{\infty} \epsilon_{-n} q^{-n} z_2^{n+1} . \tag{9.4.8b}$$

The difference of the sewn expressions is then proportional to

$$\mathscr{A}_i^{(z_1')} \mathscr{G}^{ij} \mathscr{A}_j^{(z_2')} - \mathscr{A}_i^{(z_1)} \mathscr{G}^{ij} \mathscr{A}_j^{(z_2)} = - \sum_{n=-\infty}^{\infty} \epsilon_n \Big[L_n \cdot \mathscr{A}_i^{(z_1)} \mathscr{G}^{ij} \mathscr{A}_j^{(z_2)} - \mathscr{A}_i^{(z_1)} \mathscr{G}^{ij} L_{-n} \cdot \mathscr{A}_j^{(z_2)} \Big] + (\text{anti}) , \tag{9.4.9}$$

where we have used the conformal transformation (2.9.6), and 'anti' denotes the same form with $\epsilon_n \to \epsilon_n^*$ and $L_n \to \tilde{L}_n$. The vanishing of the quantity in braces can be shown by an argument parallel to that used to derive eq. (9.4.4). This is just the statement that one can pull the contour

[1] We should note that when there is more than one (2,0) plus (0,2) field that could play the role of T_{ab}, as in the free-field examples of section 2.5, it is necessary to specify which is the energy-momentum tensor. This choice determines the conformal transformations of operators, which enter implicitly into the sewing relation.

integral that defines the generator through the sewn handle; the change from L_n to L_{-n} is from an inversion of coordinates. All other contour integrals for conserved charges can be pulled through as well.

There are also sets of curves that are not related by a continuous deformation. The four-point amplitude on the sphere can be written in terms of three-point functions by cutting on curves circling different pairs of operators. Equality between different sewings is the associativity of the OPE, figure 2.8.

Another example is given by the torus: we can reduce the torus to the sphere by cutting on the A-cycle, or the B-cycle, or any other non-self-intersecting closed curve. In terms of the plane with identifications, a closed curve C that begins at $(\sigma^1, \sigma^2) = (0,0)$ must end at an identified point

$$2\pi(m,n)\,. \tag{9.4.10}$$

All curves with the same m and n values can be continuously deformed into one another, so we need to consider the relation between sewing on curves of different m and n. One can show that for a non-self-intersecting curve, m and n must be relatively prime, which implies that there are integers p and q such that $mp - nq = 1$. There is then a modular transformation

$$(\sigma^1, \sigma^2) = (\sigma'^1, \sigma'^2) \begin{bmatrix} m & n \\ q & p \end{bmatrix}\,, \tag{9.4.11}$$

such that the curve C comes back to

$$(\sigma'^1, \sigma'^2) = 2\pi(1,0)\,, \tag{9.4.12}$$

which is the A-cycle. So cutting on a general non-self-intersecting curve is equivalent to cutting on the A-cycle for a modular-equivalent value of τ. The consistency of the sewing construction is thus equivalent to modular invariance.

Every set of sewing curves may be transmuted into any other by a series of two moves: the four-point associativity transformation on the sphere and the modular transformation on the torus with a *single* vertex operator. Figure 9.4 makes this plausible. For example, figures 9.4(a) and 9.4(b) are related by one associativity condition. Repeated applications of associativity can bring the sewing to the form of figure 9.4(c), where all loops are expressed in terms of the one-loop, one-point function. So these are the conditions for an arbitrary set of operator product coefficients to define a sensible conformal field theory. The program of solving the duality constraints is known as the *conformal bootstrap*. It has been carried out only in cases where additional restrictions have been introduced, as we will discuss in chapter 15.

All the considerations of this section extend to world-sheets with boundary, where in addition to degenerating handles there may be degenerating strips. The consistency conditions have been generalized to this case.

9.5 General amplitudes

We now have assembled all the ingredients that we need in order to understand the consistency of general string amplitudes.

Having given a graphical decomposition of moduli space, the sewing formalism allows us to give a diagrammatic expression for the general string amplitude. Consider first the propagator, from the plumbing fixture. The ghost insertion for the modulus q is

$$\oint \frac{dz}{2\pi i} b_{z_1 z_1} \left.\frac{\partial z_1}{\partial q}\right|_{z_2} = \frac{b_0}{q}\,, \tag{9.5.1}$$

and similarly for $\bar{q}$ it is $\tilde{b}_0/\bar{q}$. Including these insertions, and using the sewing expression (9.4.6), the factors associated with the propagator are

$$\int_{|q|<1} \frac{d^2q}{q\bar{q}}\, q^{\alpha'(k_i^2+m_i^2)/4} \bar{q}^{\alpha'(k_i^2+\tilde{m}_i^2)/4} \mathscr{G}^{ij} b_0 \tilde{b}_0 = \frac{8\pi \delta_{m_i^2,\tilde{m}_i^2} \mathscr{G}^{ij} b_0 \tilde{b}_0}{\alpha'(k_i^2 + m_i^2 - i\epsilon)}\,. \tag{9.5.2}$$

The n-string coupling is given by inserting vertex operators into the surfaces appearing in the vertices of the sewing construction,

$$g(i_1, \ldots, i_n) = \sum_g \int_{\mathscr{V}_{g,n}} d^m t \left\langle \prod_{k=1}^{m} B_k \prod_{l=1}^{n} \hat{\mathscr{V}}_{i_l} \right\rangle_g\,. \tag{9.5.3}$$

where $m = 6g + 2n - 6$ is the number of moduli. The Feynman graphs constructed from the couplings (9.5.3) and propagators (9.5.2) then reproduce the full string amplitude. The label i again stands for the state and momentum of each string. The tree amplitude in section 9.1 illustrates this structure.

The first issue is the finiteness of the amplitude. We assert that divergences can arise only from the boundary of moduli space, as in the various low order examples. At finite values of the moduli, the sewing formula gives a sum over an infinite number of string states, but this sum converges because it is ultimately just the expansion of a quantum state in a complete set. Actually, convergence is guaranteed only in a unitary CFT, but the nonunitary parts of the string CFT are quite tame. The X^0 CFT is defined by analytic continuation from a unitary CFT, while the ghost CFT just generates a Faddeev–Popov determinant and so is finite if we have made a good choice of gauge. Now, the boundary has been made explicit as the $q \to 0$ limit of the various propagators, and the integral (9.5.2) is one that we have encountered several times before. It is

defined via analytic continuation, giving an ordinary field-theoretic propagator. The only singularities are then the poles in propagators, which represent production of on-shell particles in the intermediate states. In particular they are long-, not short-, distance divergences.

The other issue is unitarity. We studied this for the four-point tree amplitude in some detail. Having brought the general amplitude to a Feynman-diagram form, rather than work out the combinatorics in detail we will simply cite standard field theory results on the unitarity of the diagrammatic expansion. Also, the demonstration at the end of section 9.1 that the sum over intermediate states reduces to a sum over physical states generalizes as in field theory.

In summary, the issues involved in the perturbative finiteness and unitarity of string theory appear to be well understood. Observe once again that we have not assumed 26 flat dimensions, but merely d flat dimensions plus a compact positive norm CFT of central charge $26 - d$. Of course, any discussion of the unitarity of the flat spacetime S-matrix for the bosonic string is somewhat formal, since flat spacetime is not a solution of the theory — first there is the tachyon instability, then the cosmological constant induced by loops.

An alternative approach to the finiteness and unitarity of string theory, based on the light-cone gauge, has also been extensively developed. This gauge gives up covariance, but has certain advantages — it makes possible a rather explicit description of moduli space, and the amplitudes can be represented in a field-theoretic form with a Hamiltonian that is local in the light-cone time x^+. Covariance has then been established by showing that the perturbation theory is equivalent to the covariant perturbation theory.

String divergences

Before moving on, let us expand somewhat further this point that string divergences are long-distance effects. Consider the one-loop amplitude with n vertex operators, taking the limit that m operators come together with τ held fixed as in figure 9.8(a). The string loop of course includes gravity, and figure 9.8(a) is reminiscent of the graviton loop of figure 1.1(c), which was UV divergent when the interaction vertices came together in spacetime. There is indeed a divergence in this string amplitude, but it is from long spacetime distances, not short ones. Figure 9.8(b) shows a conformally equivalent picture where the m vertex operators are connected to the rest of the torus by a long cylinder. Cutting the cylinder open and inserting a complete set of states, we obtain the usual expression (9.5.2), with $q \to 0$ being the limit where the vertex operators become coincident. So the only singularities are the usual poles from propagation over long spacetime

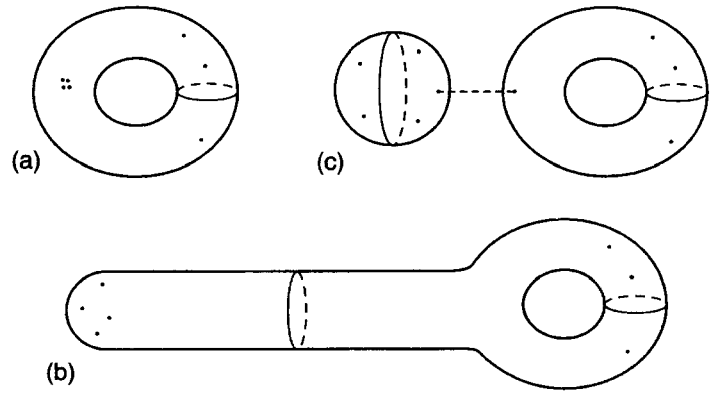


Fig. 9.8. (a) Torus with n vertex operators, in the limit that m of the operators come together. (b) Conformally equivalent picture: the m vertex operators at the end of a long cylinder. (c) Sewing construction, in terms of the product of a sphere with $m+1$ vertex operators and a torus with $n-m+1$.

distances. Conformal invariance once again changes short distances on the world-sheet into long distances in spacetime.

The cases $m = n$ and $m = n - 1$ present special problems. When $m = n$, meaning that all the vertex operators are coming together, momentum conservation forces k^μ to vanish. For massless states, the pole is then $1/0$ and the usual analytic continuation does not help because momentum conservation forces the amplitude to sit right on the massless pole. This is a tadpole, just as in section 7.4, and the resolution is the same, a correction to the background fields. The divergence here is proportional to the torus amplitude with one vertex operator.

In the other problematic case, $m = n - 1$, all the vertex operators but one, say $\mathscr{V}_n$, come together. Momentum conservation requires that the momentum flowing through the cylinder be $k^\mu = -k_n^\mu$. However, conformal invariance requires that $k_n^2 = -m_n^2$, so the intermediate momentum is fixed on top of a pole for those intermediate states with $m_j^2 = m_n^2$. Once again, there is an analogous divergence in field theory, shown in figure 9.9: the internal propagator is $1/0$. This divergence occurs whenever the mass of the internal line j is the same as that of the on-shell external particle n. The interpretation here is simple. The loop represents a radiative correction to the mass of the particle n. The S-matrix is defined with the external momentum taken on the *true* mass shell, including radiative corrections, and it is again free of this IR divergence. Note that the issue here is not some UV divergence internal to the loop, but the IR divergence on the

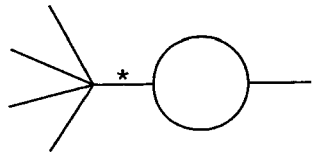


Fig. 9.9. Field-theory graph analogous to the case $m = n - 1$. The kinematics force the starred propagator to diverge. The interpretation is a (finite) mass shift due to the loop.

starred propagator which appears even when the mass correction is finite. The resolution in the string case is the same. The momentum appearing in the vertex operator must be at the true physical mass; this is not in general the same as the tree-level mass, the one that was always a multiple of $1/\alpha'$. Note that for certain massless fields — in particular gauge bosons — gauge invariance guarantees that they will remain massless to all orders. Incidentally, for massive states the mass shift is usually complex, because these states are unstable to decay into lighter string states.

There is a puzzle here. We saw early on that consistency of string theory, the condition that the world-sheet theory be conformally invariant, required the beta functions (3.7.14) to vanish. These beta functions contain no information about higher genus surfaces — they correspond to the tree-level string field equations. What happens when we try to expand around a loop-corrected background? In fact things work out just right. The dangerous limit in figure 9.8 gives rise not only to a divergence, but also to a conformal anomaly. To see this heuristically, picture the limit as in figure 9.3(c), as a small handle glued into the sphere. The small handle can be represented by a series of local operators, and in particular from the massless level there is

$$\partial X^\mu \bar{\partial} X_\mu \frac{dq d\bar{q}}{q\bar{q}} . \tag{9.5.4}$$

Now, cut off the divergence in a coordinate-invariant way by requiring the size of the handle to be greater than a; that is, $|q|e^\omega > a$, where the metric on the sphere is $dzd\bar{z}e^{2\omega}$. The lower limit on the integral (9.5.4) gives

$$-4\pi \partial X^\mu \bar{\partial} X_\mu \ln(ae^{-\omega}) . \tag{9.5.5}$$

Thus we have new dependence on the Weyl factor of the metric: this corrects the sigma-model Weyl anomaly $\beta^G_{\mu\nu}$, eq. (3.7.14). To obtain a Weyl-invariant theory, the background fields must satisfy the full *string-loop-corrected* Weyl invariance conditions (or, equivalently, BRST invariance). The Weyl anomaly from the shifted background cancels that from the small handle. This is the *Fischler–Susskind mechanism*. From figure 9.3(c),

we can think of the divergence and the Weyl anomaly as world-sheet UV effects, not from the usual short-wavelength fluctuations of the fields, but from a small topological fluctuation. We should emphasize that on a *fixed* Riemann surface of any genus, the local structure is unaffected by the topology; it is the integration over Riemann surfaces that produces the divergence and Weyl anomaly.

This can also be regarded as a breakdown of the canceled propagator argument. Because the kinematics forces the amplitude to be at the pole, we cannot appeal to analytic continuation to eliminate possible surface terms under a BRST variation. In fact they appear, corresponding to the Weyl anomaly (9.5.5). In the loop-corrected background, the BRST anomaly from the shift of the background cancels that from the boundary of moduli space — the stringy technicalities all fit together to give a consistent spacetime picture.

To implement this in string perturbation theory is rather clumsy, but fortunately it is rarely necessary. First, in most supersymmetric theories the backgrounds are uncorrected. Second, for the analysis of long-distance physics in string theory it is usually most efficient to proceed in two steps, first using string theory to derive the form of the low energy effective action and then using the effective field theory at long distance. One then does not need to deal with the IR divergences at the string level. In the present case, the tadpoles imply a potential energy term

$$\delta S = -\sum_\chi \Lambda_\chi \int d^d x (-G)^{1/2} e^{-\chi\Phi} . \tag{9.5.6}$$

in the effective action. Here Λ_χ is the tadpole from Euler number χ, and the dependence of the action on Φ follows from the fact that e^Φ is the string coupling. The separate tadpoles for the graviton and dilaton can be extracted from the string amplitude, with some effort.

The same idea extends to the vertex operator correction from the mass shift, and to other IR divergences such as those of exclusive amplitudes in theories with massless particles. These field-theory divergences must also appear in the corresponding string-theory amplitudes, since string theory reduces to field theory at long distance. Again we emphasize that while UV divergences in field theory generally signify a fundamental limitation on the range of validity of the theory, IR divergences usually mean that one is simply asking the wrong question (wrong background, wrong mass, exclusive versus inclusive cross-section, etc.) in what is otherwise a perfectly consistent theory. Any such divergences in string theory should then be dealt with precisely as in field theory.

Let us review what we have learned. Modulo problems particular to the bosonic string, the S-matrix is finite and unitary. The basic mechanism is modular invariance, which cuts off the UV divergences without spoiling

the spacetime gauge invariance. Of course, knowing that string theory makes sense in weak coupling perturbation theory around flat spacetime is hardly enough. We would like to know that the theory makes sense beyond perturbation theory, and have some means of calculating nonperturbative corrections. Also we would like to know what string theory implies for the nature of spacetime at short distance. In the remainder of this chapter we briefly describe some of the attempts that have been made in these directions.

9.6 String field theory

It is interesting to compare the approach we have taken in string theory to the way quantum field theory is formulated. The world-sheet path integral is analogous to a sum over particle paths, often called the *first-quantized* description. In quantum field theory, on the other hand, we start with a path integral (or its Hamiltonian equivalent) over spacetime fields. One could instead develop quantum field perturbation theory in the first-quantized form; it is necessary to introduce explicit interaction vertices, spacetime points where n particle paths end. In string theory this second step is not necessary — there are no special points, and the sum over topologies produces the interactions. However, the result is only the perturbation expansion, and there are important phenomena that are not contained in this expansion. In the Standard Model, these include quark confinement and the finite-temperature baryon number violation of the weak interaction. There does not seem to be any useful sense in which these are contained in a resummation of the perturbation series; rather they require a more general formulation of the theory.

In the previous section we have expressed string amplitudes in a propagator and vertex form that looks like the perturbation theory for an infinite-component field theory. Could we instead derive this from a path integral over an infinite set of spacetime fields? And is there some principle, extending gauge invariance and general coordinate invariance, that determines the form of the theory? This is the subject of *string field theory*. We will be able to sketch here only a few of the ideas, ones that are particularly close to the approach that we have been taking. The reader should consult the literature for more details.

We will start with the guess that we should take the string wavefunctional $\Psi[X]$ and elevate it to a field operator, or equivalently to a path integral variable of integration — so-called second quantization. Now, $\Psi[X]$ depends on the curve $X^\mu(\sigma)$ traced out by the string in spacetime, and roughly speaking creates or destroys a string in this configuration. However, we need to be more precise. For example, should we require Ψ

to be invariant under reparameterization of the curve, and should it also depend on the Polyakov metric g_{ab}? Things seem to work much more nicely if one starts with the BRST-invariant form of the theory, including the Faddeev–Popov ghosts, and then second quantizes. From the commutation relations, b and c are conjugate to one another (as are $\tilde{b}$ and $\tilde{c}$), so if for example we treat c and $\tilde{c}$ as coordinates the wavefunctional is $\Psi[X, c, \tilde{c}]$. We then have a path integral running over all such functionals,

$$\int [d\Psi] \exp(iS[\Psi]) \,. \tag{9.6.1}$$

We will consider first the open string, describing one particular approach.

What symmetry principle will restrict the string action? From our study of the string vertex operators and null states, we might guess the symmetry

$$\delta\Psi = Q_B \Lambda \tag{9.6.2}$$

for any functional Λ. There is an extremely simple action with this invariance,

$$S_0 = \frac{1}{2}\langle\!\langle \Psi | Q_B | \Psi \rangle\!\rangle \,. \tag{9.6.3}$$

The equation of motion is just

$$Q_B \Psi = 0 \,. \tag{9.6.4}$$

One way to see that this is reasonable is by expanding the string field Ψ in terms of an infinite number of ordinary fields, one for each internal state of the string. Let $\Psi_i[X', c, \tilde{c}]$ be a complete set of wavefunctions for the internal modes — all but the center-of-mass x^μ. Then a general state can be expanded:

$$\Psi[X, c, \tilde{c}] = \sum_i \Phi_i(x) \Psi_i[X', c, \tilde{c}] \,, \tag{9.6.5}$$

the expansion coefficients being functions of the remaining variables x^μ. The string field path integral becomes an infinite product of path integrals over the component functions,

$$\int [d\Psi] \;\to\; \prod_i \int [d\Phi_i] \,. \tag{9.6.6}$$

As usual, we can start with the ground state functional Ψ_0 and build the remaining states with the raising operators. For example, keeping all

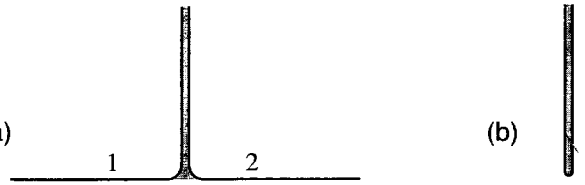


Fig. 9.10. (a) $\Psi_1 * \Psi_2$, convoluting the right half of Ψ_1 with the left half of Ψ_2. One can think of this as defined by a path integral over the shaded region, whose width is then taken to zero. (b) $\int \Psi$, convoluting the right and left halves of Ψ.

states at the first two levels,

$$\Psi = \Big[\varphi(x) + A_\mu(x)\alpha^\mu_{-1} + B(x)b_{-1} + C(x)c_{-1} + \varphi'(x)c_0 + A'_\mu(x)\alpha^\mu_{-1}c_0 + B'(x)b_{-1}c_0 + C'(x)c_{-1}c_0 + \ldots\Big]\Psi_0 \,. \quad (9.6.7)$$

Here, φ is the tachyon field and A_μ the gauge field, while B and C are the Faddeev–Popov ghost fields associated with the spacetime gauge symmetry. The primed fields can all be gauged away or are auxiliary (that is, they satisfy algebraic rather than differential equations and can be eliminated). One can then check that the symmetry (9.6.2) includes the ordinary gauge transformation on A_μ, while the action is of Klein–Gordon form for φ and of Maxwell form for A_μ. The BRST string field formalism thus gives a very compact generalization of spacetime gauge invariance.

Interactions also work out rather nicely. Define a product operation $\Psi_1 * \Psi_2$ on string fields by the functional integral depicted in figure 9.10(a), forming the overlap of the right half of Ψ_1 with the left half of Ψ_2. Define also an operation $\int$, giving the overlap of the right and left halves of the single string wavefunctional as shown in figure 9.10(b). To make an interacting theory, add a nonlinear term to the gauge transformation,

$$\delta\Psi = Q_B\Lambda + g\Psi * \Lambda - g\Lambda * \Psi \,. \quad (9.6.8)$$

The $*$ product is associative, and one can show by a contour argument that Q_B acts as a derivation,

$$Q_B(\Psi_1 * \Psi_2) = (Q_B\Psi_1) * \Psi_2 + \Psi_1 * (Q_B\Psi_2) \,. \quad (9.6.9)$$

From these it follows that the symmetry (9.6.8) closes. Using the cyclic property,

$$\int \Psi_1 * \Psi_2 = \int \Psi_2 * \Psi_1 \,, \quad (9.6.10)$$

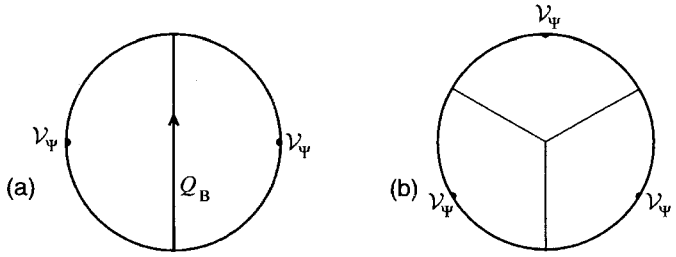


Fig. 9.11. CFT representation of the string field action: (a) the kinetic term; (b) the interaction. A $z^{2/3}$ conformal transformation is used to turn the usual semi-circle around each vertex operator into a wedge.

we can write the invariant action

$$S = \frac{1}{2}\int \Psi * Q_B\Psi + \frac{2g}{3}\int \Psi * \Psi * \Psi \,. \quad (9.6.11)$$

This is also usefully written in terms of CFT, replacing Ψ with a half-disk with corresponding vertex operator $\mathcal{V}_\Psi$,

$$S = \frac{1}{2}\langle \mathcal{V}_\Psi Q_B \cdot \mathcal{V}_\Psi \rangle_{D_2} + \frac{2g}{3}\langle \mathcal{V}_\Psi \mathcal{V}_\Psi \mathcal{V}_\Psi \rangle_{D_2} \,. \quad (9.6.12)$$

The vertex operator positions and conformal frames are not explicitly denoted in eq. (9.6.12) but are as shown in figure 9.11. A $z^{2/3}$ mapping is used to fit the three vertex operator half-disks into the disk. This action passes an important check. The path integral on the disk is nonzero only for insertions of total $Q^g = 3$. The vertex operators for the physical states all have ghost number 1, so both terms in the action (9.6.12) are nonvanishing. Since the action (9.6.11) is invariant, we might expect that for the gauge field it reduces to the non-Abelian Yang–Mills action, and this is the case.

The Siegel gauge is $b_0\Psi = 0$. In the kinetic action, only the term c_0L_0 from Q_B survives. Since in components $L_0 \propto p^2 + m^2$, this is analogous to the familiar covariant gauge for electrodynamics. The propagator

$$\int_0^\infty dt\, \exp(-tL_0) = L_0^{-1} \quad (9.6.13)$$

is simply the integral over strips of any length. The vertex is figure 9.11(b) with the wedges removed, joining the ends of three strips in a 'Y.' This has been shown to be equivalent to open string theory described as a sum over Riemann surfaces, including in particular a complete single cover of moduli space. On the one hand this is not surprising: the string field theory has the spacetime gauge invariance (9.6.8), so that null states will

decouple from the S-matrix. We have argued that the decoupling of null states determines the form of the string amplitudes including the need for a complete cover of moduli space (else total derivative terms will give null state couplings). On the other hand moduli space is complicated and few explicit descriptions are known. The one produced by this string field theory, in terms of a sum over Feynman graphs and an integral over strip lengths for each, was discovered by mathematicians only shortly before its construction through string field theory.[2]

Let us point out a strange feature of this field theory. We have stated that the string field theory covers the entire moduli space, but we know from the annulus that this includes limits that produce *closed string* poles, even though there are no explicit closed string fields. The closed string poles come from strips going to zero, rather than infinite, length.[3] The precise statement is that the open string field path integral (9.6.1) generates the amplitudes of open *plus* closed string theory (except that surfaces with no boundaries at all, to which open strings cannot couple, are not generated). In this respect it is as though closed strings were in some imprecise sense 'bound states' of open strings.

For a theory with only closed strings we can try to repeat the above steps with a closed string field, though it is somewhat more complicated. Already for the free action (9.6.3) the ghost number does not work out: we need an extra c ghost to make the matrix element between physical states nonvanishing. There is a remedy, which is to require the string field Ψ and the gauge transformation Λ to be annihilated by $b_0 - \tilde{b}_0$ and by $L_0 - \tilde{L}_0$. The action

$$S_0' = \frac{1}{2}\langle\!\langle \Psi | (c_0 - \tilde{c}_0) Q_B | \Psi \rangle\!\rangle \tag{9.6.14}$$

is then invariant under $\delta\Psi = Q_B\Lambda$. Going on to the interacting theory there is no associative product analogous to $*$, and correspondingly no simple cover of moduli space. As in the previous section we need to add in higher vertices $\mathscr{V}_{g,n}$, with more powers of the field and also with internal handles, in order to cover moduli space. The resulting action, like the open string action, can be obtained from a symmetry principle using what is known as the *Batalin–Vilkovisky formalism*. Gauge fixing then leads to the perturbation theory described in the previous section.

[2] Another physically motivated description of moduli space comes from the light-cone gauge, where the moduli are the interaction times and the momenta p^+ of the intermediate states, and also a twist angle for each intermediate closed string.

[3] For this reason, this form of string field theory would not be suitable for our purposes of the previous section. We would need to add stubs to the vertex to eliminate the unexpected divergences, and then to add in closed string propagators and vertices to cover the omitted region of moduli space.

The derivation of string theory from spacetime gauge symmetry is extremely elegant, particularly for the open string. The inclusion of closed strings in the open string field theory is intriguing. However, it remains to be seen whether string field theory is a useful tool beyond perturbation theory. It has not played a role in the recent nonperturbative progress based on spacetime supersymmetry, which will be described in chapter 14. It is quite possible that strings, or string fields, are simply not the correct degrees of freedom for a nonperturbative formulation of 'string' theory.

9.7 Large order behavior

There is an interesting result about the large order behavior of string perturbation theory. Let us start with field theory. Consider the integral

$$\int_{-\infty}^{\infty} dy \exp\left[-\frac{y^2}{2!} - \lambda\frac{y^3}{3!}\right] = \lambda^{-1}\int_{-\infty}^{\infty} dz \exp\left[-\frac{1}{\lambda^2}\left(\frac{z^2}{2!} + \frac{z^3}{3!}\right)\right], \tag{9.7.1}$$

where $z = y\lambda$. This integral diverges for $\lambda \neq 0$, but for now we can ignore that and calculate it as a formal expansion in powers of λ,

$$\sum_{n=0}^{\infty} \frac{(-\lambda)^n}{6^n n!}\int_{-\infty}^{\infty} dy\, y^{3n}\exp(-y^2/2) = (2\pi)^{1/2}\sum_{k=0}^{\infty}\lambda^{2k}C_{2k}\,, \tag{9.7.2}$$

with

$$C_{2k} = \frac{2^k\,\Gamma(3k+1/2)}{\pi^{1/2}\,3^{2k}(2k)!}\,. \tag{9.7.3}$$

This integral is the zero-dimensional limit of a scalar field theory, with a mass term and cubic coupling. The expansion in λ can be represented in terms of Feynman graphs, with propagator 1 and cubic vertex $-\lambda$. The constant C_{2k} then simply counts the total number of vacuum graphs with $2k$ cubic vertices, or equivalently $k+1$ loops, weighted by the appropriate symmetry factor. For example, the reader can check that there are two vacuum graphs with two vertices, with symmetry groups of order 8 and 12, and indeed $C_2 = \frac{5}{24} = \frac{1}{8} + \frac{1}{12}$. To get the large order behavior, we use Stirling's approximation $n! \approx n^{n+1/2}e^{-n}$. Then

$$C_{2k} \approx k^k \approx k!\,, \tag{9.7.4}$$

where the approximation holds up to factors of the form k^a and b^k.

One can also include a quartic term $\lambda^2 y^4/4!$, which makes the integral converge. The power of λ is determined by requiring that all terms in the action for z be proportional to λ^{-2}; equivalently, all $k+1$ loop vacuum graphs are weighted by λ^{2k}. This additional term does not substantially change the result; the estimate (9.7.4) still holds to the given accuracy.

When considering field theory, the propagator and vertices become more complicated. However, the dominant behavior at large orders usually comes from simple counting of graphs — the propagators and vertices only contribute factors of order b^k, unless there is unusual IR or UV behavior. The series is of the form

$$\sum_{k=0}^{\infty} \lambda^{2k} k! f_{2k} \,, \tag{9.7.5}$$

where $f_{2k} \approx k^a b^k$ varies more slowly than $k!$. This diverges unless f_{2k} has unexpected cancellations. For small λ the terms become smaller for a while and then grow as the $k!$ dominates. The smallest terms in the sum are of order

$$\exp[-O(1/\lambda^2)] \,. \tag{9.7.6}$$

This is then an indication of the accuracy of perturbation theory as an asymptotic series. Indeed, effects like instantons and dynamical symmetry breaking, which vanish to all orders of perturbation theory, are typically of this order. For example, the large order behavior of the series (9.7.2) is related to the 'instanton' (stationary point) at $z = -2$, with action $2/3\lambda^2$.

In string theory, the Feynman graph decomposition of moduli space makes it possible to establish the same bounds. It has been shown explicitly that the series is of the form (9.7.5) with f_{2k} positive and bounded below by b^k. Thus the series is divergent, and is not even Borel summable. Borel summation rewrites the series (9.7.5) as the Laplace transform of a more convergent series

$$\int_0^\infty dt \, \exp(-t) \sum_{k=0}^{\infty} (t\lambda^2)^k f_{2k} \,. \tag{9.7.7}$$

However, the positivity of the f_{2k} implies that this new sum, even if it converges, has poles on the positive t-axis, so that the integral is ambiguous. Thus, we should expect that string theories will have interesting nonperturbative phenomena. This is not surprising because even in the low energy limit they reduce to nontrivial field theories.

The estimate (9.7.4) for closed string theory is a lower bound, which can be obtained even keeping only the tree-level cubic vertex $\mathscr{V}_{0,3}$. Because the closed string vertex receives corrections from Riemann surfaces of all genus, one might wonder whether perturbation theory grows even more rapidly. One way to see that in fact it does is to recall that the open string field theory contains all open plus closed string Riemann surfaces with at least one boundary. Let us consider the surfaces with one boundary, in the region where the boundary has shrunk to a point. This limit of open string moduli space is equivalent to the moduli space for *closed* surfaces

with one vertex operator. One can show that this region is not vastly smaller than the full open string moduli space, so the same estimate holds, the perturbation series going as

$$\sum_{k=0}^{\infty} g_o^{2k} O(k!) = \sum_{k=0}^{\infty} g_c^k O(k!) \,. \tag{9.7.8}$$

The genus-g term, proportional to g_c^{2g}, has a coefficient of order $(2g)!$. This is larger than the $g!$ characteristic of field theory, which was also the lower bound obtained from $\mathscr{V}_{0,3}$ alone. Evidently the largest part of the large order amplitudes is contained within the vertices $\mathscr{V}_{g,n}$ and comes from the interior of moduli space. One then expects nonperturbative effects of order

$$\exp[-O(1/g_c)] \,, \tag{9.7.9}$$

much larger at weak coupling than the $\exp[-O(1/g_c^2)]$ characteristic of field theory. This large order behavior was first observed in some simplified models to be discussed in section 9.9. As we will see in chapter 13, these effects are now known to be related to D-branes, at least in some supersymmetric string theories.

9.8 High energy and high temperature

In order to obtain a better understanding of string dynamics, one might look at the behavior in different limits. In this section we consider various high energy limits of string theory.

Hard scattering

Consider again the hard scattering behavior, which we obtained from the Veneziano and Virasoro–Shapiro amplitudes by Stirling's approximation. Go back to the integral form (6.6.5),

$$\frac{8\pi i g_c^2}{\alpha'} (2\pi)^{26} \delta^{26}(\textstyle\sum_i k_i) \int_C d^2 z_4 \, |z_4|^{-\alpha' u/2 - 4} |1 - z_4|^{-\alpha' t/2 - 4} \,. \tag{9.8.1}$$

In the hard scattering region $s \to \infty$ with t/s fixed, the exponents are large and the integral is dominated by the stationary points of the exponent.[4] The saddle point, for $s, t, u \gg 1$, is given by

$$\frac{\partial}{\partial z_4} \left(u \ln |z_4|^2 + t \ln |1 - z_4|^2 \right) = \frac{u}{z_4} + \frac{t}{z_4 - 1} = 0 \,, \tag{9.8.2}$$

[4] As usual, the integral diverges at $z_4 \to \infty$ and is defined by analytic continuation. The saddle point approximation can be justified for the continuation.

with solution $z_4 = -u/s$; recall that $s + t + u = O(1)$. The saddle point approximation then gives an amplitude

$$S \approx \exp[-\alpha'(s \ln s + t \ln t + u \ln u)/2] , \tag{9.8.3}$$

as given in eq. (6.6.13).

Let us also back up one step, from the integral over z_4 to the path integral over X^μ. This is dominated by the saddle point

$$X^\mu_{\rm cl}(\sigma) = i \sum_i k_i^\mu G'(\sigma, \sigma_i) , \tag{9.8.4}$$

where the σ_i are the vertex operator positions. Recall that the Green's function is proportional to α'. At high energy the size of the scattering region grows linearly with energy, $\delta X \approx \alpha' k$. In field theory, higher energy hard scattering probes shorter distances, $\delta X \approx 1/k$ by the uncertainty principle. In string theory this is not the case beyond the string scale. It appears to be impossible in string perturbation theory to probe distances below this scale, giving rise to the idea that it is a limiting distance in spacetime. The effective lower limit on compactification $R \approx \alpha'^{1/2}$ also supports this. However, we will see in chapter 13 that recent advances in nonperturbative string theory reveal the existence of shorter length scales.

At general orders in the string loop expansion, the dominant term in the amplitude will be

$$\exp\left[-\sum_{i<j} k_i \cdot k_j G'(\sigma_i, \sigma_j)\right] . \tag{9.8.5}$$

We need to find the saddle points of the exponent, with respect to both the vertex operator positions and the moduli of the Riemann surface. This is like an electrostatics problem: the exponent is the energy of n point charges k_i^μ on a Riemann surface, where μ labels d different electrostatic fields and $\mu = 0$ has the opposite sign in the energy. The background charge density vanishes by momentum conservation, and the renormalized self-energy is negligible because k_i^2 remains on-shell and small. Remarkably, one can find what is probably the dominant saddle point at each genus. Consider the Riemann surface

$$z_2^N = \frac{(z_1 - a_1)(z_1 - a_2)}{(z_1 - a_3)(z_1 - a_4)} . \tag{9.8.6}$$

Like the hypergeometric curve (9.2.9), it is a subspace of $S_2 \times S_2$. It can be regarded as N sheets glued together at branch cuts, and has genus $N - 1$. For the four-point amplitude, if the charges are placed at the branch cuts a_i then the amplitude is stationary with respect to their positions — the surface is 'vertical' at these points, $\partial z_2/\partial z_1 = \infty$, and the symmetry $z_2 \to \exp(2\pi i/N) z_2$ requires the vertical component of the gradient to vanish. It remains to extremize with respect to the positions a_i, but this is easy: the electric field is the same as on the sphere with charges at $z = a_i$, except that it is divided between N sheets. Thus the energy is $N \cdot N^{-2} = N^{-1}$ times that on the sphere. Three of the positions can be fixed by Möbius invariance, for example to $0, 1, \infty$, and the fourth is given by the same condition (9.8.2) as on the sphere. Thus at genus g,

$$S_g \approx \exp[-\alpha'(s \ln s + t \ln t + u \ln u)/2(g + 1)] . \tag{9.8.7}$$

With additional effort one can include the Gaussian fluctuations around the saddle, which determine the coefficient of S_g. For the N-sheeted saddle points, $X^\mu_{\rm cl}$ is simply divided by $N = g + 1$.

There is a check that this result makes sense. In the small-angle region $-t/s \ll 1$ it becomes

$$\exp[\alpha' t \ln s/2(g + 1)] . \tag{9.8.8}$$

Now, $t^{1/2}$ is the total momentum transferred from the incoming particle 1 to the final particle 3. At small angles it is plausible that the genus-g scattering process can be decomposed into $g + 1$ successive scatterings. Each separate scattering has the same center of mass s, while the t_i are constrained so that the total momentum transfer is $t^{1/2}$. The individual scatterings go as $\exp(\alpha' t_i \ln s/2)$. It is then easy to check that it is most efficient to take the momentum transfers to be equal, $t_i^{1/2} = t^{1/2}/(g + 1)$, leading to the result (9.8.8) for the product of the amplitudes. This is different from particle scattering, where it is most efficient to transfer momentum in a small number of hard scatterings.

For excited string states, the exponential part of the vertex operator is the same as for the tachyon. The same saddle point then dominates, with $X^\mu_{\rm cl}$ inserted in the vertex operators. Since this same saddle point dominates for all states and, up to the factor $g + 1$, all orders, there are linear relations between the amplitudes for different states that hold for all genus and therefore plausibly for the exact amplitudes. In quantum field theory, high energy scattering often reveals symmetries that are not obvious at low energy. It has been suggested that the relations that emerge here are of the same sort.

It is not clear to what extent it makes sense to sum the series (9.8.7). It is interesting to contrast the result here with that in the previous section. In both cases it is the interior of moduli space that is producing the interesting physics (in contrast to the field-theoretic behavior found near the boundary). However, the large order behavior came from the whole of moduli space, dominated in particular by its large volume. The high energy behavior at fixed genus is dominated by a single point.

Regge scattering

The analysis of the Regge limit can also be carried to higher orders. We will not do this, but will develop an interesting interpretation of the tree-level Regge physics.

Regge scattering is less familiar to most physicists than hard scattering. It is a gentle probe of a highly boosted system. When a system is boosted, time dilation slows down its internal processes. One might then expect to see more of the virtual degrees of freedom as one goes to higher energies.

Let us in particular consider the quantum fluctuations of a single string, in particular the root mean size of the ground state. The mode expansion gives

$$\begin{aligned}\langle 0|[X^1(\sigma)-x^1]^2|0\rangle &= \sum_{n=1}^{\infty}\frac{\alpha'}{n^2}\langle 0|(\alpha_n\alpha_{-n}+\tilde{\alpha}_n\tilde{\alpha}_{-n})|0\rangle \\ &= \alpha'\sum_{n=1}^{\infty}\frac{1}{n}\,. \end{aligned} \tag{9.8.9}$$

This is divergent but has no direct physical significance. A measurement on a time-scale $\delta\tau$ is sensitive only up to modes of frequency less than $\delta\tau^{-1}$, essentially $n\alpha'^{-1/2} < \delta\tau$. The logarithmic divergence becomes $\alpha'\ln(\alpha'^{1/2}/\delta\tau)$, and the size is the square root of this.

This growth in the effective size is directly visible in the Regge amplitude (6.6.12) at small t,

$$s^{2+\alpha' t/2}\frac{\Gamma(-\alpha' t/4-1)}{\Gamma(\alpha' t/4+2)} \approx \frac{s^2}{t}\exp\left[-\frac{q^2\alpha'}{2}\ln(\alpha' s)\right], \tag{9.8.10}$$

where $q^2 = -t$ is the momentum transfer. This is the gravitational amplitude s^2/t modified by a form-factor whose Fourier transform corresponds to an object of size $(\alpha'\ln\alpha' s)^{1/2}$. From the point of view of one string the other has a boost of order $\alpha' s$, so the time resolution of its internal processes is inverse to this, in agreement with the earlier estimate for the size.

The square root of a logarithm is a very slow function and ordinarily would be of little importance, but it has been argued that this behavior is very significant for strings near black holes. In chapter 14 we will revisit, from a different point of view, the idea that strings reveal their true nature when very highly boosted.

High temperature

A similar limit with interesting stringy behavior is high temperature. From the generic behavior (7.2.30) of the partition function at high weight h,

and the relation $m^2 = 4h/\alpha'$, the density of string states $n(m)$ satisfies

$$\int_0^{\infty} dm\, n(m)\exp(-\alpha'\pi m^2\ell) \approx \exp(-4\pi/\ell)\,, \tag{9.8.11}$$

implying $n(m) \approx \exp(4\pi m\alpha'^{1/2})$. The thermal partition function of a single string is then

$$\int_0^{\infty} dm\, \exp(4\pi m\alpha'^{1/2})\exp(-m/T)\,. \tag{9.8.12}$$

Because the density of states grows exponentially, the partition function (9.8.12) diverges for temperatures greater than the *Hagedorn temperature*

$$T_{\rm H} = \frac{1}{4\pi\alpha'^{1/2}}\,. \tag{9.8.13}$$

One can understand the Hagedorn divergence as follows. The energy of a long string is proportional to its length. So also is its entropy, being extensive. At low temperature the energy dominates, but at high temperature the entropy dominates and it is favorable to produce strings of arbitrary length. In an ensemble of such strings the density will become large, and the interactions will be important near the transition.

This divergence suggests a phase transition. Indeed, in the strong interaction there is a similar growth in the density of states of highly excited hadrons, and there is a phase transition from a low temperature hadron phase to a high temperature quark–gluon phase. It is interesting to wonder whether the Hagedorn divergence indicates some analogy to the quark–gluon structure in the fundamental string, but there is no clear answer.

To conclude, let us discuss the partition function of free strings in somewhat more detail. The thermal partition function of a bosonic field of mass m in dimension d is

$$\begin{aligned}F(T,m^2) &= T\int\frac{d^{d-1}\mathbf{k}}{(2\pi)^{d-1}}\ln[1-\exp(-\omega_k/T)] \\ &= -\int_0^{\infty}\frac{dt}{t}(2\pi t)^{-d/2}\sum_{r=1}^{\infty}\exp\left(-\frac{m^2 t}{2}-\frac{r^2}{2T^2 t}\right). \end{aligned} \tag{9.8.14}$$

In the second form, one can readily sum over the string spectrum. Following the derivation of the vacuum amplitudes (7.3.12) and (7.3.6), the free energy density in 26 flat dimensions is

$$F(T) = -\int_R\frac{d\tau d\bar{\tau}}{4\tau_2}(4\pi^2\alpha'\tau_2)^{-13}|\eta(\tau)|^{-48}\sum_{r=1}^{\infty}\exp\left(-\frac{r^2}{4\pi T^2\alpha'\tau_2}\right). \tag{9.8.15}$$

The integral runs over the full region

$$R: \quad \tau_1 > 0\,, \quad |\tau_2| \leq \frac{1}{2}\,. \tag{9.8.16}$$

The derivation from a modular-invariant path integral is given as exercise 9.7. There is a possible divergence as $\tau_2 \to 0$; the integrand is largest when the real part of τ is zero, so let $\tau = i\tau_2$. From the modular transformation (7.2.44), the asymptotic behavior of the eta function in this limit is

$$\begin{aligned}\eta(i\tau_2) &= \eta(i/\tau_2)\tau_2^{-1/2} \\ &\approx \exp(\pi/12\tau_2) \text{ as } \tau_2 \to 0\,.\end{aligned} \tag{9.8.17}$$

The τ integral thus converges for $T < T_H$ and diverges for $T > T_H$.

The following curious observation has been made. In field theory the partition function is given by a Euclidean path integral, with time having period $1/T$. The same is true in string theory (exercise 9.7). T-duality should then relate the high and low temperature behaviors, and in fact the Hagedorn temperature is just the self-dual point. Define

$$\tilde{F}(T) = F(T) + \rho_0\,, \tag{9.8.18}$$

the thermal part (9.8.15) of the free energy plus the vacuum part (7.3.6). The logarithm of the path integral is $-\tilde{F}(T)/T$, so T-duality implies

$$\frac{1}{T}\tilde{F}(T) = \frac{T}{T_H^2}\tilde{F}(T_H^2/T)\,. \tag{9.8.19}$$

This relates the partition function below the Hagedorn temperature to that above it. The Hagedorn divergence for $T > T_H$ is a reflection of the tachyon divergence in the low temperature theory, but a similar formal result holds for the supersymmetric string, which has no tachyonic divergence. If we take this relation seriously, we obtain

$$\tilde{F}(T) \to \frac{T^2}{T_H^2}\rho_0 \tag{9.8.20}$$

as $T \to \infty$. If this is a measure of the number of high energy degrees of freedom, they are very few. A single quantum field in d spacetime dimensions has a free energy of order T^d, so it is as though the high energy limit of string theory looks like a field theory in two *spacetime* dimensions. This is entirely unlike the exponentially growing spectrum below the Hagedorn transition, and may be a clue to the nature of the phase transition.

9.9 Low dimensions and noncritical strings

Field theories often simplify as the number of spacetime dimensions is reduced, and these low-dimensional theories have been useful models for understanding dynamics. We might try to do the same in string theory, but string theory requires a critical dimension. We could compactify, but this does not reduce the number of degrees of freedom: the string still oscillates in the compact directions. We have, however, encountered a consistent background in which the string actually does move in fewer dimensions. This is the linear dilaton background (3.7.18),

$$G_{\mu\nu}(x) = \eta_{\mu\nu}\,, \quad B_{\mu\nu}(x) = 0\,, \quad \Phi(x) = V_\mu x^\mu\,, \tag{9.9.1a}$$

$$V_\mu = \delta_\mu^{\,1}\left(\frac{26-D}{6\alpha'}\right)^{1/2}\,. \tag{9.9.1b}$$

For $D < 26$ the dilaton gradient is spacelike and we have taken it to lie in the 1-direction.

The effective string coupling e^Φ is a function of x^1, diverging at large x^1. This makes it hard to use string perturbation theory, but fortunately there is a slightly different background for which perturbation theory is good. Inserting the background (9.9.1) into the tachyon action (6.6.16) leads to the linearized tachyon field equation,

$$-\partial^\mu\partial_\mu T(x) + 2V^\mu\partial_\mu T(x) - \frac{4}{\alpha'}T(x) = 0\,. \tag{9.9.2}$$

This is the condition for the exponential vertex operator to be Weyl-invariant, with the linear dilaton energy-momentum tensor (2.5.1). The solutions are

$$T(x) = \exp(q\cdot x)\,, \quad (q-V)^2 = \frac{2-D}{6\alpha'}\,. \tag{9.9.3}$$

If we look for solutions which, like the dilaton, depend only on x^1, then

$$q_1 = \alpha_\pm = \left(\frac{26-D}{6\alpha'}\right)^{1/2} \pm \left(\frac{2-D}{6\alpha'}\right)^{1/2}\,. \tag{9.9.4}$$

For $D > 2$ this is complex, so the tachyon oscillates, but for $D \leq 2$ one gets a real exponential (times a linear function at the critical value $D = 2$). Of course, the tachyon field equation will have nonlinear corrections. It can be shown that this does not affect the qualitative form of the background, except that for rather subtle reasons only the solution α_- leads to a nonsingular background.

For $D \leq 2$ the term 'tachyon' for the string ground state is actually a misnomer, because this field is stable (exercise 9.8). Also for $D \leq 2$, the real exponential $T(x) = T_0\exp(\alpha_- x^1)$ prevents propagation to the region

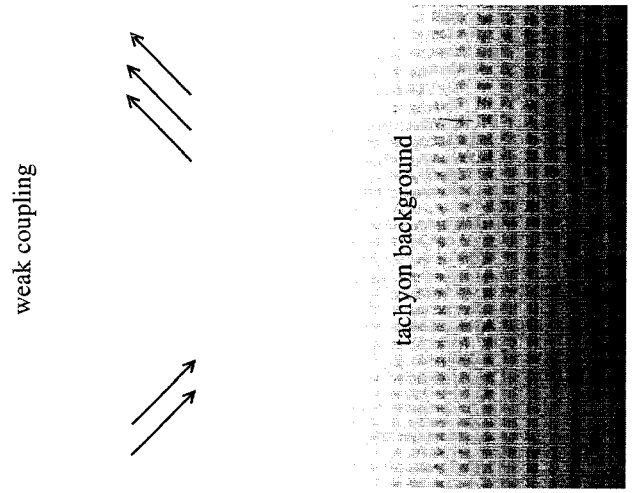


Fig. 9.12. Spacetime with dilaton and tachyon backgrounds. The interaction strength grows toward the right but is cut off by an effective potential from the tachyon background.

of large positive x^1 where the coupling diverges. To see this, consider the world-sheet action

$$S_\sigma = \frac{1}{4\pi\alpha'}\int_M d^2\sigma\, g^{1/2}\left[g^{ab}\eta_{\mu\nu}\partial_a X^\mu \partial_b X^\nu + \alpha' R V_1 x^1 + T_0 \exp(\alpha_- x^1)\right] . \tag{9.9.5}$$

At large x^1 the exponential tachyon dominates the linear dilaton and the path integral is suppressed, corresponding to an effective potential for the string due to the tachyon background. Physically one has an asymptotic region $x^1 \to -\infty$, where the coupling goes to zero. Strings propagating toward positive x^1 interact with one another and reflect off the potential and back to the asymptotic region. The maximum effective coupling is set by the energy and the strength T_0 of the potential. This picture, shown in figure 9.12, is a bit exotic, but in the interest of finding a solvable model it is worth pursuing. A model with rather similar physics, the static model, played a role in unraveling the nature of quantum field theory.

For $D > 2$ the physical effect of the tachyon background is less clear, though it has been argued that it still acts as a barrier. However, the focus has been on $D \leq 2$ for another reason, namely that these theories are actually solvable. Let us first observe that when we quantize the string, the gauge symmetries take away two families of oscillators, so that in $D \leq 2$ none will be left — there is only the center-of-mass motion, a single

scalar.[5] Nevertheless, to solve the string theory directly is still quite hard. First one has to evaluate the world-sheet path integral with action (9.9.5). This is no longer a free theory, because of the exponential tachyon field. It is known as the *Liouville field theory*. It is classically integrable, and with substantial effort some quantum amplitudes have been determined. However, to solve the string theory one would still need to integrate the result at each genus over the moduli space of Riemann surfaces; we have seen that this is a complicated space. Remarkably, the full amplitude has been determined, to all orders, by very different means. To describe this, we must first introduce another idea.

Noncritical strings

Consider a statistical mechanical system defined by a sum over two-dimensional surfaces in a d-dimensional Euclidean space. Let us also suppose that the surfaces have some internal degree of freedom that can be interpreted as a measure of length. The natural fields to describe the system are the embedding $X^\mu(\sigma)$ and a metric $g_{ab}(\sigma)$. However, there is no particular reason for the length to drop out of the problem — that is, for the theory to be Weyl-invariant. This is therefore called the *noncritical string*. The lowest-dimensional action (or Boltzmann weight, since we are thinking of the path integral in the context of statistical mechanics) for this system would then be the Polyakov action with an additional world-sheet cosmological constant term, which in the critical string is forbidden by Weyl invariance,

$$S = \frac{1}{4\pi\alpha'}\int_M d^2\sigma\, g^{1/2}\left(g^{ab}\partial_a X^\mu \partial_b X_\mu + \mu\right) . \tag{9.9.6}$$

Without Weyl invariance, we can use diff invariance to fix the metric partially, but one degree of freedom remains:

$$g_{ab}(\sigma) = \exp(2\phi(\sigma))\,\hat{g}_{ab}(\sigma) , \tag{9.9.7}$$

where $\hat{g}_{ab}$ is a gauge choice and the scale ϕ of the metric is still path-integrated. The Faddeev–Popov determinant for this gauge-fixing is the same as in eq. (3.4.19), the Weyl-fixing contribution to the latter being trivial. Inserting the metric (9.9.7), for which the curvature is

$$g^{1/2}R = \hat{g}^{1/2}\left(\hat{R} - 2\hat{\nabla}^2\phi\right) , \tag{9.9.8}$$

and dropping a field-independent term, the action plus Faddeev–Popov

[5] At special momenta one finds additional physical states. These can be interpreted as unbroken symmetries of string field theory, and likely play a role in the solvability of the theory.

determinant becomes

$$S = \frac{1}{4\pi\alpha'}\int_M d^2\sigma\,\hat{g}^{1/2}\left[\hat{g}^{ab}\partial_a X^\mu \partial_b X_\mu + \mu \exp 2\phi + \frac{13\alpha'}{3}\left(\hat{g}^{ab}\partial_a\phi\partial_b\phi + \hat{R}\phi\right)\right]. \tag{9.9.9}$$

Note that the Faddeev–Popov determinant has provided a kinetic term for ϕ. Comparing this with the dilaton–tachyon action (9.9.5) we see a marked resemblance. If we let $d = D - 1$ but interpret the new field ϕ as x^1, and if we rename $g_{ab} \to \hat{g}_{ab}$ in the action (9.9.5) (since it can be fully gauge-fixed), then all the same terms appear. Some coefficients are a bit different — there is no rescaling of ϕ that makes them all agree — but in fact we claim that these are the same field theory. The point is that the critical string path integral (9.9.5) is defined using $\hat{g}_{ab}$ as the metric, and the noncritical string (9.9.9) using $\exp(2\phi)\hat{g}_{ab}$. A careful treatment of the Weyl transformation of the measure then converts one to the other. One can understand this quite easily by noting that only the full metric (9.9.7) appears in the definition of the noncritical string, so nothing can change under

$$\hat{g}_{ab}(\sigma) \to \exp(2\omega(\sigma))\,\hat{g}_{ab}(\sigma)\,, \tag{9.9.10a}$$

$$\phi(\sigma) \to \phi(\sigma) - \omega(\sigma)\,. \tag{9.9.10b}$$

The noncritical string thus acquires a Weyl invariance, and this is enough to fix the various coefficients, other than the free parameter μ.

It may seem odd that Weyl-invariant and noninvariant theories can be equivalent. The point is that in field theory we can always add extra 'fake' gauge symmetries if we also add extra fields that we just use to fix the gauge. A gauge symmetry after all is just a redundancy of description, and we can always adopt a more redundant description. For the symmetry (9.9.10) it is clear that we can just cancel the Weyl symmetry against the field ϕ, but for general string backgrounds (for example flat space, where the embedding coordinates are Weyl-invariant) this is not possible and the Weyl invariance is 'real.'

In any case we are now nearing the punch line. One can approximate the noncritical string by a sum of triangulated surfaces, the number of edges between two points being a measure of length. These triangulated surfaces resemble large Feynman diagrams, and in fact one can find a quantum system, the *matrix model*, for which this is the diagram expansion. The fields of the matrix model are $N \times N$ matrices. Taking $N \to \infty$ and adjusting other couplings, there is (at least for $D \leq 2$) a critical point dominated by large smooth surfaces, so we recover the noncritical string. Moreover, the matrix model is solvable (again for $D \leq 2$) by diagonalizing the matrices. In the end one has a rather simple system (noninteracting fermions moving in the space of eigenvalues) which by retracing the above steps allows the calculation of string amplitudes to all genus.

Because these methods are so far afield from the rest of the book, we will have to leave the details to the references. In any case they have not yet been useful at $D > 2$ or in supersymmetric theories. (However, we will see in chapter 14 that a somewhat different theory of large-N matrices has recently been proposed as a nonperturbative string theory.) We should mention, though, at least one important lesson from these models. This is the $(2g)!$ behavior of perturbation theory, which was seen first in the explicit matrix model solution and then argued to hold in all string theories. In the matrix model the corresponding $\exp[O(-1/g_c)]$ effect involves tunneling of the fermions in the space of eigenvalues. It is not clear whether this can be related to the $\exp[O(-1/g_c)]$ effects in higher-dimensional string theories.

Exercises

9.1 (a) Use a conformal transformation to show that the OPE (9.1.2) and expectation value (9.1.3) imply the inner product (9.1.29).
(b) By taking matrix elements in the basis $|\hat{\mathcal{V}}_j(k)\rangle$, verify the second equality in equation (9.1.28).

9.2 (a) Write the four-point amplitude on the torus in operator form, parallel to eq. (9.1.17) for the sphere, but now involving a trace.
(b) Show that if one of the vertex operators is null, $\hat{\mathcal{V}}_1 = \{\mathcal{Q}^B, \mathcal{X}\}$, then the trace is equal to total derivatives with respect to the vertex operator positions plus total derivatives with respect to τ and $\bar{\tau}$.

9.3 (a) For the torus with four vertex operators, use the result of part (a) of the previous problem to study the unitarity in the limit that w_1, w_2, and $w_3 - w_4$ are fixed and w_3 and $2\pi\tau - w_3$ approach $i\infty$. This is two handles degenerating, and unitarity relates the amplitude to the product of two tree-level four-point amplitudes.
(b) Repeat in particular the demonstration that the sum over the full Hilbert space $\hat{\mathscr{H}}$ and that over the physical Hilbert space are equal. Show that this requires cancellation between intermediate states of different ghost numbers. This cancellation is a stringy generalization of the cancellation between Faddeev–Popov ghosts and longitudinal gauge bosons in gauge theory.

9.4 Verify the assertion in eq. (9.3.7) about the region covered by the Feynman graph of figure 9.6(a), and find the other two regions covered by this graph.

9.5 (a) For the terms (9.6.7) in the expansion of the string field, write out explicitly the gauge transformation (9.6.2) and the action (9.6.3).

(b) Eliminate auxiliary fields and fields that can be completely gauged away, and show that the resulting gauge transformation and action for A_μ have the expected form.

9.6 For $f_{2k} = k^a b^k$, show that the smallest term in the series (9.7.5) is of the order of $\exp(-C/g^2)$. Work to sufficient accuracy to find C. For the series (9.7.2), compare with the instanton action.

9.7 In quantum theory, the Euclidean path integral with Euclidean time having period T^{-1} gives the thermal partition function $\exp[-\tilde{F}(T)/T]$. Show that for the string path integral this agrees with the result (9.8.15). The one subtlety is an interplay between the region of integration and the sum over windings (Reference: McClain & Roth, (1987).)

9.8 For the tachyon wave equation (9.9.2) in the linear dilaton background (9.9.1), show that any *finite energy* initial configuration has stable (oscillatory) evolution provided that $D \leq 2$.

9.9 Study the physical state conditions (either in OCQ or BRST form) for the graviton level of the closed string in the $D = 2$ linear dilaton background. Let the momenta take complex values. You should find that the physical spectrum is trivial at generic momenta but that at two momenta there are nontrivial physical states.

Appendix A
A short course on path integrals

Path integrals are a powerful means of representing quantum theories, especially on surfaces of nontrivial topology. The introduction we present here is similar to that which the reader will find in any modern field theory text. We include it in order to emphasize certain ideas that we will need, such as the relation between the path integral and Hilbert space formalisms and the use of operator equations inside the path integral.

A.1 Bosonic fields

Consider first a quantum mechanics problem, one degree of freedom with Hamiltonian $\hat{H}(\hat{p}, \hat{q})$, where

$$[\hat{q}, \hat{p}] = i\hbar \,. \tag{A.1.1}$$

Throughout the appendix operators are indicated by hats. A basic quantity of interest is the amplitude to evolve from one q eigenstate to another in a time T,

$$\langle q_f | \exp(-i\hat{H}T/\hbar) | q_i \rangle \,. \tag{A.1.2}$$

In field theory it is generally convenient to use the Heisenberg representation, where operators have the time dependence

$$\hat{A}(t) = \exp(i\hat{H}t/\hbar)\, \hat{A}\, \exp(-i\hat{H}t/\hbar) \,. \tag{A.1.3}$$

The state $|q, t\rangle$ is an eigenstate of $\hat{q}(t)$,

$$\hat{q}(t)|q, t\rangle = |q, t\rangle q \,. \tag{A.1.4}$$

In terms of the $t = 0$ Schrödinger eigenstates this is

$$|q, t\rangle = \exp(i\hat{H}t/\hbar)|q\rangle \,. \tag{A.1.5}$$

In this notation the transition amplitude (A.1.2) is $\langle q_f, T | q_i, 0 \rangle$.

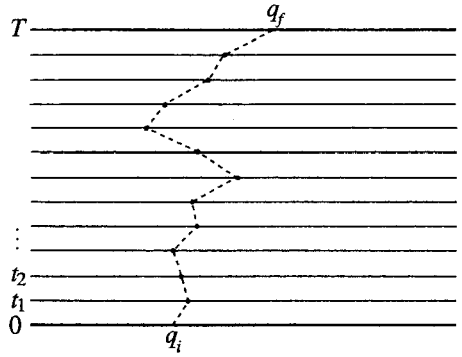


Fig. A.1. Transition amplitude broken down into time steps. The dashed line suggests a piecewise linear path, which is one of many ways to define path integration.

By inserting a complete set of states, we can write the transition amplitude as a coherent sum over all states q through which the system might pass at some intermediate time t:

$$\langle q_f, T|q_i, 0\rangle = \int dq\, \langle q_f, T|q, t\rangle\langle q, t|q_i, 0\rangle\ . \tag{A.1.6}$$

Divide the time interval further into N steps as in figure A.1,

$$t_m = m\epsilon\ , \quad \epsilon = T/N\ , \tag{A.1.7}$$

and corresponding to each intermediate time insert a complete set of states

$$\langle q_f, T|q_i, 0\rangle = \int dq_{N-1}\ldots dq_1 \prod_{m=0}^{N-1} \langle q_{m+1}, t_{m+1}|q_m, t_m\rangle\ . \tag{A.1.8}$$

Translate back to Schrödinger formalism and introduce an integral over intermediate momenta,

$$\begin{aligned}\langle q_{m+1}, t_{m+1}|q_m, t_m\rangle &= \langle q_{m+1}|\exp(-i\hat{H}\epsilon/\hbar)|q_m\rangle \\ &= \int dp_m\, \langle q_{m+1}|p_m\rangle\langle p_m|\exp(-i\hat{H}\epsilon/\hbar)|q_m\rangle\ .\end{aligned} \tag{A.1.9}$$

By commuting, we can always write $\hat{H}$ with all $\hat{p}$ to the left and all $\hat{q}$ to the right, so that

$$\langle p_m|\hat{H}(\hat{p}, \hat{q})|q_m\rangle = H(p_m, q_m)\langle p_m|q_m\rangle\ . \tag{A.1.10}$$

Use this to evaluate the matrix element in eq. (A.1.9). From expanding the exponential there will be terms of order ϵ^2 and higher that have $\hat{q}$ to the left of $\hat{p}$; drop these. Then to this order $\langle q_{m+1}, t_{m+1}|q_m, t_m\rangle$ becomes

$$\begin{aligned}&\int dp_m\, \exp[-iH(p_m, q_m)\epsilon/\hbar]\langle q_{m+1}|p_m\rangle\langle p_m|q_m\rangle \\ &= \int \frac{dp_m}{2\pi}\, \exp\left\{-\frac{i}{\hbar}\left[H(p_m, q_m)\epsilon - p_m(q_{m+1} - q_m)\right] + O(\epsilon^2)\right\}\ .\end{aligned} \tag{A.1.11}$$

Thus

$$\begin{aligned}\langle q_f, T|q_i, 0\rangle \approx &\int \frac{dp_{N-1}}{2\pi} dq_{N-1}\ldots\frac{dp_1}{2\pi}dq_1\frac{dp_0}{2\pi} \\ &\times \exp\left\{-\frac{i}{\hbar}\sum_{m=0}^{N-1}\left[H(p_m, q_m)\epsilon - p_m(q_{m+1} - q_m)\right]\right\} \\ &\to \int [dpdq]\exp\left\{\frac{i}{\hbar}\int_0^T dt\,[p\dot{q} - H(p, q)]\right\}\ .\end{aligned} \tag{A.1.12}$$

In the last line we have taken a formal $\epsilon \to 0$ limit, so the integral runs over all paths $p(t), q(t)$ with given $q(T)$ and $q(0)$. This is the *Hamiltonian path integral* (or *functional integral*), the sum over phase space paths weighted by $\exp(iS/\hbar)$ with S the action in Hamiltonian form.

Let us integrate over $p(t)$, making first the approximation that the integral is dominated by the point of stationary phase,

$$0 = \frac{\partial}{\partial p}[p\dot{q} - H(p, q)] = \dot{q} - \frac{\partial}{\partial p}H(p, q)\ . \tag{A.1.13}$$

Solving for p in terms of q and $\dot{q}$ and recalling that $L = p\dot{q} - H$, the stationary phase approximation gives

$$\langle q_f, T|q_i, 0\rangle = \int [dq]\exp\left[\frac{i}{\hbar}\int_0^T dt\, L(q, \dot{q})\right]\ . \tag{A.1.14}$$

Often the momentum integral is Gaussian, $\hat{H} = \frac{1}{2}\hat{p}^2 + V(\hat{q})$, so that the stationary phase approximation is exact up to a normalization that can be absorbed in the measure $[dq]$. In fact, we can do this generally: because the relation (A.1.13) between p and $\dot{q}$ is local, the corrections to the stationary phase approximation are local and so can be absorbed into the definition of $[dq]$ without spoiling the essential property that $[dq]$ is a product over independent integrals at each intermediate time. The result (A.1.14) is the *Lagrangian path integral*, the integral over all paths in configuration space weighted by $\exp(iS/\hbar)$, where now S is the Lagrangian action.

We have kept $\hbar$ so as to discuss the classical limit. As $\hbar \to 0$, the integral is dominated by the paths of stationary phase, $\delta S/\delta q(t) = 0$, which is indeed the classical equation of motion. Henceforth, $\hbar = 1$.

The derivation above clearly generalizes to a multicomponent q^i. One can always write a quantum field theory in terms of an infinite number of qs, either by making the spatial coordinate $\mathbf{x}$ discrete and taking a limit,

or by expanding the **x**-dependence in terms of Fourier modes. Thus the above derivation of the path integral representation applies to bosonic field theory as well.

The derivation of the path integral formula treated the continuum limit $\epsilon \to 0$ rather heuristically. Renormalization theory is the study of this limiting process, and is rather well understood. The limit exists provided the Lagrangian satisfies certain conditions (renormalizability) and provided the coefficients in the Lagrangian (bare couplings) are taken to vary appropriately as the limit is taken. The result is independent of various choices that must be made, including the operator ordering in the Hamiltonian, the precise definition of the measure, and the differences between regulators (the limit of piecewise linear paths suggested by the figure versus a frequency cutoff on the Fourier modes, for example).[1] Being local effects these can all be absorbed into the bare couplings. Actually, we will evaluate explicitly only a few Gaussian path integrals. We use the path integral primarily to derive some general results, such as operator equations and sewing relations.

Relation to the Hilbert space formalism

Starting from the path integral one may cut it open to recover the Hilbert space of the theory. In equation (A.1.6), if one writes each of the amplitudes as a path integral one finds

$$\int [dq]_{q_i,0}^{q_f,T} \exp\left(i\int_0^T L\right) = \int dq \int [dq]_{q,t}^{q_f,T} \exp\left(i\int_t^T L\right) \int [dq]_{q_i,0}^{q,t} \exp\left(i\int_0^t L\right) . \quad \text{(A.1.15)}$$

That is, the path integral on $[0, T]$ breaks up into separate path integrals on $[0, t]$ and $[t, T]$, plus an ordinary integral over $q(t)$. The boundary conditions on paths are denoted by super/subscripts. Note the restriction that $T \geq t \geq 0$. One can also replace the intermediate integral with a sum over any complete set of states. This cutting (or, inversely, sewing) principle is particularly useful on the surfaces of nontrivial topology that arise in string theory. There one may cut the surface open along many different closed curves; different choices give different Hilbert space representations of a given amplitude. This is known as *world-sheet duality*.

Consider now a path integral with the additional insertion of a factor

[1] As another example, one can freely rescale fields. The Jacobian is formally a product of separate Jacobians at each point; its logarithm is therefore a sum of a local expression at each point. Regulated, this is just a shift of the bare couplings.

of $q(t)$, where $0 < t < T$. Then, using the cutting relation (A.1.15),

$$\begin{aligned}\int [dq]_{q_i,0}^{q_f,T} \exp(iS) q(t) &= \int dq \, \langle q_f, T|q, t\rangle \, q \, \langle q, t|q_i, 0\rangle \\ &= \int dq \, \langle q_f, T|\hat{q}(t)|q, t\rangle \, \langle q, t|q_i, 0\rangle \\ &= \langle q_f, T|\hat{q}(t)|q_i, 0\rangle \, . \end{aligned} \quad \text{(A.1.16)}$$

Thus, $q(t)$ in the functional integral translates into $\hat{q}(t)$ in the matrix element. For a product of two insertions $q(t)q(t')$ (which is equal to $q(t')q(t)$, since these are simply variables of integration), where both t and t' are in the range $[0, T]$, one finds by the same method that

$$\int [dq]_{q_i,0}^{q_f,T} \exp(iS) q(t)q(t') = \langle q_f, T|\mathrm{T}[\hat{q}(t)\hat{q}(t')]|q_i, 0\rangle . \quad \text{(A.1.17)}$$

Here, T denotes the time-ordered product

$$\mathrm{T}[\hat{A}(t)\hat{B}(t')] = \theta(t - t')\hat{A}(t)\hat{B}(t') + \theta(t' - t)\hat{B}(t')\hat{A}(t) . \quad \text{(A.1.18)}$$

Because of the way the path integral is built out of successive infinitesimal time slices, two or more insertions in the path integral will always correspond to the time-ordered product of operators in the matrix element.

The operator equation of motion is $\delta S/\delta q(t) = 0$. Let us see how this works inside the path integral. Consider this equation in a path integral with other insertions $\mathscr{F}$,

$$\begin{aligned}\int [dq]_{q_i,0}^{q_f,T} \exp(iS) \frac{\delta S}{\delta q(t)} \mathscr{F} &= -i \int [dq]_{q_i,0}^{q_f,T} \left[\frac{\delta}{\delta q(t)} \exp(iS)\right] \mathscr{F} \\ &= i \int [dq]_{q_i,0}^{q_f,T} \exp(iS) \frac{\delta \mathscr{F}}{\delta q(t)} \, . \end{aligned} \quad \text{(A.1.19)}$$

We have integrated by parts in the second line.[2] If none of the fields in $\mathscr{F}$ are at time t, the right-hand side vanishes and the equation of motion holds in the path integral. We therefore say that the equation of motion holds as an *operator equation*. This is a technique that we will use extensively: an operator equation is one which holds when inserted into any path integral that has no other fields coincident with those involved in the equation. Translating to Hilbert space language, we can use the other insertions $\mathscr{F}$ to set up arbitrary initial and final states, so this corresponds

[2] The functional derivative is defined as

$$\frac{\delta}{\delta q(t)} q(t') = \delta(t - t') \, .$$

It can be regarded as the limit of the ordinary derivative

$$\frac{1}{\epsilon} \frac{\partial}{\partial q_m} , \quad m = t/\epsilon \, .$$

In particular, one can integrate a functional derivative by parts, as for an ordinary derivative.

to the ordinary notion that an operator equation is one that should hold for general matrix elements.

When additional insertions are present at t, operator equations inside the path integral hold only up to additional delta-function *source terms*. For example, let $\mathscr{F} = q(t')\dots$, where t' is near t and the other fields in '...' are far away. Then eq. (A.1.19) becomes

$$\left\langle q(t')\frac{\delta S}{\delta q(t)}\dots\right\rangle = i\langle\,\delta(t-t')\dots\rangle\ . \tag{A.1.20}$$

This is an example of the *Schwinger–Dyson equation*, the equation of motion for the expectation value of a product of fields. The same result written as an operator equation is

$$q(t')\frac{\delta S}{\delta q(t)} = i\delta(t-t')\ . \tag{A.1.21}$$

As an explicit example, take the harmonic oscillator,

$$S = \frac{1}{2}\int dt\,(\dot{q}^2-\omega^2q^2)\ , \tag{A.1.22a}$$

$$\frac{\delta S}{\delta q(t)} = -\ddot{q}(t)-\omega^2q(t)\ . \tag{A.1.22b}$$

Then the operator equations we have derived are

$$\left(\frac{\partial^2}{\partial t^2}+\omega^2\right)q(t) = 0 \tag{A.1.23}$$

and

$$\left(\frac{\partial^2}{\partial t^2}+\omega^2\right)q(t)q(t') = -i\delta(t-t')\ . \tag{A.1.24}$$

In Hilbert space language these become

$$\left(\frac{\partial^2}{\partial t^2}+\omega^2\right)\hat{q}(t) = 0\ ,$$

$$\left(\frac{\partial^2}{\partial t^2}+\omega^2\right)\mathrm{T}[\hat{q}(t)\hat{q}(t')] = -i\delta(t-t')\ . \tag{A.1.25a}$$

The first is the Heisenberg equation for the field operator, while the delta function in the second comes from the time derivative acting on the step functions in the time-ordered product,

$$\left(\frac{\partial^2}{\partial t^2}+\omega^2\right)\mathrm{T}\,[\hat{q}(t)\hat{q}(t')] = \mathrm{T}\left\{\left[\ddot{\hat{q}}(t)+\omega^2\hat{q}(t)\right]\hat{q}(t')\right\}+\delta(t-t')[\dot{\hat{q}}(t),\hat{q}(t)]$$

$$= 0\ -\ i\delta(t-t')\ . \tag{A.1.26}$$

Euclidean path integrals

In Schrödinger language, the time-ordered product (A.1.17) is

$$\langle q_f|\exp[-i\hat{H}(T-t_>)]\,\hat{q}\,\exp[-i\hat{H}(t_>-t_<)]\,\hat{q}\,\exp[-i\hat{H}t_<]|q_i\rangle\ , \tag{A.1.27}$$

where $t_>$ is the greater of t and t', and $t_<$ is the lesser. By the nature of the path integral, which builds amplitudes from successive time steps, the time differences multiplying H are always positive in amplitudes such as (A.1.27): one only encounters $\exp(-iH\Delta t)$ for positive Δt. Now continue analytically by uniformly rotating all times in the complex plane, $t \to e^{-i\theta}t$ for any phase $0 \le \theta \le \pi$. The real part of the exponent is $-E\sin\theta\Delta t$, where E is the energy of the intermediate state. Because the Hamiltonian is bounded below for a stable system, the sum over intermediate states is safely convergent and no singularities are encountered — the continued amplitude is well defined.

Taking the times to be purely imaginary, $t = -iu$ for real u, defines the *Euclidean* amplitudes, as opposed to the *Minkowski* amplitudes studied above:

$$\langle q_f,u|\mathrm{T}\left[\prod_a\hat{q}(u_a)\right]|q_i,0\rangle_{\mathrm{E}} \tag{A.1.28}$$

is defined to be the continuation of

$$\langle q_f,t|\mathrm{T}\left[\prod_a\hat{q}(t_a)\right]|q_i,0\rangle \tag{A.1.29}$$

to $t = -iu$. One can repeat the derivation of the path integral representation for the Euclidean case, but the result is extremely simple: just replace t in the action with $-iu$,

$$\langle q_f,U|\mathrm{T}\left[\prod_a\hat{q}(u_a)\right]|q_i,0\rangle_{\mathrm{E}} = \int[dq]_{q_i,U}^{q_f,0}\exp\left[\int_0^U du\,L(q,i\partial_u q)\right]\prod_a q(u_a)\ . \tag{A.1.30}$$

The weight in the Euclidean path integral is conventionally defined to be $\exp(-S_{\mathrm{E}})$, so the Euclidean action contains a minus sign,

$$S_{\mathrm{E}} = \int du\,L_{\mathrm{E}}(q,\partial_u q) = -\int du\,L(q,i\partial_u q)\ . \tag{A.1.31}$$

Euclidean path integrals are usually the more well-defined, because the exponential of the action is damped rather than oscillatory. The Minkowski matrix elements can then be obtained by analytic continuation. In the text our main concern is the Euclidean path integral, which we use to represent the string S-matrix.

One is often interested in traces, such as

$$\mathrm{Tr}\exp(-\hat{H}U) = \sum_i\exp(-E_iU)\ , \tag{A.1.32}$$

where the sum runs over all eigenvectors of $\hat{H}$ and the E_i are the corresponding eigenvalues. In the q basis the trace becomes

$$\begin{aligned}\int dq\,\langle q|\exp(-\hat{H}U)|q\rangle &= \int dq\,\langle q,U|q,0\rangle_{\mathrm{E}} \\ &= \int dq \int [dq]_{q,0}^{q,U} \exp(-S_{\mathrm{E}}) \\ &= \int [dq]_{\mathrm{P}} \exp(-S_{\mathrm{E}})\ . \end{aligned} \tag{A.1.33}$$

In the last line, $[dq]_{\mathrm{P}}$ denotes the integral over all *periodic* paths on $[0,U]$. An interesting variation on this is the integral over all *antiperiodic* paths:

$$\begin{aligned}\int [dq]_{\mathrm{A}} \exp(-S_{\mathrm{E}}) &= \int dq \int [dq]_{-q,0}^{q,U} \exp(-S_{\mathrm{E}}) \\ &= \int dq\,\langle q,U|-q,0\rangle_{\mathrm{E}} \\ &= \int dq\,\langle q,U|\hat{R}|q,0\rangle_{\mathrm{E}} \\ &= \mathrm{Tr}\left[\exp(-\hat{H}U)\hat{R}\right]\ . \end{aligned} \tag{A.1.34}$$

Here $\hat{R}$ is the reflection operator, taking q to $-q$. We see that *twisting* the boundary conditions — for example, replacing periodic with antiperiodic — is equivalent to inserting an operator into the trace. Since $\hat{R}^2 = 1$, its eigenvalues are ± 1. If we take the average of the periodic and antiperiodic path integrals, this inserts $\frac{1}{2}(1+\hat{R})$, which projects onto the $\hat{R} = +1$ space:

$$\frac{1}{2}(\text{periodic} + \text{antiperiodic}) = \sum_{R_i=+1} \exp(-E_i U)\ . \tag{A.1.35}$$

In the same way, $\frac{1}{2}$(periodic − antiperiodic) gives the trace on the $\hat{R} = -1$ space.

A minor but sometimes confusing point: if $\hat{A}$ is a Hermitean operator, the Euclidean Heisenberg operator $\hat{A}(u) = \exp(\hat{H}u)\hat{A}\exp(-\hat{H}u)$ satisfies

$$\hat{A}(u)^\dagger = \hat{A}(-u) \tag{A.1.36}$$

and so is not in general Hermitean. For this reason it is useful to define a *Euclidean adjoint*,

$$\overline{\hat{A}(u)} = \hat{A}(-u)^\dagger\ . \tag{A.1.37}$$

A Hermitean operator in the Schrödinger picture becomes Euclidean self-adjoint in the Euclidean Heisenberg picture.

Diagrams and determinants

The diagrammatic expansion is easily obtained from the path integral. This subject is well treated in modern field theory texts. We will only derive here a few results used in the text. For notational clarity we switch to field theory, replacing $q(t)$ with a real scalar field $\phi(t,\mathbf{x})$. Begin with the free action

$$\frac{1}{2}\int d^dx\,\phi(x)\Delta\phi(x)\ , \tag{A.1.38}$$

where Δ is some differential operator. Consider the path integral with a source $J(x)$, taking the Euclidean case to be specific,

$$\begin{aligned} Z[J] &= \left\langle \exp\left[i\int d^dx\,J(x)\phi(x)\right]\right\rangle \\ &= \int [d\phi]\,\exp\left[-\frac{1}{2}\int d^dx\,\phi(x)\Delta\phi(x) + i\int d^dx\,J(x)\phi(x)\right]\ . \end{aligned} \tag{A.1.39}$$

This is the generating functional for Euclidean Green's functions. Taking functional derivatives with respect to $J(x)$ pulls down factors of $i\phi(x)$ into the integral. Shifting ϕ to $\phi + i\Delta^{-1}J$ completes the square, leaving

$$\begin{aligned} Z[J] &= Z[0]\exp\left[-\frac{1}{2}\int d^dx\,d^dy\,J(x)\Delta^{-1}(x,y)J(y)\right] \\ &= Z[0]\exp\left[\frac{1}{2}\int d^dx\,d^dy\,\Delta^{-1}(x,y)\frac{\delta^2}{\delta\phi(x)\delta\phi(y)}\right] \\ &\qquad \times \exp\left[i\int d^dx\,J(x)\phi(x)\right]\Big|_{\phi=0}\ . \end{aligned} \tag{A.1.40}$$

In the second line, each ϕ derivative becomes iJ. Any functional $\mathscr{F}$ of ϕ can be written

$$\mathscr{F}[\phi] = \int [dJ]\,\exp\left[i\int d^dx\,J(x)\phi(x)\right] f[J]\ , \tag{A.1.41}$$

where $f[J]$ is the (functional) Fourier transform of $\mathscr{F}$. By linearity one obtains

$$\langle\,\mathscr{F}\,\rangle = Z[0]\exp\left[\frac{1}{2}\int d^dx\,d^dy\,\Delta^{-1}(x,y)\frac{\delta^2}{\delta\phi(x)\delta\phi(y)}\right]\mathscr{F}[\phi]\Big|_{\phi=0}\ . \tag{A.1.42}$$

Expanding the exponential generates the diagrammatic expansion, all ways of contracting pairs of fields in $\mathscr{F}$. Also, when $\mathscr{F}$ is a product of factors, $\mathscr{F}_1\mathscr{F}_2\ldots$, we can separate

$$\frac{\delta}{\delta\phi(x)} = \frac{\delta}{\delta\phi_1(x)} + \frac{\delta}{\delta\phi_2(x)} + \ldots\ , \tag{A.1.43}$$

where $\delta/\delta\phi_i$ acts only on $\mathscr{F}_i$. The *self-contractions* are then the terms

$$\frac{\delta^2}{\delta\phi_i(x)\delta\phi_i(y)} \quad \text{(no sum on } i\text{)} . \tag{A.1.44}$$

Consider now the overall factor $Z[0]$. We assume that Δ is self-adjoint and so we expand $\phi(x)$ in a complete set of eigenfunctions $\Phi_i(x)$:

$$\phi(x) = \sum_i \phi_i \Phi_i(x) , \tag{A.1.45}$$

where

$$\Delta\Phi_i(x) = \lambda_i \Phi_i(x) , \tag{A.1.46a}$$

$$\int d^d x\, \Phi_i(x)\Phi_j(x) = \delta_{ij} . \tag{A.1.46b}$$

Define the path integral measure by

$$[d\phi] = \prod_i d\phi_i . \tag{A.1.47}$$

Then

$$\begin{aligned} \langle 1 \rangle &= \prod_i \left[\int d\phi_i \exp\left(-\frac{1}{2}\phi_i^2 \lambda_i\right)\right] \\ &= \prod_i \left(\frac{2\pi}{\lambda_i}\right)^{1/2} \\ &= \left(\det \frac{\Delta}{2\pi}\right)^{-1/2} . \end{aligned} \tag{A.1.48}$$

The product of eigenvalues must be suitably regulated and renormalized. The value of $\langle 1 \rangle$ is usually given simply as $(\det \Delta)^{-1/2}$, because a rescaling of Δ is equivalent to a rescaling of ϕ, the Jacobian for which can be absorbed into the bare couplings. (In free field theory, the only 'coupling' that would be renormalized would be a c-number constant in the Lagrangian density; we will see this appearing in the example.)

All of the above discussion extends readily to multicomponent q^i or ϕ^i. One case of interest is a complex scalar field. One can break this up into two real scalars, $\phi = 2^{-1/2}(\phi^1 + i\phi^2)$, and the measure is $[d\phi\, d\phi^*] \equiv [d\phi^1\, d\phi^2]$. For complex fields, (A.1.48) becomes

$$\int [d\phi\, d\phi^*] \exp\left(-\int d^d x\, \phi^* \Delta\phi\right) = (\det \Delta)^{-1} . \tag{A.1.49}$$

One similarly finds for two real fields

$$\int [d\phi^1\, d\phi^2] \exp\left(i \int d^d x\, \phi^1 \Delta \phi^2\right) = (\det \Delta)^{-1} . \tag{A.1.50}$$

This is a functional version of the delta-function integral

$$\int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy \exp(i\lambda xy) = 2\pi \int_{-\infty}^{\infty} dx\, \delta(\lambda x) = \frac{2\pi}{\lambda} . \tag{A.1.51}$$

An example

As an example, we study the Euclidean transition amplitude for the harmonic oscillator,

$$\langle q_f, U | q_i, 0 \rangle_{\mathrm{E}} = \int [dq]_{q_i,0}^{q_f,U} \exp(-S_{\mathrm{E}}) , \tag{A.1.52a}$$

$$S_{\mathrm{E}} = \frac{1}{2}\int_0^U du \left[(\partial_u q)^2 + \omega^2 q^2\right] + S_{\mathrm{ct}} , \tag{A.1.52b}$$

where S_{ct} is a cutoff-dependent piece to be determined later. To evaluate this, separate out the 'classical' part of q,

$$q(u) = q_{\mathrm{cl}}(u) + q'(u) , \tag{A.1.53}$$

where

$$-\ddot{q}_{\mathrm{cl}}(u) + \omega^2 q_{\mathrm{cl}}(u) = 0 , \tag{A.1.54a}$$

$$q_{\mathrm{cl}}(0) = q_i , \quad q_{\mathrm{cl}}(U) = q_f . \tag{A.1.54b}$$

The boundary conditions on q' are then

$$q'(0) = q'(U) = 0 \tag{A.1.55}$$

and the action separates

$$S_{\mathrm{E}} = S_{\mathrm{cl}}(q_i, q_f) + S' + S_{\mathrm{ct}} . \tag{A.1.56}$$

Here

$$\begin{aligned} S_{\mathrm{cl}}(q_i, q_f) &= \frac{1}{2}\int_0^U du \left[(\partial_u q_{\mathrm{cl}})^2 + \omega^2 q_{\mathrm{cl}}^2\right] \\ &= \omega \frac{(q_i^2 + q_f^2)\cosh \omega U - 2q_i q_f}{2 \sinh \omega U} , \end{aligned} \tag{A.1.57a}$$

$$S' = \frac{1}{2}\int_0^U du \left[(\partial_u q')^2 + \omega^2 q'^2\right] . \tag{A.1.57b}$$

The amplitude becomes

$$\langle q_f, U | q_i, 0 \rangle = \exp[-S_{\mathrm{cl}}(q_i, q_f) - S_{\mathrm{ct}}] \int [dq']_{0,0}^{0,U} \exp(-S') . \tag{A.1.58}$$

Expand in a complete set of eigenfunctions of $\Delta = -\partial_u^2 + \omega^2$ subject to

the boundary condition (A.1.55),

$$f_j(u) = \left(\frac{2}{U}\right)^{1/2} \sin\frac{j\pi u}{U}\,, \quad \lambda_j = \frac{j^2\pi^2}{U^2} + \omega^2\,, \tag{A.1.59a}$$

$$\det\frac{\Delta}{2\pi} = \prod_{j=1}^{\infty} \frac{j^2\pi^2 + \omega^2 U^2}{2\pi U^2}\,. \tag{A.1.59b}$$

The infinite product diverges. The *Pauli–Villars regulator* is a simple way to define it. Divide by the amplitude for a *regulator* oscillator of very high frequency Ω,

$$\det\frac{\Delta}{2\pi} \overset{\text{regulate}}{\to} \prod_{j=1}^{\infty} \frac{j^2\pi^2 + \omega^2 U^2}{j^2\pi^2 + \Omega^2 U^2}\,. \tag{A.1.60}$$

For j small, the terms in (A.1.60) are the same as those in (A.1.59b) up to a multiplicative constant, while for j large the product converges. The value of the product (A.1.60) is

$$\frac{\Omega \sinh\omega U}{\omega \sinh\Omega U}\,. \tag{A.1.61}$$

This follows from a standard infinite product representation for the sinh. One can guess the result (A.1.61) by noting that it has the right poles and zeros in U, and the right value as U goes to zero.

Keeping only terms that are nonvanishing as $\Omega \to \infty$,

$$\langle q_f, U|q_i, 0\rangle \to \left(\frac{\omega}{2\sinh\omega U}\right)^{1/2} \exp\left[-S_{\rm cl}(q_i, q_f) + \frac{1}{2}(\Omega U - \ln\Omega) - S_{\rm ct}\right]\,. \tag{A.1.62}$$

To get a finite answer as $\Omega \to \infty$, we need first to include a term $\frac{1}{2}\Omega$ in the Lagrangian $L_{\rm ct}$, canceling the linear divergence in (A.1.62). That is, there is a linearly divergent bare coupling for the operator '1'. It may seem strange that we need to renormalize in a quantum mechanics problem, but the power counting is completely uniform with quantum field theory. The logarithmic divergence is a wavefunction renormalization. To determine the finite normalization, compare the path integral expression (A.1.62) to the canonical result as $U \to \infty$. The $U \to \infty$ amplitude is easily evaluated by inserting a complete set of states, since only the ground state $|0\rangle$ contributes in the limit:

$$\begin{aligned}\langle q_f, U|q_i, 0\rangle &\to \langle q_f|0\rangle\langle 0|q_i\rangle \exp(-E_0 U) \\ &= \left(\frac{\omega}{\pi}\right)^{1/2} \exp\left[-\frac{\omega}{2}(q_f^2 + q_i^2 + U)\right]\,.\end{aligned} \tag{A.1.63}$$

On the other hand, the path integral expression (A.1.62) becomes

$$\langle q_f, U|q_i, 0\rangle \to \omega^{1/2} \exp\left[-\frac{\omega}{2}(q_f^2 + q_i^2 + U) + \frac{1}{2}(\Omega U - \ln\Omega) - S_{\rm ct}\right]\,. \tag{A.1.64}$$

Thus

$$S_{\rm ct} = \frac{1}{2}\left[\int_0^U du\,\Omega\right] - \frac{1}{2}\ln\frac{\Omega}{\pi} \tag{A.1.65}$$

and

$$\langle q_f, U|q_i, 0\rangle = \left(\frac{\omega}{2\pi\sinh\omega U}\right)^{1/2} \exp(-S_{\rm cl})\,, \tag{A.1.66}$$

which is the correct result.

A.2 Fermionic fields

Consider the basic fermionic quantum system, two states $|\uparrow\rangle$ and $|\downarrow\rangle$ with raising operator $\hat{\chi}$ and lowering operator $\hat{\psi}$:

$$\hat{\psi}|\downarrow\rangle = 0\,, \quad \hat{\chi}|\downarrow\rangle = |\uparrow\rangle\,, \tag{A.2.1a}$$

$$\hat{\psi}|\uparrow\rangle = |\downarrow\rangle\,, \quad \hat{\chi}|\uparrow\rangle = 0\,, \tag{A.2.1b}$$

$$\{\hat{\psi}, \hat{\chi}\} = 1\,, \quad \hat{\psi}^2 = \hat{\chi}^2 = 0\,. \tag{A.2.1c}$$

Amplitudes will be expressed as path integrals over the classical analogs of $\hat{\chi}$ and $\hat{\psi}$. These are *anticommuting c-numbers* or *Grassmann variables*, elements χ_m and ψ_m of a Grassmann algebra

$$\{\chi_m, \psi_n\} = \{\chi_m, \chi_n\} = \{\psi_m, \psi_n\} = 0\,. \tag{A.2.2}$$

There is a natural notion of integration over Grassmann variables, the Berezin integral. Notice that if θ is a Grassmann variable then $\theta^2 = 0$, and so the most general function of θ is of the form $a + \theta b$. Define

$$\int d\theta = 0\,, \quad \int d\theta\,\theta = -\int \theta\,d\theta = 1\,. \tag{A.2.3}$$

Integrals of more general expressions are determined by requiring the integral to be linear and by defining $d\theta$ to anticommute with all Grassmann variables. The significance of the definition (A.2.3) is that the integral of a total derivative is zero:

$$\int d\theta\,\frac{d}{d\theta}(a + \theta b) = \int d\theta\, b = 0\,. \tag{A.2.4}$$

For multiple Grassmann variables the integral is

$$\int d\theta_1 \ldots d\theta_n f(\theta) = c\,, \tag{A.2.5}$$

where c is the coefficient of the highest term in the Taylor expansion of f,

$$f(\theta) = \ldots + \theta_n\theta_{n-1}\ldots\theta_1 c\,. \tag{A.2.6}$$

Looking back on the discussion of the bosonic path integral, one sees that all of the properties of ordinary integrals that were used are also true for

the Berezin integral: linearity, invariance (except for a sign) under change of order of integration, and integration by parts.[3]

In order to follow the bosonic discussion as closely as possible, it is useful to define states that are formally eigenstates of $\hat{\psi}$:

$$|\psi\rangle = |\downarrow\rangle + |\uparrow\rangle\psi \,, \tag{A.2.7a}$$

$$\hat{\psi}|\psi\rangle = |\psi\rangle\psi \,. \tag{A.2.7b}$$

Define $\langle\psi|$ to satisfy

$$\langle\psi|\psi'\rangle = \psi - \psi' \,. \tag{A.2.8}$$

The right-hand side is the Grassmann version of the Dirac delta function. It has the same properties as the ordinary delta function,

$$(\psi - \psi')f(\psi) = (\psi - \psi')f(\psi') \tag{A.2.9}$$

and

$$\int d\psi\,(\psi - \psi')f(\psi) = f(\psi') \tag{A.2.10}$$

for arbitrary $f(\psi)$. The state $\langle\psi|$ is a left eigenstate with opposite sign,

$$\langle\psi|\hat{\psi} = -\psi\langle\psi| \,, \tag{A.2.11}$$

as one verifies by taking the inner product of both sides with a general state $|\psi'\rangle$. Note also that

$$\int |\psi\rangle\, d\psi\, \langle\psi| = 1 \,. \tag{A.2.12}$$

One can verify this by taking the matrix element between arbitrary states $\langle\psi'|$ and $|\psi''\rangle$ and evaluating the Grassmann integral.

The delta function has an integral representation

$$-\int d\chi\, \exp[\chi(\psi' - \psi)] = \psi - \psi' = \langle\psi|\psi'\rangle \,. \tag{A.2.13}$$

By expanding the inner product (A.2.8) in powers of ψ' one obtains

$$\langle\psi|\downarrow\rangle = \psi \,, \quad \langle\psi|\uparrow\rangle = -1 \,, \tag{A.2.14}$$

and therefore

$$\begin{aligned}\langle\psi|\hat{\chi}|\psi'\rangle &= -\langle\psi|\uparrow\rangle \\ &= -\int d\chi\, \chi \exp[\chi(\psi' - \psi)] \,.\end{aligned} \tag{A.2.15}$$

[3] It is not useful to think of $\int d\theta$ as an integral in any ordinary sense: it is simply a linear mapping from functions of θ to the complex numbers. Since it shares so many of the properties of ordinary integrals, the integral notation is a useful mnemonic.

Using the further property that $|\psi\rangle$ is an eigenstate of $\hat{\psi}$, one obtains

$$-\langle\psi|f(\hat{\chi},\hat{\psi})|\psi'\rangle = \int d\chi\, f(\chi,\psi')\exp[\chi(\psi' - \psi)] \,. \tag{A.2.16}$$

This holds for any function f with the prescription that it is written with the $\hat{\psi}$ on the right.

We can now emulate the bosonic discussion, using the completeness relation (A.2.12) to join together successive time steps, with the result

$$\begin{aligned}\langle\psi_f, T|\psi_i, 0\rangle \approx \int d\chi_0 d\psi_0 d\chi_1 \ldots d\psi_{N-1} d\chi_{N-1} \\ \times \exp\sum_{m=0}^{N-1}\left[-iH(\chi_m,\psi_m)\epsilon - \chi_m(\psi_{m+1} - \psi_m)\right] \\ \to \int [d\chi d\psi] \exp\left\{\int_0^T dt\left[-\chi\dot{\psi} - iH(\chi,\psi)\right]\right\} \,.\end{aligned} \tag{A.2.17}$$

The inconvenient sign in the integral representation has been offset by reversing the order of Grassmann integrations. A typical fermionic Hamiltonian is $\hat{H} = m\hat{\chi}\hat{\psi}$. The corresponding Lagrangian is $L = i\chi\dot{\psi} - H$. For such first order Lagrangians the Hamiltonian and Lagrangian formulations coincide, and the result (A.2.17) is also the Lagrangian path integral.

There are two useful inner products for this system,

$$\text{(A)}: \quad \langle\uparrow|\uparrow\rangle = \langle\downarrow|\downarrow\rangle = 1 \,, \quad \langle\uparrow|\downarrow\rangle = \langle\downarrow|\uparrow\rangle = 0 \,, \tag{A.2.18a}$$

$$\text{(B)}: \quad \langle\uparrow|\downarrow\rangle = \langle\downarrow|\uparrow\rangle = 1 \,, \quad \langle\uparrow|\uparrow\rangle = \langle\downarrow|\downarrow\rangle = 0 \,. \tag{A.2.18b}$$

Inner product (A) is positive definite and gives $\hat{\psi}^\dagger = \hat{\chi}$. It is the inner product for an ordinary physical oscillator. Inner product (B) is not positive, and it makes $\hat{\psi}$ and $\hat{\chi}$ Hermitean. It is the appropriate inner product for the Faddeev–Popov ghost system, as discussed in chapter 4. The derivation of the path integral was independent of the choice of inner product, but the relations (A.2.14) determine the expansion of $\langle\psi|$ in basis states in terms of the inner product.

All of the discussion of the bosonic case applies to the fermionic path integral as well. The derivation is readily extended to multiple Fermi oscillators and to field theory. Renormalization theory works in the same way as for the bosonic path integral. Fermionic path integrals again satisfy cutting relations, which simply amount to taking one of the iterated integrals (A.2.17) and saving it for the end. Insertion of $\psi(t)$ and $\chi(t)$ in the path integral again gives the corresponding operator in the matrix element, and multiple insertions are again time-ordered. The path integral insertion

$$\psi(t)\psi(t') = -\psi(t')\psi(t) \tag{A.2.19}$$

becomes the operator

$$\mathrm{T}[\hat{\psi}(t)\hat{\psi}(t')] \equiv \theta(t-t')\hat{\psi}(t)\hat{\psi}(t') - \theta(t'-t)\hat{\psi}(t')\hat{\psi}(t) . \quad \text{(A.2.20)}$$

Note that a minus sign appears in the definition of time-ordering. The equations of motion and the Euclidean continuation work as in the bosonic case, but there is one important change. Consider a general operator $\hat{A}$,

$$\hat{A}|i\rangle = |j\rangle A_{ji}, \quad \text{(A.2.21)}$$

where i, j run over $\downarrow$ and $\uparrow$. Then by expanding $\hat{A}|\psi\rangle$ in terms of the basis states one finds

$$\begin{aligned}\int d\psi \, \langle\psi|\hat{A}|\psi\rangle &= A_{\downarrow\downarrow} - A_{\uparrow\uparrow} \\ &= \mathrm{Tr}\left[(-1)^{\hat{F}}\hat{A}\right] , \end{aligned} \quad \text{(A.2.22)}$$

where the fermion number $\hat{F}$ has been defined to be even for $|\downarrow\rangle$ and odd for $|\uparrow\rangle$. In particular,

$$\int d\psi \, \langle\psi, U|\psi, 0\rangle = \mathrm{Tr}\left[(-1)^{\hat{F}} \exp(-\hat{H}U)\right] . \quad \text{(A.2.23)}$$

In contrast to the bosonic results (A.1.33) and (A.1.34), the *periodic* path integral gives the trace weighted by $(-1)^{\hat{F}}$ and the *antiperiodic* path integral gives the simple trace.

Feynman diagrams are as in the bosonic case, but the determinants work out differently. Consider

$$\langle 1 \rangle = \int [d\psi \, d\chi] \exp\left(\int d^d x \, \chi \Delta \psi\right) , \quad \text{(A.2.24)}$$

where we have again switched to the notation of field theory. Expand ψ in eigenfunctions of Δ, and χ in eigenfunctions of Δ^T,

$$\psi(x) = \sum_i \psi_i \Psi_i(x) , \quad \chi(x) = \sum_i \chi_i \Upsilon_i(x) , \quad \text{(A.2.25a)}$$

$$\Delta\Psi_i(x) = \lambda_i \Psi_i(x) , \quad \Delta^T \Upsilon_i(x) = \lambda_i \Upsilon_i(x) , \quad \text{(A.2.25b)}$$

$$\int d^d x \, \Upsilon_i(x)\Psi_j(x) = \delta_{ij} . \quad \text{(A.2.25c)}$$

Define $[d\psi d\chi] = \prod_i d\psi_i d\chi_i$. Then

$$\begin{aligned}\langle 1 \rangle &= \prod_i \left[\int d\psi_i d\chi_i \exp(\lambda_i \chi_i \psi_i)\right] \\ &= \prod_i \lambda_i \\ &= \det \Delta . \end{aligned} \quad \text{(A.2.26)}$$

Similarly, one finds for a single field ψ that

$$\int [d\psi] \exp\left(\int d^d x \, \psi \Delta \psi\right) = (\det \Delta)^{1/2} , \quad \text{(A.2.27)}$$

where Δ must be antisymmetric. These fermionic path integrals are just the inverse of the corresponding bosonic integrals.

For finite-dimensional integrals we have to be careful about 2πs,

$$\int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy \, \exp(2\pi i \lambda x y) = \frac{1}{\lambda} = \left[\int d\psi \int d\chi \, \exp(\lambda \chi \psi)\right]^{-1} . \quad \text{(A.2.28)}$$

Curiously, we can be more cavalier for functional determinants, where these can be eliminated by rescaling the fields. Bosonic path integrals actually give the absolute values of determinants, while fermionic path integrals give the determinant with a sign that depends on the precise order of Grassmann factors in the measure. We will fix the sign by hand.[4]

If Δ has zero eigenvalues, the path integral vanishes in the fermionic case and diverges in the bosonic case. When this occurs there is always a good reason, as we will see in practice, and the integrals of interest will have appropriate insertions to give a finite result.

A useful result about Berezin integrals: from the definition it follows that if $\theta' = \theta a$ then $d\theta' = a^{-1} d\theta$. That is, the Jacobian is the inverse of the bosonic case. More generally, with bosonic variables x_i and Grassmann variables θ_j, under a general change of variables to $x_i'(x, \theta)$, $\theta_j'(x, \theta)$ the infinitesimal transformation of the measure has an extra minus sign,

$$\delta(d^m x \, d^n \theta) = d^m x \, d^n \theta \left(\sum_i \frac{\partial}{\partial x_i} \delta x^i - \sum_j \frac{\partial}{\partial \theta_j} \delta \theta_j\right) . \quad \text{(A.2.29)}$$

Exercises

A.1 (a) Evaluate the path integral (A.1.33) for the harmonic oscillator by expanding $[dq]_{\mathrm{P}}$ in a set of periodic eigenfunctions (not by inserting the result already obtained for the transition function into the left-hand side). Compare with an explicit evaluation of the sum over states (A.1.32).
(b) Do the same for the antiperiodic path integral (A.1.34).

A.2 Consider two harmonic oscillators of the same frequency, and form the complex coordinate $q = 2^{-1/2}(q_1 + iq_2)$. Evaluate the following path

[4] A fine point: when the world-sheet theory is parity-asymmetric, even the phase of the path integral is ambiguous. The point is that Δ is not Hermitian in this case, because the Euclidean adjoint includes a time-reversal (which is the same as a parity flip plus Euclidean rotation). One must instead take the determinant of the Hermitian operator $\Delta^\dagger\Delta$, which determines that of Δ only up to a phase. We will encounter this in section 10.7, for example, where we will have to check that there is a consistent definition of the phase.

integral on $[0, U]$ by using a suitable mode expansion:

$$\int [dq\, dq^*]_\theta \exp(-S_E) ,$$

where the integral is to run over all paths such that $q(U) = q(0)e^{i\theta}$. Again compare with the answer obtained by summing over states: it is useful here to form complex combinations of oscillators, $a_1 \pm i a_2$. This is a more exotic version of exercise A.1, and arises in various string theories such as orbifolds.

A.3 Consider a free scalar field X in two dimensions, with action

$$S = \frac{1}{4\pi\alpha'} \int d^2\sigma \left(\partial_1 X \partial_1 X + \partial_2 X \partial_2 X + m^2 X^2\right) .$$

Evaluate the path integral with both coordinates periodic,

$$0 \le \sigma_1 \le 2\pi , \quad 0 \le \sigma_2 \le T .$$

[Expanding the spatial dependence in Fourier modes, the answer takes the form of an infinite product of harmonic oscillator traces.]

A.4 Consider a free particle on a circle,

$$H = -\frac{1}{2}\frac{\partial^2}{\partial\phi^2} .$$

By means of path integrals, evaluate $\langle \phi_f, T | \phi_i, 0 \rangle$. Be careful: there are topologically inequivalent paths, distinguished by the number of times they wind around the circle.

A.5 The fermionic version of exercise A.1: evaluate the fermionic path integral (A.2.23) and the corresponding antiperiodic path integral by using a complete set of periodic or antiperiodic functions, and compare with the sum over states.

References

The literature on the subjects covered in these two volumes easily exceeds 5000 papers, and probably approaches twice that number. To represent accurately the contributions of all who have worked in these many areas is an undertaking beyond the scope of this book. Rather, I have tried to assemble a list that is short enough to be useful in developing specific subjects and in giving points of entry to the literature. The resulting list includes many review articles, plus papers from the original literature where I felt they were needed to supplement the treatment in the text.

Papers listed as 'e-print: hep-th/yymmnnn' are available electronically at the Los Alamos physics e-print archive,

http://xxx.lanl.gov/abs/hep-th/yymmnnn .

For a list of corrections and other useful information see

http://www.itp.ucsb.edu/~joep/bigbook.html .

General references

The reader may find the two volume set by Green, Schwarz, & Witten (1987) (henceforth GSW) a useful supplement. The mix of covariant and light-cone methods used in that book, and the advanced topics emphasized, are somewhat complementary to the approach taken here. The book by Polyakov (1987) emphasizes conformal field theory and the path integral approach to strings, but the focus is less on unified string theories and more on string theories of the strong interaction and on the statistical mechanics of random surfaces. Some other lectures on string theory are Peskin (1987), Lust & Theisen (1989), Alvarez-Gaumé & Vazquez-Mozo (1992), D'Hoker (1993), Ooguri & Zin (1997), and Kiritsis (1997).

A brief history

There have been three periods of intense activity in the history of string theory, covering roughly the years 1968–73, 1984–9, and 1995 to the present. The first of these began with the Veneziano amplitude as a phenomenological proposal for the S-matrix of the strong interaction. This led to *dual resonance models*, where the Veneziano formula was generalized and systematized. Further discoveries were the connection of dual models to strings, the critical dimension (26), superconformal invariance, and the existence of strings with critical dimension 10. This first period of activity was ended by the discovery of QCD as the correct theory of the strong interaction, but those who had worked on string theory were left with the belief that they had discovered a rich physical theory that should have some application in nature.

During the succeeding 'quiet' period, a small group of theorists developed string theory as a unified theory of gravity and gauge interactions. String theories with spacetime supersymmetry were discovered during this period, and the argument made that superstrings were a finite theory of quantum gravity. Three discoveries in 1984 established that string theory was also a strong candidate for the unification of the Standard Model. These discoveries were the existence of anomaly-free chiral theories in ten dimensions, the heterotic string, and Calabi–Yau compactifaction. The ensuing period of activity was largely devoted to working out their consequences. The main difficulty that remained, and still remains, was the enormous number of potential vacua and an insufficient understanding of the dynamics that chooses among them.

The latest period of activity is marked by new dynamical tools. Again, many important results date from the preceding quiet period, but their importance and their relation to one another took time to emerge. The key steps were the systematic analysis of the consequences of supersymmetry in gauge theories and then in string theory. This led to many new discoveries: weak–strong duality in string theory, the unification of all string theories as limits of a single theory, a limit in which a new, eleventh, dimension appears, the significance of D-branes, and new understanding of the quantum mechanics of black holes.

The developments from 1984 on are covered in the literature and review articles below and in volume two, but the earlier period is somewhat neglected. Many of the key papers up to 1985 can be found in Schwarz (1985), which also includes summaries of the history. Chapters 1 and 2 of the reference section of GSW give a more detailed history of the dual model period, with extensive references. The review by Scherk (1975) also covers this period.

Chapter 1

There are now many treatments of the Standard Model; Quigg (1983) gives a survey and Cheng & Li (1991) a more detailed treatment, to name just two. Weinberg (1979) discusses the UV problem of quantum gravity in field theory. References for the various unifying ideas will be given in the appropriate chapters.

The classical motion of the string and light-cone quantization are discussed in Goddard, Goldstone, Rebbi, & Thorn (1973) and in the reviews by Rebbi (1974) and Scherk (1975). The first two of these include a careful treatment of Lorentz invariance. The heuristic argument that we have given is based on Brink & Nielsen (1973). Weinberg (1964) discusses the couplings of massless vector and tensor particles.

Chapter 2

In addition to the general references, Friedan, Martinec, & Shenker (1986) covers many of the subjects in this chapter. Belavin, Polyakov, & Zamolodchikov (1984) is a classic reference on CFT, but directed at the advanced topics to be treated in chapter 15. Ginsparg (1990) is a review of CFT; Gatto (1973) and Fradkin & Palchik (1978) treat conformal invariance in four-dimensional quantum field theory; although this is much smaller than the two-dimensional symmetry, many of the same ideas apply.

Chapter 3

Additional references on the Polyakov path integral are Polyakov (1981), Friedan (1982), and Alvarez (1983). The Faddeev–Popov procedure is treated in all modern field theory textbooks. The basic steps follow the original reference Faddeev & Popov (1967). The book by Birrell & Davies (1982) discusses Weyl anomalies in various dimensions (though it uses the common terminology *conformal anomaly* instead). Gravitational anomalies are discussed in Alvarez-Gaumé & Witten (1983).

An LSZ-like derivation of the vertex operators in light-cone gauge is in Mandelstam (1973); see also chapter 11 of GSW (1987). For calculation of beta functions in general nonlinear sigma models see Alvarez-Gaumé, Freedman, & Mukhi (1981) and Friedan (1985). A general reference on strings in background fields is Callan, Friedan, Martinec, & Perry (1985); reviews include Callan & Thorlacius (1989) and Tseytlin (1989). The linear dilaton theory is discussed as a CFT by Chodos & Thorn (1974) and as a string background by Myers (1987).

Chapter 4

Classic proofs of the OCQ no-ghost theorem are Brower (1972) and Goddard & Thorn (1972). The basic reference on quantization with constraints is Dirac (1950); a review is Henneaux (1985). The BRST symmetry in gauge theory is discussed in modern field theory texts. The structure of its cohomology is discussed in Kugo & Ojima (1979). The BRST operator for field theory is derived in Kato & Ogawa (1983); our proof of the no-ghost theorem is loosely patterned on the one in that paper. Other proofs in the BRST context are Freeman & Olive (1986), Frenkel, Garland, & Zuckerman (1986), and Thorn (1987).

Chapter 5

D'Hoker & Phong (1988) is an extensive review of the subjects in this and subsequent chapters. The measure on moduli space is obtained in Moore & Nelson (1986) and D'Hoker & Phong (1986); see also Alvarez (1983). The representation of the measure in terms of the transition functions is discussed in Martinec (1987) and Alvarez-Gaumé, Gomez, Moore, & Vafa (1988). General properties of the amplitude are also discussed in Friedan, Martinec, & Shenker (1986) and Mansfield (1987).

Chapter 6

Some of the physics of the tree-level amplitudes is discussed in Veneziano (1974). The appearance of gauge and gravitational interactions in the long-distance limit is developed in Neveu & Scherk (1972), Yoneya (1974), and Scherk & Schwarz (1974). Constraints on open string gauge groups are discussed in Marcus & Sagnotti (1982) and Schwarz (1982). For more on operator methods see GSW and the dual model references Schwarz (1973) and Mandelstam (1974). The relation of closed to open string tree-level amplitudes is given in Kawai, Lewellen, & Tye (1986).

Chapter 7

The two volumes by Mumford (1983) are a beautiful introduction to theta functions; the one-loop functions are discussed in the first half of volume one. Constraints from modular invariance on general conformal field theories are in Cardy (1986). The zero-point energy in quantum field theory is discussed in Coleman & Weinberg (1973). The cosmological constant problem is reviewed in Weinberg (1989).

Cancellation of divergences for open superstring gauge group $SO(32)$ is discussed in Green & Schwarz (1985). The analysis using the boundary state formalism is in Callan, Lovelace, Nappi, & Yost (1987) and Polchinski & Cai (1988).

Chapter 8

Kaluza–Klein theory is reviewed in Duff, Nilsson, & Pope (1986) and in the reprint volume Appelquist, Chodos, & Freund (1987). Giveon, Porrati, & Rabinovici (1994) review toroidal compactification and T-duality. Polchinski (1997) reviews T-duality of open and unoriented strings, and D-branes. Further references on toroidal compactification are Narain (1986) and Narain, Sarmadi, & Witten (1987).

Dixon, Harvey, Vafa, & Witten (1985) is an introduction to orbifolds. Twisted state vertex amplitudes are obtained in Dixon, Friedan, Martinec & Shenker (1987) and Hamidi & Vafa (1987). More extensive orbifold references are given with chapter 16. The space of $c = 1$ CFTs is described in Dijkgraaf, Verlinde, & Verlinde (1988) and Ginsparg (1988) and in the review Ginsparg (1990).

Chapter 9

Unitarity of tree-level amplitudes is discussed in the dual model reviews Schwarz (1973) and Mandelstam (1974) (these discussions are in OCQ, whereas section 9.1 gives the BRST version of the argument). One-loop amplitudes were originally constructed by imposing unitarity on the tree-level amplitudes; see Scherk (1975) for a review.

The moduli space of higher genus Riemann surfaces is reviewed in Martinec (1987) and D'Hoker & Phong (1988). Some more details of the Schottky construction can be found in Mandelstam (1974) and references therein. Sewing of CFTs has been discussed by many authors. The treatment here is based on Vafa (1987), Polchinski (1988), Sonoda (1988a,b,1989) and LeClair, Peskin, & Preitschopf (1989a,b). The extension to the open string is in Lewellen (1992). The graphical decomposition of moduli space and the construction of minimal area metrics are reviewed in Zwiebach (1993). A discussion of the combinatorics of unitarity in quantum field perturbation theory can be found in 't Hooft & Veltman (1973). The Fischler–Susskind mechanism is reviewed in Polchinski (1988). For a different approach to string unitarity, based on showing the equivalence between the light-cone and Polyakov amplitudes, see D'Hoker & Giddings (1987). The light-cone formalism is discussed in Mandelstam (1973) and GSW, chapter 11.

The form of open string field theory that we have described is from Witten (1986). String field theory is reviewed by Siegel (1988), Thorn (1989), and Zwiebach (1993). Borel summability of string perturbation theory is discussed in Gross & Periwal (1988). The $2g!$ behavior is discussed in Shenker (1991). High energy hard scattering is examined in Gross & Mende (1988) and Gross (1988). Another approach to high energy scattering is Amati, Ciafaloni, & Veneziano (1989). The spreading of the boosted string is described by Susskind (1993). The Hagedorn transition has been a source of much speculation since the original work of Hagedorn (1965). The discussion here is based in part on Atick & Witten (1988).

The review Polchinski (1996) covers many of the subjects in this chapter, and in particular gives an introduction to matrix models. Other matrix model reviews are Kazakov & Migdal (1988), Klebanov (1992), Martinec (1992), and an extensive review by Ginsparg & Moore (1993).

References

Alvarez, O. (1983). Theory of strings with boundaries. *Nuclear Physics*, **B216**, 125.

Alvarez-Gaumé, L., Freedman, D. Z., & Mukhi, S. (1981). The background field method and the ultraviolet structure of the supersymmetric nonlinear sigma model. *Annals of Physics*, **134**, 85.

Alvarez-Gaumé, L., Gomez, C., Moore, G., & Vafa, C. (1988). Strings in the operator formalism. *Nuclear Physics*, **B303**, 455.

Alvarez-Gaumé, L., & Vazquez-Mozo, M. A. (1992). Topics in string theory and quantum gravity. In *Les Houches Summer School on Gravitation and Quantizations*, eds. J. Zinn-Justin & B. Julia, pp. 481–636. Amsterdam: North-Holland. E-print hep-th/9212006.

Alvarez-Gaumé, L., & Witten, E. (1983). Gravitational anomalies. *Nuclear Physics*, **B234**, 269.

Amati, D, Ciafaloni, M., & Veneziano, G. (1989). Can spacetime be probed below the string size? *Physics Letters*, **B216**, 41.

Appelquist, T., Chodos, A., & Freund, P. G. O. (1987). *Modern Kaluza–Klein Theories*. Menlo Park: Addison Wesley.

Atick, J. J., & Witten, E. (1988). The Hagedorn transition and the number of degrees of freedom of string theory. *Nuclear Physics*, **B310**, 291.

Belavin, A. A., Polyakov, A. M., & Zamolodchikov, A. B. (1984). Infinite conformal symmetry in two-dimensional quantum field theory. *Nuclear Physics*, **B241**, 333.

Birrell, N. D., & Davies, P. C. W. (1982). *Quantum Fields in Curved Space*. Cambridge: Cambridge University Press.

Brink, L., & Nielsen, H. B. (1973). A simple physical interpretation of the critical dimension of spacetime in dual models. *Physics Letters*, **B45**, 332.

Brower, R. C. (1972). Spectrum-generating algebra and no-ghost theorem for the dual model. *Physical Review*, **D6**, 1655.

Buscher, T. (1987). A symmetry of the string background field equations. *Physics Letters*, **B194**, 59.

Buscher, T. (1988). Path integral derivation of quantum duality in nonlinear sigma models. *Physics Letters*, **B201**, 466.

Callan, C. G., Friedan, D., Martinec, E. J., & Perry, M. J. (1985). Strings in background fields. *Nuclear Physics*, **B262**, 593.

Callan, C. G., Lovelace, C., Nappi, C. R., & Yost, S. A. (1987). Adding holes and crosscaps to the superstring. *Nuclear Physics*, **B293**, 83.

Callan, C. G., & Thorlacius, L. (1989). Sigma models and string theory. In *Particles, Strings, and Supernovae, TASI 1988*, eds. A. Jevicki & C.-I. Tan, pp. 795–878. Singapore: World Scientific.

Cardy, J. L. (1986). Operator content of two-dimensional conformally invariant theories. *Nuclear Physics*, **B270**, 186.

Cheng, T.-P., & Li, L.-F. (1991). *Gauge Theory of Elementary Particle Physics*. Oxford: Oxford University Press.

Chodos, A., & Thorn, C. B. (1974). Making the massless string massive. *Nuclear Physics*, **B72**, 509.

Coleman, S., & Weinberg, E. (1973). Radiative corrections as the origin of spontaneous symmetry breaking. *Physical Review*, **D7**, 1888.

D'Hoker, E. (1993). TASI Lectures on critical string theory. In *Recent Directions in Particle Theory, TASI 1992*, eds. J. Harvey & J. Polchinski, pp. 1–98. Singapore: World Scientific.

D'Hoker, E., & Giddings, S. B. (1987). Unitarity of the closed bosonic Polyakov string. *Nuclear Physics*, **B291**, 90.

D'Hoker, E., & Phong, D. H. (1986). Multiloop amplitudes for the bosonic Polyakov string. *Nuclear Physics*, **B269**, 205.

D'Hoker, E., & Phong, D. H. (1988). The geometry of string perturbation theory. *Reviews of Modern Physics*, **60**, 917.

Dijkgraaf, R., Verlinde, E., & Verlinde, H. (1988). $c = 1$ conformal field theories on Riemann surfaces. *Communications in Mathematical Physics*, **115**, 649.

Dirac, P. A. M. (1950). Generalized Hamiltonian dynamics. *Canadian Journal of Mathematics*, **2**, 129.

Dixon, L., Friedan, D., Martinec, E., & Shenker, S. (1987). The conformal field theory of orbifolds. *Nuclear Physics*, **B282**, 13.

Dixon, L., Harvey, J. A., Vafa, C., & Witten, E. (1985). Strings on orbifolds. *Nuclear Physics*, **B261**, 678.

Duff, M. J., Nilsson, B. E. W., & Pope, C. N. (1986). Kaluza–Klein supergravity. *Physics Reports*, **130**, 1.

Faddeev, L. D., & Popov, V. N. (1967). Feynman diagrams for the Yang–Mills field. *Physics Letters*, **B25**, 29.

Fradkin, E. S., & Palchik, M. Ya. (1978). Recent developments in conformal invariant quantum field theory. *Physics Reports*, **44**, 249.

Freeman, M. D., & Olive, D. I. (1986). BRS cohomology in string theory and the no-ghost theorem. *Physics Letters*, **B175**, 151.

Frenkel, I. B., Garland, H., & Zuckerman, G. J. (1986). Semi-infinite cohomology and string theory. *Proceedings of the National Academy of Sciences*, **83**, 8442.

Friedan, D. (1982). Introduction to Polyakov's string theory. In *Recent Advances in Field Theory and Statistical Mechanics, Les Houches 1982*, eds. J.-B. Zuber & R. Stora, pp. 839–67. Amsterdam: North-Holland.

Friedan, D. (1985). Nonlinear models in $2 + \epsilon$ dimensions. *Annals of Physics*, **163**, 318.

Friedan, D., Martinec, E., & Shenker, S. (1986). Conformal invariance, supersymmetry, and string theory. *Nuclear Physics*, **B271**, 93.

Gatto, R. (1973). *Scale and Conformal Symmetry in Hadron Physics*. New York: Wiley.

Ginsparg, P. (1988). Curiosities at $c = 1$. *Nuclear Physics*, **B295**, 153.

Ginsparg, P. (1990). Applied conformal field theory. In *Fields, Strings, and Critical Phenomena, Les Houches 1988*, eds. E. Brezin & J. Zinn-Justin, pp. 1–168. Amsterdam: North-Holland.

Ginsparg, P., & Moore, G. (1993). Lectures on 2D gravity and 2D string theory. In *Recent Directions in Particle Theory, TASI 1992*, eds. J. Harvey & J. Polchinski, pp. 277–469. Singapore: World Scientific. E-print hep-th/9304011.

Giveon, A., Porrati, M., & Rabinovici, E. (1994). Target space duality in string theory. *Physics Reports*, **244**, 77. Eprint hep-th/9401139.

Goddard, P., Goldstone, J., Rebbi, C. B., & Thorn, C. (1973). Quantum dynamics of a massless relativistic string. *Nuclear Physics*, **B56**, 109.

Goddard, P., & Thorn, C. B. (1972). Compatibility of the dual Pomeron with unitarity and the absence of ghosts in the dual resonance model. *Physics Letters*, **B40**, 235.

Green, M. B., & Schwarz, J. H. (1985). Infinity cancellations in $SO(32)$ superstring theory. *Physics Letters*, **B151**, 21.

Green, M. B., Schwarz, J. H., & Witten, E. (1987). *Superstring Theory*, in two volumes. Cambridge: Cambridge University Press.

Gross, D. J. (1988). High energy symmetries of string theory. *Physical Review Letters*, **60**, 1229.

Gross. D. J., & Mende, P. F. (1988). String theory beyond the Planck scale. *Nuclear Physics*, **B303**, 407.

Gross, D. J., & Periwal, V. (1988). String perturbation theory diverges. *Physical Review Letters*, **60**, 2105.

Hagedorn, R. (1965). Statistical thermodynamics of strong interactions at high energies. *Nuovo Cimento Supplement*, **3**, 147.

Hamidi, S., & Vafa, C. (1987). Interactions on orbifolds. *Nuclear Physics*, **B279**, 465.

Henneaux, M. (1985). Hamiltonian form of the path integral for theories with a gauge freedom. *Physics Reports*, **126**, 1.

Kato, M., & Ogawa, K. (1983). Covariant quantization of string based on BRS invariance. *Nuclear Physics*, **B212**, 443.

Kawai, H., Lewellen, D. C., & Tye, S.-H. H. (1986). A relation between tree amplitudes of closed and open strings. *Nuclear Physics*, **B269**, 1.

Kazakov, V.A., & Migdal A. A. (1988). Recent progress in the theory of noncritical strings. *Nuclear Physics*, **B311**, 171.

Kiritsis, E. (1997). *Introduction to Superstring Theory*. Leuven: Leuven University Press, in press. E-print hep-th/9709062.

Klebanov, I. R. (1992). String theory in two dimensions. In *String Theory and Quantum Gravity, Trieste 1991*, eds. J. Harvey, *et. al.*, pp. 30–101. Singapore: World Scientific.

Kugo, T., & Ojima, I. (1979). Local covariant operator formulation of non-Abelian gauge theories and quark confinement problem. *Supplement to Progress in Theoretical Physics*, **66**, 1.

LeClair, A., Peskin, M. E., & Preitschopf, C. R. (1989a). String field theory on the conformal plane. 1. Kinematical principles. *Nuclear Physics*, **B317**, 411.

LeClair, A., Peskin, M. E., & Preitschopf, C. R. (1989b). String field theory on the conformal plane. 2. Generalized gluing. *Nuclear Physics*, **B317**, 464.

Lewellen, D. C. (1992). Sewing constraints for conformal field theories on surfaces with boundaries. *Nuclear Physics*, **B372**, 654.

Lovelace, C. (1971). Pomeron form factors and dual Regge cuts. *Physics Letters*, **B34**, 500.

Lust, D., & Theisen, S. (1989). *Lectures on String Theory*. Berlin: Springer-Verlag.

Mandelstam, S. (1973). Interacting-string picture of dual-resonance models. *Nuclear Physics*, **B64**, 205.

Mandelstam, S. (1974). Dual-resonance models. *Physics Reports*, **13**, 259.

Mansfield, P. (1987). Nilpotent BRST invariance of the interacting Polyakov string. *Nuclear Physics*, **B283**, 551.

Marcus, N., & Sagnotti, A. (1982). Tree-level constraints on gauge groups for type I superstrings. *Physics Letters*, **B119**, 97.

Martinec, E. (1987). Conformal field theory on a (super-)Riemann surface. *Nuclear Physics*, **B281**, 157.

Martinec, E. (1992). An introduction to 2d gravity and solvable string models. In *String Theory and Quantum Gravity, Trieste 1991*, eds. J. Harvey, *et. al.*, pp. 1–29. Singapore: World Scientific.

McClain, B., & Roth, B. D. B. (1987). Modular invariance for interacting bosonic strings at finite temperature. *Communications in Mathematical Physics*, **111**, 539.

Misner, C. W., Thorne, K. S., & Wheeler, J. A. (1973). *Gravitation*. San Francisco: Freeman.

Moore, G., & Nelson, P. (1986). Measure for moduli. *Nuclear Physics*, **B266**, 58.

Mumford, D. (1983). *Tata lectures on theta*, in three volumes. Boston: Birkhauser.

Myers, R. C. (1987). New dimensions for old strings. *Physics Letters*, **B199**, 371.

Narain, K. S. (1986). New heterotic string theories in uncompactified dimensions $D < 10$. *Physics Letters*, **B169**, 41.

Narain, K. S., Sarmadi, M. H., & Witten, E. (1987). A note on toroidal compactification of heterotic string theory. *Nuclear Physics*, **B279**, 369.

Neveu, A., & Scherk, J. (1972). Connection between Yang–Mills fields and dual models. *Nuclear Physics*, **B36**, 155.

Ooguri, H., & Zin, Y. (1997). Lectures on perturbative string theories. In *Fields, Strings, and Duality, TASI 1996*, eds. C. Efthimiou & B. Greene, pp. 5–82. Singapore: World Scientific. E-print hep-th/9612254.

Peskin, M. (1987). Introduction to string and superstring theory. In *From the Planck Scale to the Weak Scale, TASI 1986*, ed. H. Haber, pp. 277–408. Singapore: World Scientific.

Polchinski, J. (1986). Evaluation of the one loop string path integral. *Communications in Mathematical Physics*, **104**, 37.

Polchinski, J. (1988). Factorization of bosonic string amplitudes. *Nuclear Physics*, **B307**, 61.

Polchinski, J. (1996). What is string theory? In *Fluctuating Geometries in Statistical Mechanics and Field Theory, Les Houches 1994*, eds. F. David, P. Ginsparg, & J. Zinn-Justin, pp. 287–422. Amsterdam: North-Holland.

Polchinski, J. (1997). TASI lectures on D-branes. In *Fields, Strings, and Duality, TASI 1996*, eds. C. Efthimiou & B. Greene, pp. 293–356. Singapore: World Scientific. E-print hep-th/9611050.

Polchinski, J., & Cai, Y. (1988). Consistency of open superstring theories. *Nuclear Physics*, **B296**, 91.

Polyakov, A. M. (1981). Quantum geometry of bosonic strings. *Physics Letters*, **B103**, 207.

Polyakov, A. M. (1987). *Gauge Fields and Strings*. Chur, Switzerland: Harwood.

Quigg, C. (1983). *Gauge Theories of the Strong, Weak, and Electromagnetic Interactions*. Reading: Benjamin-Cummings.

Rebbi, C. (1974). Dual models and relativistic quantum strings. *Physics Reports*, **12**, 1.

Roček, M., & Verlinde, E. (1992). Duality, quotients and currents. *Nuclear Physics*, **B373**, 630. E-print hep-th/9110053.

Scherk, J. (1975). An introduction to the theory of dual models and strings. *Reviews of Modern Physics*, **47**, 123.

Scherk, J., & Schwarz, J. H. (1974). Dual models for non-hadrons. *Nuclear Physics*, **B81**, 118.

Schwarz, J. H. (1973). Dual resonance theory. *Physics Reports*, **8**, 269.

Schwarz, J. H. (1982). Superstring theory. *Physics Reports*, **89**, 223.

Schwarz, J. H. (1985). *Superstrings. The First 15 Years of Superstring Theory*, in two volumes. Singapore: World Scientific.

Shapiro, J. A. (1972). Loop graph in the dual-tube model. *Physical Review*, **D5**, 1945.

Shenker, S. (1991). The strength of nonperturbative effects in string theory. In *Random Surfaces and Quantum Gravity, Cargese 1990*, eds. O. Alvarez, E. Marinari, & P. Windey, pp. 191–200. New York: Plenum.

Siegel, W. (1988). *Introduction to String Field Theory*. Singapore: World Scientific.

Sonoda, H. (1988a). Sewing conformal field theories. 1. *Nuclear Physics*, **B311**, 401.

Sonoda, H. (1988b). Sewing conformal field theories. 2. *Nuclear Physics*, **B311**, 417.

Sonoda, H. (1989). Hermiticity and CPT in string theory. *Nuclear Physics*, **B326**, 135.

Susskind, L. (1993). String theory and the principle of black hole complementarity. *Physical Review Letters*, **71**, 2367. E-print hep-th/9307168.

't Hooft, G., & Veltman, M. (1973). Diagrammar. CERN report, reprinted in *Under the Spell of the Gauge Principle*, collected work of G. 't Hooft, pp. 28–173. Singapore: World Scientific, 1994.

Thorn, C. B. (1987). A detailed study of the physical state conditions in covariantly quantized string theories. *Nuclear Physics*, **B286**, 61.

Thorn, C. B. (1989). String field theory. *Physics Reports*, **175**, 1.

Tseytlin, A. A. (1989). Sigma model approach to string theory. *International Journal of Modern Physics*, **A4**, 1257.

Vafa, C. (1987). Conformal field theory and punctured surfaces. *Physics Letters*, **B199**, 195.

Veneziano, G. (1974). An introduction to dual models of strong interactions and their physical motivations. *Physics Reports*, **9**, 199.

Weinberg, S. (1964). Photons and gravitons in S matrix theory. *Physical Review*, **135**, B1049.

Weinberg, S. (1979). Ultraviolet divergences in quantum theories of gravitation. In *General Relativity. An Einstein Centenary Survey*, eds. S. W. Hawking & W. Israel, pp. 790–831. Cambridge: Cambridge University Press.

Weinberg, S. (1989). The cosmological constant problem. *Reviews of Modern Physics*, **61**, 1.

Witten, E. (1986). Noncommutative geometry and string field theory. *Nuclear Physics*, **B268**, 253.

Yoneya, T. (1974). Connection of dual models to electrodynamics and gravidynamics. *Progress of Theoretical Physics*, **51**, 1907.

Zwiebach, B. (1993). Closed string field theory: quantum action and the Batalin-Vilkovisky master equation. *Nuclear Physics*, **B390**, 33. E-print hep-th/9206084.

Glossary

(0, 2) compactification a generic heterotic string vacuum having $d = 4$, $N = 1$ supersymmetry. The world-sheet CFT has $N = 2$ right-moving superconformal invariance.

(2, 0) theory in current usage, this refers to a family of nontrivial fixed point theories with tensionless strings and $d = 6$ (2, 0) supersymmetry. These arise on coincident M5-branes and IIA NS5-branes, and on the IIB theory at an A–D–E singularity.

(2, 2) compactification one of a special subset of $d = 4$, $N = 1$ heterotic string vacua, which includes the Calabi–Yau compactifications. The world-sheet CFT has both right-moving and left-moving $N = 2$ superconformal invariance. In the type II string theories, these CFTs give vacua with $d = 4$, $N = 2$ supersymmetry.

A–D–E singularity a singularity of a four-(real)-dimensional complex manifold, resulting from the collapse of one or more two-spheres to zero volume. The terminology A–D–E refers to the Dynkin diagrams of the simply-laced Lie algebras, which describe the intersection numbers of the collapsed spheres.

Abelian differential a globally defined holomorphic (1,0)-form on a Riemann surface.

abstruse identity one of a set of quartic theta function identities due to Jacobi, it implies the degeneracy of bosons and fermions in GSO-projected string theories as required by supersymmetry.

affine Lie algebra see *current algebra.*

anomaly the violation of a classical symmetry by quantum effects. A *gravitational anomaly* is an anomaly in coordinate invariance. A *global anomaly* is an anomaly in a large symmetry transformation (one not continuously connected to the identity).

anomaly polynomial a formal $(d + 2)$-form in d-dimensions, which encodes the gauge and gravitational anomalies.

asymptotically locally Euclidean (ALE) space a space which at long distance approaches flat Euclidean space identified under a discrete group. This is the geometry in the neighborhood of an orbifold fixed point (or blown-up fixed point).

Atiyah–Drinfeld–Hitchin–Manin (ADHM) construction a method for the construction of all Yang–Mills field configurations having self-dual field strength.

auxiliary field a nonpropagating field, one whose field equation is algebraic rather than differential. In many supersymmetric theories, the transformations can be simplified by introducing such fields.

axion a Goldstone boson associated with spontaneously broken PQ symmetry. The *model-independent axion* appears in every perturbative string theory, and is closely related to the graviton and dilaton.

***bc* CFT** a free CFT of anticommuting fields with an action of first order in derivatives. There is a family of such CFTs, parameterized by the weight $h_b = 1 - h_c$. For $h_b = 2$ this CFT describes the Faddeev–Popov ghosts associated with conformal invariance.

***βγ* CFT** a free CFT of commuting fields with an action of first order in derivatives. There is a family of such CFTs, parameterized by the weight $h_\beta = 1 - h_\gamma$. For $h_\beta = \frac{3}{2}$ this CFT describes the Faddeev–Popov ghosts associated with superconformal invariance.

Batalin–Vilkovisky formalism an extension of the BRST formalism, for quantizing more general theories with constraints. This has been useful in string field theory.

Becchi–Rouet–Stora–Tyutin (BRST) invariance a nilpotent symmetry of Faddeev–Popov gauge-fixed theories, which encodes the information contained in the original gauge symmetry.

Beltrami differential the derivative with respect to the moduli of the complex structure of a Riemann surface.

Berezin integration a linear operation taking functions of Grassmann variables to complex numbers, with many of the key properties of ordinary integration.

beta function 1. in quantum field theory, the derivative of the effective strength of an interaction with respect to length scale; 2. a special function involving a ratio of gamma functions, which appears in the Veneziano amplitude.

Betti numbers the number of nontrivial p-forms in de Rham cohomology, denoted B_p.

black hole entropy a quantity S proportional to the area of the horizon of a black hole, $S = 2\pi A/\kappa^2$. This has the properties of a thermodynamic entropy: it is nondecreasing in classical general relativity, and the sum of the black hole entropy and the ordinary entropy is nondecreasing even with the inclusion of Hawking radiation. To find a statistical mechanical derivation of this entropy has been a major goal, partly realized in recent work.

black hole evaporation the emission of thermal (Hawking) radiation by a black hole, due to pair production near the horizon.

black hole information paradox a conflict between quantum mechanics and general relativity. Information falling into a black hole is lost and does not reappear when the black hole evaporates; this is inconsistent with ordinary quantum mechanical evolution. It apparently requires either a significant modification of quantum mechanics, or a significant breakdown of the usual understanding of locality.

black *p*-brane a p-dimensional extended object with an event horizon: a space that is translationally invariant in p directions and has a black hole geometry in the remaining directions.

blow up to deform a singular manifold into a smooth manifold.

Bogomolnyi–Prasad–Sommerfield (BPS) state a state that is invariant under a nontrivial subalgebra of the full supersymmetry algebra. Such states always carry conserved charges, and the supersymmetry algebra determines the mass of the state exactly in terms of its charges. BPS states lie in smaller supersymmetry representations than non-BPS states, so-called *short representations*. When there are short representations of different sizes, one also distinguishes *ultrashort representations*, which are the smallest possible (generally their dimension is the square root of the non-BPS dimension).

Borel summation a method of defining the sum of a divergent series. This has been used as a means of studying nonperturbative effects in field and string theories, but it should be understood that most nonperturbative effects are not usefully studied in terms of the perturbation series.

Born–Infeld action a generalization of the usual gauge field action which is nonpolynomial in the gauge field strength. This was originally proposed as a possible short-distance modification of electromagnetism. It arises as the low energy effective action of the gauge fields on D-branes.

bosonization the exact equivalence of a theory of fermionic fields and a theory of bosonic fields, possible in two dimensions. The boson is a fermion–antifermion pair; the fermion is a coherent state of bosons.

***c*-map** a method for constructing the hypermultiplet moduli space of a type II string theory compactified on a Calabi–Yau three-fold from the vector multiplet moduli space of the other type II theory on the same three-fold.

***c*-theorem** the existence, in unitary CFTs in two dimensions, of a positive quantity c that is monotonically nonincreasing with increasing length scale, and which at fixed points is stationary and equal to the central charge. This is a strong constraint on the global form of the renormalization group flow; no simple analog seems to exist in higher dimensions. Also known as the *Zamolodchikov c-theorem*.

***CPT* symmetry** the combined operation of parity-reversal, time-reversal, and charge conjugation, which is a symmetry of all Lorentz-invariant local quantum field theories.

Calabi–Yau manifold a Kähler manifold with vanishing first Chern class. A *Calabi–Yau n-fold* has $2n$ real $= n$ complex coordinates. Yau's theorem guarantees the existence of a Ricci-flat metric of $SU(n)$ holonomy.

canceled propagator argument a general principle implying, under broad conditions, the vanishing of surface terms on the moduli space of Riemann surfaces and therefore the decoupling of unphysical states in string amplitudes. Such amplitudes are defined by analytic continuation from a regime where the integrand falls rapidly at the boundary and the surface term is identically zero; its continuation is therefore also identically zero.

Casimir energy a shift in the ground state energy of a quantum field theory due to boundary conditions on the fields.

center-of-mass mode the zeroth spatial Fourier component of a quantum field.

central charge an operator (which might be a constant) that appears on the right-hand side of a Lie algebra and commutes with all operators in the algebra. Prominent examples include the constant term in the Virasoro algebra and the charges appearing on the right-hand sides of many supersymmetry algebras.

Chan–Paton degrees of freedom degrees of freedom localized at the endpoints of open strings. These are now interpreted as designating the D-brane on which the string ends.

Chan–Paton factor the vertex operator factor for the state of the Chan–Paton degrees of freedom.

Chern–Simons term a term in the action which involves p-form potentials as well as field strengths. Such a term is gauge-invariant as a consequence of the Bianchi identity and/or the modification of the p-form gauge transformation. These terms usually have a close connection to topology and to anomalies.

chiral 1. acting in a parity asymmetric fashion; see *chiral multiplet, chiral symmetry, chiral theory, chirality, extended chiral algebra*; 2. invariant under part of the supersymmetry algebra; see *chiral field, chiral multiplet, chiral primary, chiral ring.*

chiral field in supersymmetry, a local operator that is invariant under part of the algebra: the operator analog of a *BPS state.*

chiral multiplet the multiplet of $d = 4$, $N = 1$ supersymmetry with two real scalars. The quarks and leptons are contained in such multiplets. This multiplet is connected to both senses of *chiral:* it contains a fermion with chiral couplings, and the integral of the associated superfield is invariant under half of the supersymmetry algebra.

chiral primary in an $N = 2$ SCFT, a primary field that is also annihilated by one of the rigid supersymmetries $G^{\pm}_{-1/2}$.

chiral ring the closed OPE algebra of chiral fields.

chiral symmetry a symmetry whose action on spinor fields is not parity-symmetric.

chiral theory a gauge theory in which the gauge couplings are not parity-symmetric.

chirality in $d = 2k$ dimensions, the eigenvalue of the operator Γ^d which anti-commutes with the Γ^μ. This eigenvalue distinguishes the two Weyl representations, which are related to one another by parity.

Christoffel connection in general relativity, the connection that is constructed from the metric.

critical behavior the behavior of a quantum field theory at an IR fixed point with massless fields, and the approach to this behavior.

closed string a string with the topology of a circle.

cocycle in a vertex operator, an operator-valued phase factor which multiplies the creation–annihilation normal ordered exponential. This in needed in some cases in order to give the operator the correct commuting or anticommuting property.

coefficient functions the position-dependent coefficients of local operators appearing in the expansion of an operator product.

cohomology in any vector space with a nilpotent operator Q (one such that $Q^2 = 0$), the kernel of Q modulo the image of Q. That is, the space of *closed* states (those annihilated by Q) with the *exact* states (those of the form $Q\psi$) defined to be equivalent to zero. *De Rham cohomology* is the cohomology of the exterior derivative d acting on differential forms. On a complex manifold, *Dolbeault cohomology* is the cohomology of ∂ and $\bar{\partial}$ (the $(1,0)$ and $(0,1)$ parts of d) on (p,q)-forms. *Homology* is the cohomology of the boundary operator. *BRST cohomology* is the cohomology of the BRST operator, and defines the physical space of a gauge-invariant theory.

Coleman–Weinberg formula the expression for the vacuum energy density of a free quantum field, from the renormalized sum of the zero-point energies of its modes.

collapsing cycle a cycle whose volume vanishes in a limit, usually giving rise to a singular manifold.

collective coordinate in quantizing a soliton or other extended object, the degrees of freedom corresponding to its position or configuration.

compact CFT a CFT in which the number of states with energy less than any given value is finite. This is defined by analogy with the spectrum of a differential operator on a compact space.

compactification scale the characteristic mass scale of states whose wavefunctions have a nontrivial dependence on the compact dimensions.

compactify to consider a field theory or string theory in a spacetime, some of whose spatial dimensions are compact.

complex manifold a manifold with an assigned system of complex coordinates, modulo holomorphic reparameterizations of these coordinates.

complex structure an equivalence class of complex coordinates. A given differentiable manifold may have many inequivalent complex structures.

complex structure moduli the moduli that parameterize the inequivalent complex structures on a manifold. In compactification on a Calabi–Yau 3-fold, these are associated with (2, 1)-forms.

conformal block in CFT, the contribution of a single conformal family to a sum over states.

conformal bootstrap the partially successful program to construct all CFTs by using only symmetry and consistency conditions.

conformal family the set of states obtained by acting on a highest weight state with Virasoro raising generators in all inequivalent ways; or, the corresponding set of local operators. A *degenerate conformal family* contains null states, which are orthogonal to all states in the family.

conformal field theory (CFT) a conformally invariant quantum field theory.

conformal gauge a choice of coordinates in two dimensions, such that the metric is proportional to the unit metric.

conformal Killing vector a globally defined infinitesimal diff×Weyl transformation that leaves the metric invariant.

conformal transformation a mapping of Euclidean or Minkowski space to itself that leaves the flat metric invariant up to a position-dependent rescaling; equivalently, the subgroup of diff × Weyl that leaves invariant the flat metric. In $d \geq 3$ dimensions this has $\frac{1}{2}(d+1)(d+2)$ parameters. In two dimensions it is the set of all holomorphic maps. Finite transformations require the inclusion of points at infinity, as in the case of the Möbius transformations of the sphere.
This usage has become standard in string theory and quantum field theory, but in general relativity *conformal transformation* is defined to be any position-dependent rescaling of the metric, now called a *Weyl transformation* in string theory.

conifold a Calabi–Yau manifold with a singular complex structure, corresponding to the collapse of a three-cycle. The string theory on this space is singular; this is now understood to be due to the quantum effects of a massless 3-brane wrapped on the cycle.

conifold transition a change of topology due to condensation of massless 3-brane fields. Under appropriate conditions, the potential for the massless 3-brane fields on a conifold with multiple collapsed cycles has a flat direction for these fields; this corresponds to a change of topology, blowing up a 2-cycle rather than the collapsed 3-cycle.

constraint a symmetry generator whose matrix elements are required to vanish in physical states, either in the BRST or OCQ sense. These can usually be understood as arising from a gauge symmetry (for so-called *first class constraints*, which are all that we consider), and consist of those gauge symmetry generators that do not vanish by the equations of motion.

coset CFT a CFT constructed as one of the factors of a known CFT, when the energy-momentum tensor of the latter can be written as a sum of commuting pieces. In the classic example the full CFT G is a current algebra, as is one of the factors (H). This can also be thought of as gauging the symmetry H.

cosmological constant the energy density of the vacuum. In a nonsupersymmetric quantum theory (including one with spontaneously broken supersymmetry), there are many effects that give rise to such an energy density. The *cosmological constant problem* is the problem that the cosmological constant in nature is many orders of magnitude smaller than known nonzero effects.

critical dimension the dimension in which a perturbative string theory is consistent in flat Minkowski spacetime; the value is 26 for the bosonic string and 10 for the supersymmetric string theories.

current algebra in quantum field theory, the algebra of the currents associated with a continuous symmetry group g (or of their Fourier modes). As used here, it is the specific algebra that occurs in two-dimensional CFTs, with the energy-momentum tensor defined to be of Sugawara form. The terms *affine Lie algebra* and *affine Kac–Moody algebra* are also used for this algebra, though like *current algebra* they both have broader definitions as well. The term *affine* refers to the c-number (Schwinger) term. An *untwisted* current algebra is the algebra of periodic currents, with integer modes. An algebra can be *twisted* by any automorphism of g.

cycle a topologically nontrivial submanifold (in the sense of homology); a *p-cycle* is p-dimensional. The *A-* and *B-cycles* are a standard basis for the nontrivial one-cycles on a Riemann surface.

D-brane in the type I, IIA, and IIB string theories, a dynamical object on which strings can end. The term is a contraction of *Dirichlet brane*. The coordinates of the attached strings satisfy Dirichlet boundary conditions in the directions normal to the brane and Neumann conditions in the directions tangent to the brane. A *Dp-brane* is p-dimensional, with p taking any even value in the IIA theory, any odd value in the IIB theory, and the values 1, 5, and 9 in the type I theory; a D9-brane fills space and so corresponds to an ordinary Neumann boundary condition. The Dp-brane is a source for the $(p+1)$-form R–R gauge field. The mass or tension of a D-brane is intermediate between that of an ordinary quantum or a fundamental string and that of a *soliton*. The low energy fluctations of D-branes are described by supersymmetric gauge theory, which is non-Abelian for coincident branes.

D-instanton an object localized in (Euclidean) time as well as space, defined by Dirichlet conditions on all coordinates of attached strings. This is similar to

a field-theoretic instanton, corresponding to a tunneling process that changes the value of an R–R field strength. More generally, the $(p+1)$-dimensional world-volume of a Dp-brane, when localized in time and wrapped on a $(p+1)$-cycle in space, has similar effects.

D-string a D1-brane, in the type I and IIB string theories.

***D*-term** 1. in gauge theories with four or eight supersymmetries, the auxiliary field in the gauge multiplet; 2 the potential term proportional to the square of this auxiliary field, which depends only on the gauge couplings and, in the $U(1)$ case, the value of the Fayet–Iliopoulos parameter.

Del Guidice–Di Vecchia–Fubini (DDF) operators operators satisfying an oscillator algebra, which create a complete set of physical states in OCQ.

descendant a state obtained by acting on a highest weight state with Virasoro raising generators.

diagonal modular invariant a modular-invariant CFT formed by imposing common boundary conditions on the left- and right-moving fields.

diff invariance general coordinate (reparameterization) invariance, usually applied to the world-sheet coordinates.

dilaton the massless scalar with gravitational-strength couplings, found in all perturbative string theories. An exactly massless dilaton would violate limits on nongravitational interactions, but a mass for the dilaton is not forbidden by any symmetry and so dynamical effects will generate one in vacua with broken supersymmetry (the same holds for other moduli). The string coupling constant is determined by the value of the dilaton field.

dimensional reduction in the simplest cases, toroidal compactification retaining only the states of zero compact momentum. More generally (and less physically) the construction of a lower-dimensional field theory by requiring all fields to be invariant under a set of symmetries; this may have no interpretation in terms of compactification.

Dirac quantization condition for an electric and a magnetic charge, the condition that the product be quantized, $\mu_e \mu'_m = 2\pi n$. For two dyons, which have both charges, the condition is $\mu_e \mu'_m - \mu_m \mu'_e = 2\pi n$. These conditions generalize to objects of dimension p and $d-p-4$ in d dimensions, where one is the source of a $(p+2)$-form field strength and the other of the Poincaré dual $(d-p+2)$-form field strength.

Dirac spinor the unique irreducible representation of the algebra of Dirac matrices, which is also known as a Clifford algebra. This is also a representation of the Lorentz group; in even dimensions it is reducible to two Weyl representations of the Lorentz group. The Dirac spinor is complex; in certain dimensions a further Majorana (reality) condition is compatible with Lorentz invariance.

Dirichlet boundary condition the condition that the value of a field be fixed at a boundary. This is the relevant usage in string theory, but in other contexts only the tangent derivative need be fixed.

discrete torsion in forming a *twisted CFT*, a change in the phases of the path integral sectors and therefore in the projection on the Hilbert space.

doubling trick the representation of holomorphic and antiholomorphic fields on a manifold with boundary by holomorphic fields alone, on the doubled copy of the manifold obtained by reflecting through the boundary.

dual resonance model a phenomenological model of the strong interaction, which developed into string theory. The *dual* refers here to *world-sheet duality*.

duality the equivalence of seemingly distinct physical systems. Such an equivalence often arises when a single quantum theory has distinct classical limits. One classic example is *particle–wave duality*, wherein a quantum field theory has one limit described by classical field theory and another described by classical particle mechanics. Another is the high-temperature–low-temperature duality of the Ising model. Here, low temperature is the statistical mechanical analog of the classical limit, the Boltzmann sum being dominated by the configurations of lowest energy. See *Montonen–Olive duality*, *S-duality*, *string–string duality*, *T-duality*, *U-duality*, *world-sheet duality*.

effective field theory the description of a physical system below a given energy scale (or equivalently, above a given length scale).

Einstein metric the metric whose leading low energy action is the *Hilbert action*

$$\frac{1}{2\kappa^2}\int d^d x (-G)^{1/2} R \ ;$$

this is independent of other fields. Here κ is the gravitational coupling, related to the Planck length by $\kappa = (8\pi)^{1/2} L_P$. This metric is related to other metrics such as the *sigma-model metric* by a field-dependent Weyl transformation. The existence of distinct metrics would appear to violate the equivalence principle, but when the dilaton and other moduli are massive the distinction disappears.

electroweak scale the mass scale of electroweak symmetry breaking, roughly 10^2 GeV.

enhanced gauge symmetry a gauge symmetry appearing at special points in moduli space, which is not evident in the original formulation of a theory. The classic examples are the gauge symmetries that arise at special radii of toroidal compactification, whose gauge bosons are winding states. Many other mechanisms are now known: D-branes and black p-branes wrapped on collapsing cycles, F-strings or D-strings stretched between various branes in the limit that the latter become coincident, and the gauge symmetry appearing on a zero size $SO(32)$ instanton.

Euclidean having a metric of strictly positive signature. The original connotation that the metric be flat is somewhat disregarded; thus one refers to *Euclidean quantum gravity*, a conjectured analytic continuation of the Minkowskian theory. For the metric itself, the term *Riemannian* for a curved metric of Euclidean signature is more precise.

Euclidean adjoint in a Euclidean quantum theory, the Hermitean adjoint combined with time-reversal. The latter operation undoes the time-reversing effect of the adjoint, so that the combined operation is local.

Euler number the topological invariant

$$\chi = \sum_{p=0}^{d} (-1)^p B_p \ ,$$

where B_p is the pth Betti number. It is equal to $2(1-g)$ for a Riemann surface of genus g. More properly, the *Euler characteristic.*

expectation value a path integral with specified insertions. This is the term that we have chosen to use, but *correlation function* and *correlator* are also in common usage.

extended chiral algebra the full set of holomorphic operators in a CFT.

extended supersymmetry a supersymmetry algebra in which the supercharges comprise more than one copy of the smallest spinor representation for the given spacetime dimension.

F-term 1. the auxiliary field in the chiral multiplet of $d = 4$, $N = 1$ supersymmetry; 2. the potential term proportional to the square of this field.

F theory 1. a systematic description of IIB superstring states with nontrivial dilaton and R–R scalar backgrounds, which relates these fields to the modulus τ of an auxiliary two-torus; 2. a conjectured twelve-dimensional quantum theory underlying the IIB string. The name is inspired by M-theory, with F for father.

Faddeev–Popov determinant the Jacobian determinant arising from the reduction of a gauge-invariant functional integral to an integral over a gauge slice.

Faddeev–Popov ghosts the wrong-statistics quantum fields used to give a functional integral representation of the Faddeev–Popov determinant.

Fayet–Iliopoulos term in $U(1)$ gauge theories with four or eight supersymmetries, a term in the action which is linear in the auxiliary D-term field.

Feigin–Fuchs representation a representation of minimal model expectation values in terms of free fields.

fibration a space which is locally the product of a fiber F and a base B. The geometry of the fiber varies as one moves over the base, and may become singular. Typical fibers are tori and the $K3$ manifold.

first Chern class on a complex manifold, the Dolbeault cohomology class of the Ricci form $R_{i\bar{\jmath}} dz^i d\bar{z}^{\bar{\jmath}}$.

first-quantized description the representation of a quantized particle theory as a sum over particle paths, or of a string theory as a sum over world-sheets. *Second-quantized* refers to the representation in terms of a functional integral over ordinary or string fields. The term second-quantized implies the reinterpretation of the first-quantized *wavefunction* as a *field operator.* This terminology is in common usage, but it has been argued that it is unsatisfactory, in that it implies a deep principle where none may exist. Since the sum over world-sheets is itself a quantum field theory, one can equally well call it second-quantized, in which case string field theory is *third-quantized.* Third quantization of an ordinary field theory would describe operators that create and destroy universes, a concept which may or may not be useful.

Fischler–Susskind mechanism the cancellation of divergences and anomalies in the world-sheet quantum field theory against divergences and anomalies from integration over small topological features at higher orders of string perturbation theory. This is needed for the consistency of string perturbation theory in a quantum-corrected background.

fixed point 1. (*in geometry*) a point left invariant by a given symmetry transformation. This becomes a boundary point or a singularity if the space is identified under the transformation; 2. (*in quantum field theory*) a quantum theory whose physics is independent of length scale (scale-invariant). Usually such a theory is conformally invariant as well. A *UV fixed point* is the theory governing the short-distance physics of a quantum field theory; an *IR fixed point* is the theory governing the long-distance physics of a quantum field theory. A *trivial IR fixed point* has no massless fields. A *nontrivial IR fixed point* has massless fields with nonvanishing interactions. A theory whose IR limit is a massless free field theory is therefore described by neither of these terms; it is a *noninteracting IR fixed point.*

flat direction in scalar field space, a line of degenerate local minima. The field corresponding to this direction is a *modulus.*

flop a change of topology which can occur in weakly coupled string theory, where a two-cycle collapses and then a different two-cycle blows up.

fractional charge an unconfined particle whose electric charge is not a multiple of that of the electron. These exist in most $d = 4$ string theories, although in many cases all are superheavy.

fractional string theory a proposed generalization of string theory having constraints whose spin is not a multiple of $\frac{1}{2}$. No complete construction exists.

Fuchsian group a discrete subgroup Γ of the $SL(2,\mathbf{R})$ Möbius transformations of the complex upper-half-plane H, with additional conditions such that H/Γ is a manifold and in particular a Riemann surface.

functional integral in our usage, synonymous with *path integral.*

fundamental region in relation to a coset space M/Γ where Γ is a discrete group, a region F such that every point in M is identified with exactly one point in the interior of F or with one or more points on the boundary of F.

fundamental string (F-string) the original string whose quantization defines a weakly coupled string theory, as distinguished from D-strings and solitonic strings.

fusion rule the specification of which conformal families appear in the operator product of any two primary fields in a given CFT.

gauge-fixing the reduction of a redundant (gauge-invariant) description of a quantum theory to a description with a single representative from each equivalence class.

gaugino a spin-$\frac{1}{2}$ fermion in the same supersymmetry multiplet as a gauge boson.

gaugino condensation a strong coupling effect where a product of gaugino fields acquires a vacuum expectation value. This generally breaks a chiral symmetry but does not directly break supersymmetry; however, in combination with other fields it often induces supersymmetry breaking.

generation a family of quarks and leptons, described by spinor fields in a chiral but anomaly-free set of $SU(3) \times SU(2) \times U(1)$ representations. In $SU(5)$ grand unification these become a $\mathbf{5} + \overline{\mathbf{10}}$, in $SO(10)$ they are contained in a **16**, and in E_6 they are contained in a **27**. An *antigeneration* is the conjugate representation; the distinction between generation and antigeneration is a matter of convention.

genus the number g of handles on a closed oriented Riemann surface: $g = 0$ is a sphere, $g = 1$ is a torus, and so on.

Gepner model a string model based on $N = 2$ *minimal model* CFTs.

ghosts see *Faddeev–Popov ghosts.*

Gliozzi–Scherk–Olive (GSO) projection a construction of modular-invariant string theories by summing over R and NS boundary conditions on the fermion fields and projecting onto states of definite world-sheet fermion number. In supersymmetric string theories there are independent GSO projections on the left-movers and right-movers. The diagonal projection, which acts simultaneously on both sides, produces a nonsupersymmetric theory.

goldstino the massless spin-$\frac{1}{2}$ Goldstone fermion associated with spontaneously broken supersymmetry. In supergravity it combines with the gravitino to form a massive fermion.

Goldstone boson the massless scalar corresponding to fluctuations of the direction of spontaneous symmetry breaking.

grand unification the unification of the $SU(3) \times SU(2) \times U(1)$ gauge symmetries in a simple group.

grand unification scale the mass scale of spontaneous breaking of the grand unified group. Proton stability and the unification of the couplings require that it be within two or three orders of magnitude of the gravitational scale.

Grassmann variable the elements θ_i of an algebra with the relation $\theta_i\theta_j = -\theta_j\theta_i$. These are used to give a path integral representation of fermionic fields, and to define *superspace.* They are also called *anticommuting c-numbers.*

gravitational scale the mass scale at which the dimensionless gravitational coupling becomes of order 1, $m_{\text{grav}} = \kappa^{-1} = 2.4 \times 10^{18}$ GeV; this is $(8\pi)^{-1/2}$ times the Planck mass.

gravitino a spin-$\frac{3}{2}$ fermion in the same supersymmetry multiplet as the graviton.

Green–Schwarz mechanism the cancellation of an anomaly by the modified transformation law of a p-form potential in a *Chern–Simons term.*

Green–Schwarz superstring a manifestly supersymmetric formulation of the supersymmetric string theories, with a spacetime-fermionic gauge invariance known as κ *symmetry.* There is no simple covariant gauge fixing.

***H*-monopole** a monopole carrying the magnetic charge of the antisymmetric tensor gauge field $B_{\mu n}$.

heterotic 5-brane the 5-brane carrying the magnetic charge of the massless heterotic string 2-form potential. It is obtained as the limit of a zero size instanton in the heterotic string gauge fields. The instanton configuration is localized in four spatial dimensions, and is therefore a 5-brane in nine spatial dimensions.

heterotic string a string with different constraint algebras acting on the left- and right-moving fields. The case of phenomenological interest has a $(0, 1)$ superconformal constraint algebra, with spacetime supersymmetry acting only on the right-movers and with gauge group $E_8 \times E_8$ or $SO(32)$.

Hagedorn temperature the temperature at which the thermal partition function of free strings diverges, due to the exponential growth of the density of states of highly excited strings.

hidden sector the fields that couple to the Standard Model only through gravitational-strength interactions. In *hidden sector models,* these include the fields responsible for supersymmetry breaking.

highest weight state in CFT, a state annihilated by all Virasoro lowering operators, or more generally by all lowering operators in a given algebra.

Hodge number the number of nontrivial (p, q)-forms in Dolbeault cohomology, denoted $h^{p,q}$.

holographic principle the conjecture that the states of quantum gravity in d dimensions have a natural description in terms of a $(d-1)$-dimensional theory. This radical departure from local field theory was motivated by the black hole information problem, and has played a role in attempts to formulate M-theory.

holomorphic analytic, as used in the theory of complex variables. The Minkowskian continuation is *left-moving.* An *antiholomorphic* field is analytic in the conjugate variable, and its continuation is *right-moving.*

holomorphic quadratic differential a globally defined holomorphic $(2, 0)$-form on a Riemann surface.

holomorphic vector field a globally defined holomorphic $(-1,0)$-form on a Riemann surface.

holonomy consider the parallel transport of a vector around a closed loop on a d-dimensional manifold: it returns to an $O(n)$ rotation of its original value. The set of all rotations that are obtained in this way for a given manifold is a subgroup of $O(n)$; this is the *holonomy group*.

homology see *cohomology*.

hyperelliptic surface a Riemann surface with a $\mathbf{Z}_2$ symmetry. This can be represented as a two-sheeted cover of the sphere with branch cuts.

hyper-Kähler manifold a $4k$-dimensional manifold of holonomy $Sp(k) \subset SO(4k)$. This is the geometry of the moduli space of hypermultiplets in $d = 6$, $N = 1$ or $d = 4$, $N = 2$ supersymmetry, in the limit in which gravity decouples.

hypermultiplet in $d = 6$, $N = 1$ or $d = 4$, $N = 2$ supersymmetry, the multiplet whose bosonic content is four real massless scalars.

identify to define two points (or other objects) to be equivalent, thus producing a coset space.

infrared (IR) divergence a divergence arising from long distances in spacetime, usually signifying that one has calculated the wrong thing.

inheritance principle in twisted (orbifold) theories, the principle that the tree-level amplitudes of untwisted states are the same as in the untwisted theory.

insertion the integrand of a path integral, excluding the weight $\exp(iS)$ or $\exp(-S)$.

instanton in a Euclidean path integral, a nonconstant configuration that is a local but not a global minimum of the action. Such configurations are usually localized in spacetime, are usually topologically nontrivial, and are of interest when they give rise to effects such as tunneling that are not obtained from small fluctuations around a constant configuration. *Spacetime instantons* are instantons in the effective field theory in spacetime. *World-sheet instantons* are instantons in the world-sheet quantum field theory, and correspond to world-sheets wrapping around nontrivial two-cycles of spacetime.

intersection number the number of points at which a set of surfaces intersect, weighted by the orientation of the intersection.

irrelevant interaction an interaction whose dimensionless strength decreases with increasing length scale. In perturbation theory, this is equivalent to a *non-renormalizable* interaction.

K3 manifold the unique nontrivial Calabi–Yau manifold of four (real) dimensions. To be precise, it is topologically unique, but possesses complex structure and Kähler moduli. Its holonomy is $SU(2)$, so that half of the supersymmetries of a theory are broken upon compactification on K3.

Kac determinant the determinant of the matrix of inner products of states at a given L_0 level of a Verma module.

Kac–Moody algebra see *current algebra*.

Kähler form the $(1,1)$-form $G_{i\bar{j}}dz^i d\bar{z}^{\bar{j}}$, formed from a Hermitean metric on a complex manifold.

Kähler manifold a complex manifold of $U(n)$ holonomy in n complex dimensions.

Kähler moduli the moduli parameterizing the Kähler form.

Kähler potential the potential $K(z,\bar{z})$, in terms of which the metric of a Kähler manifold is determined, $G_{i\bar{j}} = \partial_i \partial_{\bar{j}} K$. This is not globally defined, being determined only up to a *Kähler transformation* $K(z,\bar{z}) \to K(z,\bar{z}) + f(z) + f(z)^*$.

Kaluza–Klein gauge field in a compactified theory, a gauge field originating from the metric of the higher-dimensional theory. The gauge group is the isometry group of the compact space.

Kaluza–Klein monopole a monopole carrying the magnetic charge of a $U(1)$ Kaluza–Klein gauge symmetry. The monopole configuration is the smooth *Taub–NUT* spacetime. It is localized in three spatial dimensions, and is therefore a 6-brane in nine spatial dimensions.

Kaluza–Klein states states with nonzero momentum in a compact spatial direction.

Knizhnik–Zamolodchikov (KZ) equation the differential equation determining the expectation values of the primary fields of a current algebra.

Landau–Ginzburg model a scalar field theory which has a nontrivial IR fixed point when the potential is appropriately tuned. In particular, this gives a Lagrangian representation of the minimal model CFTs.

large coordinate transformation a coordinate transformation that is not continuously connected to the identity.

lattice the set Γ of integer linear combinations of n linearly independent basis vectors in n dimensions. Given a Euclidean or Lorentzian metric, an *even lattice* is one whose points have even length-squared. The dual lattice Γ^* is the set of points v such that $v \cdot w \in \mathbf{Z}$ for all $w \in \Gamma$. The *root lattice* is the set of integer linear combinations of the roots of a Lie algebra.

level 1. the quantized c-number term in a current algebra, also known as the *Schwinger term*; 2. the total oscillator excitation number in free field theory; 3. in a conformal family, the difference between the L_0 eigenvalue of a given state and that of the highest weight state.

level-matching the modular invariance condition that $L_0 - \tilde{L}_0 \in \mathbf{Z}$.

Lie algebra an algebra with an antisymmetric product that satisfies the Jacobi identity. A *simple* Lie algebra has no subalgebra that commutes with its complement. A *simply-laced* Lie algebra has all roots of equal length. A

graded Lie algebra has odd and even elements, with a symmetric product between odd elements.

light-cone gauge in string theory, the choice of world-sheet time coordinate to coincide with a particular spacetime null coordinate. In theories with local symmetries, a gauge choice such that the connection in a given null direction vanishes.

linear dilaton theory a scalar CFT in which the energy-momentum tensor includes a term proportional to the second derivative of the scalar. This arises in string theory when the dilaton is a linear function of position.

linear sigma model a scalar field theory whose kinetic term is field independent, but whose long-distance physics is governed by a nonlinear sigma model.

Liouville field theory the CFT of a scalar field with an exponential interaction. This arises in various situations, including the *noncritical string*. It corresponds to bosonic string theory in a linear dilaton plus exponential tachyon background.

little string theory one of several interacting string theories without gravity, notably found on NS5-branes in the limit of zero string coupling.

loop expansion in quantum field theory, the Feynman graph expansion, which is equivalent to the expansion in powers of $\hbar$. The *string loop expansion* is the sum over Riemann surfaces, with dimensionless string coupling g. The *world-sheet loop expansion* is the nonlinear sigma model perturbation expansion, in powers of α'/R_c^2 with R_c the compactification radius.

Lorentzian having a mixed signature $(-,\ldots,-,+,\ldots,+)$.

lowering operator operators that reduce the energy of a given state. In CFT, operators that reduce the Virasoro generator L_0 (or $\tilde{L}_0$) by n units carry a grading (subscript) n.

M-theory 1. (*narrow*) the limit of strongly coupled IIA theory with eleven-dimensional Poincaré invariance; 2. (*broad; most common usage*) the entire quantum theory whose limits include the various weakly coupled string theories as well as M-theory in the narrow sense. The name is deliberately ambiguous, reflecting the unknown nature of the theory; M has variously been suggested to stand for membrane, matrix, mother, and mystery.

M2-brane the 2-brane of M-theory, which couples to the potential A_3 of eleven-dimensional supergravity.

M5-brane the 5-brane of M-theory, which carries the magnetic charge of the potential A_3 of eleven-dimensional supergravity.

macroscopic string a string whose length is much greater than the characteristic string length scale. In particular, it is sometimes useful to consider an infinite string stretching across spacetime.

Majorana condition a Lorentz-invariant reality condition on a spinor field. This can be imposed only if the spacetime dimension is 1, 2, 3, 4, or 8 (mod 8).

marginal interaction an interaction whose dimensionless strength is independent of the length scale. In general this might hold only to first order in the coupling of the interaction; a *truly marginal* interaction is one that remains marginal even with finite coupling.

matrix models quantum mechanical systems with matrix degrees of freedom, with critical points governed by noncritical string theories.

matrix theory a quantum mechanical system with matrix degrees of freedom and 32 supercharges, obtained by dimensional reduction of $d = 10$ supersymmetric $U(n)$ Yang–Mills theory. In the large-n limit this is conjectured to define M-theory (in the broad sense).

minimal models several families of solvable CFTs, in which every conformal family is degenerate. There are infinite series of unitary minimal models having $N = 0$, $N = 1$, and $N = 2$ superconformal symmetries, which converge from below on the central charges 1, $\frac{3}{2}$, and 3 respectively.

Minkowskian having a signature $(-,+,+,\ldots,+)$.

mirror symmetry an equivalence between string theories compactified on distinct manifolds. The equivalence reverses the sign of the $U(1)$ charge of one $N = 2$ superconformal algebra, and therefore changes the sign of the Euler number of the manifold.

Möbius group the globally defined $SL(2,\mathbf{C})$ conformal symmetry of the sphere; or, the globally defined $SL(2,\mathbf{R})$ conformal symmetry of the disk.

mode operators the spatial Fourier components of a quantum field.

model see *vacuum*.

modular group the group of large coordinate transformations (often applied to a Riemann surface but also applicable to spacetime).

modular invariance the invariance of the string path integral under large coordinate transformations.

moduli 1. the parameters labeling the geometry of a manifold. Notable examples are the parameters for the complex structure of the string world-sheet, and the parameters for the geometry of compactification; 2. the parameters labeling a space of degenerate (and, usually, physically inequivalent) vacua in quantum field theory. This is closely related to the compactification example: in expanding around the classical limit, each compact solution of the field equations gives a vacuum of the quantum theory, to leading order; 3. the massless fields corresponding to position dependence of these parameters. Contrast *Goldstone boson*.

moduli space the space of geometries or vacua, whose coordinates are the moduli.

monodromy for a quantity which is locally single-valued, the multi-valuedness around nontrivial closed paths.

Montonen–Olive duality the weak–strong duality of $d = 4$, $N = 4$ Yang–Mills theory.

Nambu–Goto action a string action, which is proportional to the invariant area of the world-sheet in spacetime.

Narain compactification the abstract description of toroidal compactification in terms of the lattice of left- and right-moving momenta.

naturalness problem the problem of explaining why a constant of nature takes a value much smaller than estimated nonzero contributions. Examples are the Higgs scalar mass, the cosmological constant, and the QCD θ-angle.

Neumann boundary condition the condition that the normal derivative of a field vanish at a boundary; the value of the field is free to fluctuate.

Neveu–Schwarz algebra the world-sheet algebra of the Fourier modes of the supercurrent and energy-momentum tensor, in a sector where the supercurrent is antiperiodic and its moding therefore half-integer-valued.

Neveu–Schwarz (NS) boundary condition the condition that a fermionic field on the world-sheet be antiperiodic, in the closed string or in the double of the open string (see *doubling trick*). Its Fourier moding is then half-integer-valued.

Neveu–Schwarz 5-brane in the type I and type II superstring theories, the 5-brane that carries the magnetic charge of the NS–NS 2-form potential.

Neveu–Schwarz–Neveu–Schwarz (NS–NS) states in type I and type II superstring theories, the bosonic closed string states whose left- and right-moving parts are bosonic. These include the graviton and dilaton, and in the type II case a 2-form potential.

no-ghost theorem 1. the theorem that the OCQ or BRST Hilbert space has a positive inner product; 2. the further theorem that the string amplitudes are well defined and unitary in this space.

no-scale model a field theory that has, to some approximation, a line of degenerate vacua with broken supersymmetry.

Noether's theorem the theorem that an invariance of the Lagrangian implies a conserved quantity.

noncommutative geometry a generalization of ordinary geometry, focusing on the algebra of functions on a space. The noncommutative collective coordinates of D-branes suggest the need for such a generalization.

noncritical string theory 1. a Weyl-noninvariant string theory — one with a measure of world-sheet distance that is independent of the embedding in spacetime. These include strings with an independent world-sheet metric field, and strings with a short-distance cutoff; 2. more recently, the term has been applied to any string theory that does not have a weakly coupled limit with a Weyl-invariant world-sheet theory. In this form it includes various theories with stringlike excitations in which the coupling is fixed to be of order 1, such as the *(2,0) theory* and the *little string theories*. Such a theory does not have a well-defined world-sheet, because processes that change the world-sheet topology cannot be turned off.

nonlinear sigma model a scalar field theory in which the kinetic term has a field-dependent coefficient. This has a natural interpretation in terms of a curved field space, and corresponds to string theory in curved spacetime, or more generally one with position-dependent background fields.

nonrenormalization theorem a theorem restricting the form of quantum corrections to a given amplitude, or to the effective action. It may require that these corrections vanish, that they arise only at specific orders of perturbation theory, or that they arise only nonperturbatively.

normal ordering a prescription for defining products of free fields by specific subtractions of divergent terms. *Conformal normal ordering*, denoted $: \; :$, produces operators with simple conformal properties. *Creation–annihilation normal ordering*, denoted $: \; :$, where lowering operators are put to the right of raising operators, produces operators with simple matrix elements. *Boundary normal ordering*, denoted $: \; :$, is conformal normal ordering with an additional image charge subtraction to produce operators that are finite as they approach a boundary.

null state a physical state that is orthogonal to all physical states including itself. Or, a descendant in a conformal family which is orthogonal to all states in the family.

old covariant quantization (OCQ) a method of quantizing string theory, similar to the Gupta–Bleuler quantization of electrodynamics. It is equivalent to the light-cone and BRST quantizations.

one-loop the leading quantum correction, coming from surfaces of Euler number zero in string perturbation theory.

open string a string that is topologically a line segment.

operator equation in quantum theory, an equality between operators that holds in arbitrary matrix elements; equivalently, an equality that holds when inserted into a functional integral with arbitrary boundary conditions.

operator product expansion (OPE) the expansion of a product of operators as a sum of local operators. This provides an asymptotic expansion, as the separation of the operators vanishes, for an arbitrary expectation value containing the product. In CFT the expansion is convergent.

orbifold 1. (*noun*) a coset space M/H, where H is a group of discrete symmetries of a manifold M. The coset is singular at the fixed points of H; 2. (*noun*) the CFT or string theory produced by the gauging of a discrete world-sheet symmetry group H. If the elements of H are spacetime symmetries, the result is a theory of strings propagating on the coset space M/H. A *non-Abelian orbifold* is one whose point group is non-Abelian. An *asymmetric orbifold* is one where H does not have a spacetime interpretation and which in general acts differently on the right-movers and left-movers of the string; 3. (*verb*) to produce such a CFT or string theory by gauging H; this is synonymous with the second definition of *twist*.

oriented string theory a string theory in which the world-sheet has a definite orientation; world-sheet parity-reversal is not treated as a gauge symmetry.

orientifold a string theory produced by the gauging of a world-sheet symmetry group H, where H includes elements that combine the world-sheet parity-reversal Ω with other symmetries.

orientifold plane a plane (of any dimension p) consisting of fixed points of the orientifold group H (specifically, of an element of H that includes Ω).

***p*-brane** a p-dimensional spatially extended object. Examples are *black p-branes, Dp-branes, M2-* and *M5-branes, NS5-branes, heterotic 5-branes,* and (in $d > 4$) *Kaluza–Klein monopoles.*

***p*-form** a fully antisymmetric p-index tensor, usually written in an index-free notation.

***p*-form gauge field** a generalization of Abelian gauge theory, with a p-form potential A, a $(p-1)$-form gauge parameter λ, and a $(p+1)$-form field strength F. For $p = 0$ this is an ordinary massless scalar; for $p = 1$ it is an Abelian gauge field. A $(p+1)$-form potential couples naturally to a p-brane, through the integral of the form over the world-volume. A *self-dual p-form theory* is one where $*F = F$; this requires the spacetime dimension to be $d = 2p + 2$, and further d must be 2 mod 4 (in the Minkowskian case) in order that $** = 1$.

***p*-*p'* string** an open string with one endpoint on a Dp-brane and the other on a Dp'-brane.

(p, q)-form on a complex manifold, a tensor that is completely antisymmetric in p holomorphic indices and q antiholomorphic indices.

(p, q) string a bound state of p F-strings and q D-strings in the IIB theory.

parafermion CFT a family of coset CFTs with $\mathbf{Z}_n$ symmetry, generalizing the $\mathbf{Z}_2$-invariant free fermion theory. These describe the generic critical behavior of a system with $\mathbf{Z}_n$ symmetry.

parity transformation an operation that reflects one spatial dimension, or any odd number. One distinugishes *spacetime parity, P* (or β) and *world-sheet parity* Ω.

partition function a sum over the spectrum of a quantum system, weighted by $e^{-H/T}$ where H is the Hamiltonian and T the temperature. Often additional charges are included in the exponent. This is the basic object in equilibrium statistical mechanics. In string theory it is given by a path integral on the torus or the cylinder, and so arises in one-loop amplitudes.

path integral a representation of the transition amplitudes of a quantum system as a coherent sum over all possible histories. In quantum mechanics the history is a particle path; in quantum field theory it is a path in field space; in first-quantized string theory it is the embedding of the string world-sheet in spacetime.

Pauli–Villars regulator a means of regulating quantum field theories by introducing a very massive wrong-statistics field.

Peccei–Quinn (PQ) symmetry an approximate symmetry, violated only by anomalies.

period matrix a $g \times g$ matrix characterizing the complex structure of a genus-g Riemann surface.

perturbation theory the expansion of the amplitudes of a quantum system in powers of the coupling.

physical state in a quantum system with constraints, a state annihilated by the constraints. In OCQ this is a state annihilated by the Virasoro lowering generators and having a specified L_0 eigenvalue.. In BRST quantization it is a state annihilated by the BRST operator. In both these cases, the true physical spectrum is the space of physical states with an additional equivalence relation, physical states differing by a null state being identified.

picture in the RNS superstring, one of several isomorphic representations of the vertex operators. The q-picture consists of vertex operators of $\beta\gamma$ ghost charge q. The natural pictures are $q = -1$ and $-\frac{1}{2}$, with higher pictures including partial integrations over supermoduli space. The *picture changing operator* increases q by one.

Planck length the natural length scale of quantum gravity, $L_{\mathrm{P}} = M_{\mathrm{P}}^{-1} = 1.6 \times 10^{-33}$ cm, constructed from $\hbar$, c, and G_{N}.

Planck mass the natural mass scale of quantum gravity, $M_{\mathrm{P}} = 1.22 \times 10^{19}$ GeV, constructed from $\hbar$, c, and G_{N}.

plumbing fixture a procedure for constructing higher genus Riemann surfaces from lower ones by sewing in a handle. The construction includes a parameter q, such that when q goes to 0 the handle degenerates or *pinches.* This gives a canonical representation of the boundary of the moduli space of Riemann surfaces.

Poincaré dual a map from p-forms to $(d-p)$-forms, given by contraction with the completely antisymmetric tensor.

Poincaré invariance the invariance group of the flat metric, consisting of translations and Lorentz transformations.

point group the orbifold group H, with translations ignored (applicable only for orbifolds having a spacetime interpretation).

Polyakov path integral a representation of first-quantized string theory as a path integral with an independent world-sheet metric. A local Weyl symmetry guarantees that the classical degrees of freedom are the same as those of the Nambu–Goto theory.

primary field in CFT, a local operator annihilated by all of the lowering generators of a given algebra, such as the Virasoro algebra. The corresponding state is a *highest weight state.*

projective space a compact n-dimensional space constructed from a linear $(n+1)$-dimensional space by identifying points under the overall rescaling $(x_1,\ldots,x_{n+1}) \cong (\lambda x_1,\ldots,\lambda x_{n+1})$. For x_i and λ real this produces RP^n, and for x_i and λ complex it produces CP^n (which has n complex $= 2n$ real dimensions).

pseudospin in a current algebra, an $SU(2)$ subalgebra not contained in the Lie algebra of the center-of-mass modes.

puncture a marked point on a Riemann surface, the position of a vertex operator.

QCD string a reformulation of non-Abelian gauge theory as a string theory, conjectured to exist at least in the limit of a large number of colors.

quaternionic manifold a $4k$-dimensional manifold with the holonomy group $Sp(k) \times SU(2) \subset SO(4k)$, with specific $SU(2)$ curvature. This is the geometry of the moduli space of hypermultiplets in $d = 6$, $N = 1$ or $d = 4$, $N = 2$ supergravity.

***R* symmetry** a symmetry that acts nontrivially on the supercurrent(s).

raising operator operators that reduce the energy of a given state. In CFT, operators that increase the Virasoro generator L_0 (or $\tilde{L}_0$) by n units carry a grading (subscript) $-n$.

Ramond algebra the world-sheet algebra of the Fourier modes of the supercurrent and energy-momentum tensor, in a sector where the supercurrent is periodic and its moding therefore integer-valued.

Ramond (R) boundary condition the condition that a fermionic field on the world-sheet be periodic, in the closed string or in the double of the open string (see *doubling trick*). Its Fourier moding is then integer-valued.

Ramond–Neveu–Schwarz (RNS) superstring the formulation of type I and II superstrings that has superconformal invariance but not manifest spacetime supersymmetry. The latter emerges after imposing the GSO projection on the string Hilbert space.

Ramond–Ramond (R–R) states in type I and type II superstring theories, the bosonic closed string states whose left- and right-moving parts are fermionic. These include p-form potentials C_p, with p taking all odd values in the IIA string and all even values in the IIB string.

rank 1. the maximal number of commuting generators of a Lie algebra; 2. the number of indices on a tensor.

rational CFT a CFT with a finite number of primary fields under an extended chiral algebra, a generalization of the *minimal models*. Such CFTs are highly constrained.

refermionization after bosonization, the construction of new spin-$\frac{1}{2}$ fields from linear combinations of the bosonic fields. These fermions are nonlocal, nonlinear functions of the original fermions.

Regge behavior in scattering at large center-of-mass energy-squared s and fixed momentum transfer-squared $-t$, the scaling of the amplitude as $s^{\alpha(t)}$. The values of t where $\alpha(t) = j$ is a nonnegative integer correspond to exchange of a particle of spin j and mass-squared $-t$.

Regge slope denoted α', the square of the characteristic length scale of perturbative string theory. The tension of the fundamental string is $1/2\pi\alpha'$.

relevant interaction an interaction whose dimensionless strength increases with distance. In perturbation theory, this is equivalent to a *superrenormalizable* interaction.

renormalization group equation the differential equation governing the change of physics with length scale.

renormalization theory the calculus of path integrals.

Riemann–Roch theorem the theorem that the number of metric moduli minus the number of conformal Killing vectors on a Riemann surface is -3χ, with χ being the Euler number of the surface; and, generalizations of this result.

Riemann surface a two-(real)-dimensionsional complex manifold, equivalent to a Weyl equivalence class of Riemannian manifolds.

root a vector of the eigenvalues of the maximal set of commuting generators of a Lie algebra, associated with a state in the adjoint representation.

***S*-duality** a duality under which the coupling constant of a quantum theory changes nontrivially, including the case of weak–strong duality. Important examples are the $SL(2,\mathbf{Z})$ self-dualities of IIB string theory and of $d = 4$, $N = 4$ supersymmetric Yang–Mills theory. More loosely, it is used for weak–strong dualities between different theories, such as IIA–M-theory (on a circle) duality, $SO(32)$ heterotic–type I duality, and E_8 heterotic–M-theory (on an interval) duality. In compactified theories, the term S-duality is limited to those dualities that leave the radii invariant, up to an overall coupling-dependent rescaling; contrast *T-duality* and *U-duality*.

S-matrix the overlap amplitude between states in the infinite past and states in the infinite future; the scattering amplitude. In coordinate-invariant quantum theories this is generally the simplest invariant. The term usually implies a basis of free particle states; this is problematic in theories with massless particles due to IR divergences, and meaningless in theories at nontrivial IR fixed points.

scale transformation a rigid rescaling of spacetime, or of the world-sheet.

Scherk–Schwarz mechanism the breaking of supersymmetry by dimensional reduction that includes a spacetime rotation.

Schottky group a discrete subgroup Γ of the $SL(2,\mathbf{C})$ Möbius transformations of the sphere S_2, with additional conditions such that S_2/Γ is a manifold and in particular a Riemann surface.

Schwarzian the combination of derivatives appearing in the finite conformal transformation of the energy-momentum tensor.

Schwinger–Dyson equation the operator equations of a quantum theory, expressed as equations for the expectation values.

short multiplet see *BPS state.*

sigma model metric the metric appearing in the string world-sheet action. Also known as the *string metric,* this differs from the *Einstein metric* by a dilaton-dependent Weyl transformation.

simple current an operator J such that, for any primary field $\mathcal{O}$, the operator product $J\mathcal{O}$ contains only a single conformal family.

soliton a state whose classical limit is a smooth, localized, and (usually) topologically nontrivial classical field configuration; this includes particle states, which are localized in all directions, as well as extended objects. By contrast, a state of ordinary *quanta* is represented near the classical limit by small fluctuations around a constant configuration. In a theory with multiple classical limits (*dualities*), solitons and quanta may exchange roles.

special Kähler geometry the geometry of the moduli space of vector multiplets in $d = 4$, $N = 2$ supergravity. It is most simply defined (section B.7) in terms of *special coordinates,* which are fixed up to a symplectic transformation. *Rigid special geometry* is obtained in the limit where gravity decouples and the supersymmetry becomes global.

spectral flow the adiabatic change in the spectrum produced by a continuous change in the boundary conditions.

spin the behavior of a field or a state under rotations. In a CFT this is given in terms of the conformal weights by $h - \tilde{h}$.

spin field the vertex operator for a Ramond ground state, which produces a branch cut in the spinor fields.

spin structure one of a set of inequivalent ways of defining a spinor field globally on a manifold. Roughly speaking, it corresponds to a choice of signs in the square roots of the transition functions.

spurious state in OCQ, a state produced by Virasoro raising operators.

state–operator isomorphism in CFT, a one-to-one isomorphism between states of the theory quantized on a circle and local operators. Also, a one-to-one isomorphism between states of the theory quantized on an interval and local operators on a boundary. In d dimensions the circle becomes a $(d-1)$-sphere and the interval a $(d-1)$-hemisphere.

string coupling the dimensionless parameter g governing the weights of different Riemann surfaces in string perturbation theory, the contribution from surfaces of Euler number χ being weighted by $g^{-\chi}$. The string coupling is related to the dilaton by $g = e^{\Phi}$. This definition corresponds to the amplitude to emit a closed string; the amplitude to emit an open string is proportional to $g^{1/2}$.

string field theory the representation of string theory as theory of fields, the fields being maps from a circle (or interval) into spacetime. This corresponds to an infinite number of ordinary quantum fields. This formalism can reproduce string perturbation theory, but it is unclear whether it can be defined beyond perturbation theory.

string metric see *sigma model metric.*

string scale the mass scale $\alpha'^{-1/2}$ characterizing the tower of string excitations.

string–string duality a term sometimes used to denote a weak–strong duality between different string theories, in particular between the heterotic string compactified on T^4 and the IIA string compactified on K3.

string tension the mass per unit length of a string at rest, related to the Regge slope by $1/2\pi\alpha'$.

Sugawara construction in current algebra, the construction of the energy-momentum tensor as a product of two currents. Originally proposed as a phenomenological model in four dimensions, this was later found to be an exact result in two dimensions.

superconformal algebra an extension of the conformal (Virasoro) algebra to include anticommuting spinor generators. The $(N, \tilde{N})$ superconformal algebra has N left-moving and $\tilde{N}$ right-moving supercurrents.

superconformal current algebra an extension of the conformal transformations to include both spin-$\frac{3}{2}$ and spin-1 currents.

superconformal field theory (SCFT) a quantum field theory that is invariant under superconformal transformations.

supercurrent a conserved spinor current. This includes the world-sheet current T_F associated with superconformal transformations, and the spacetime current associated with spacetime supersymmetry.

superfield a field on superspace, with specific transformation properties under a change of coordinates.

supergravity the union of general relativity and supersymmetry, implying also the promotion of supersymmetry to a local symmetry.

supermanifold (or *superspace*) a formal extension of the concept of manifold to include both commuting and anticommuting (Grassmann) coordinates.

supermoduli the anticommuting parameters characterizing a super-Riemann surface.

superpartner scale the mass scale of the superpartners of the Standard Model particles. This is expected to be between 10^2 and 10^3 GeV if supersymmetry solves the naturalness problem of the Higgs scalar mass.

superpotential in $d = 4$, $N = 1$ supersymmetry, the holomorphic function of the superfields that determines the nongauge interactions.

super-Riemann surface a supermanifold defined in terms of superconformal transition functions between patches.

supersymmetry a symmetry whose charge transforms as a spinor, which relates the masses and couplings of fermions and bosons.

supersymmetry breaking scale the mass scale of the expectation value that breaks supersymmetry. The superpartner scale is the supersymmetry breaking scale times the strength of the coupling of the Standard Model fields to the supersymmetry breaking fields.

***T*-duality** a duality in string theory, usually in a toroidally compactified theory, that leaves the coupling constant invariant up to a radius-dependent rescaling and therefore holds at each order of string perturbation theory. Most notable is $R \to \alpha'/R$ duality, which relates string theories compactified on large and small tori by interchanging winding and Kaluza–Klein states. More generally it includes shifts of antisymmetric tensor backgrounds and large coordinate transformations in spacetime. Contrast *S-duality* and *U-duality*.

't Hooft–Polyakov monopole a classical solution with magnetic charge, which exists whenever a simple group is spontaneously broken to a group with a $U(1)$ factor.

tachyon a particle (almost always a scalar) with a negative mass-squared, signifying an instability of the vacuum.

tadpole an amplitude for creation of a single particle from the vacuum, induced by quantum effects.

target space the space in which a function takes its values. This is usually applied to the nonlinear sigma model on the string world-sheet, where the target space is itself spacetime.

Teichmüller parameters the moduli for the complex structure of a Riemann surface (strictly speaking, points in *Teichmüller space* are not identified under the modular group).

tensionless string theory an interacting theory with tensionless strings. These can arise as p-branes with $p-1$ directions wrapped on a collapsing cycle, as various 2-branes with one direction stretched between higher-dimensional branes when the latter become coincident, and on zero-size $E_8 \times E_8$ instantons. In general the coupling is fixed to be of order one, so there is no perturbation expansion. *Tensionless* implies that the string tension in units of the gravitational scale goes to zero; it is not applied to the fundamental string, which becomes noninteracting in that limit.

tensor multiplet 1. the multiplet of $d = 6$, $(1,0)$ supersymmetry whose bosonic content is one self-dual tensor and one scalar. This reduces to a vector multiplet of $d = 4$, $N = 2$ supersymmetry; 2. the multiplet of $d = 6$, $(2,0)$ supersymmetry whose bosonic content is one self-dual tensor and five scalars. This reduces to a vector multiplet of $d = 4$, $N = 4$ supersymmetry.

tensor operator in CFT, a local operator whose conformal or superconformal transformation involves only the first derivative of the transformation; synonymous with a *primary field* of the conformal or superconformal symmetry. Such an operator is mapped to a *highest weight state* by the state–operator isomorphism.

tetrad a basis of d orthonormal vector fields in d dimensions. The term tetrad originates in $d = 4$ (as does the equivalent term *vierbein*) but there is no other convenient term for general d.

the theory formerly known as strings see *M-theory*, second definition.

theory see *vacuum*.

theta functions the family of holomorphic functions having simple periodicity properties on a torus.

Thirring model the solvable quantum field theory of a single Dirac fermion with a quartic interaction in $1+1$ dimensions. This is equivalent under bosonization to a free scalar field at a general real radius.

threshold correction a correction to the low energy effective action, and in particular to the gauge coupling, due to virtual massive particles.

topological string theory a modification of string theory without local dynamics; all observables are topological.

toric geometry a generalization of the idea of projective space. Roughly speaking, this corresponds to the most general linear sigma model.

toroidal compactification the periodic identification of one or more flat dimensions.

torsion a term applied to various 3-form field strengths, so called because they appear in covariant derivatives in combination with the Christoffel connection.

tree-level the Feynman graphs which become disconnected if one propagator is cut, or the analogous string amplitudes, the sphere and disk. These correspond to classical terms in the effective action.

twist 1. (*verb*) to define a field in a periodic space to be aperiodic by a symmetry transformation h; 2. (*noun*) the aperiodicity h; 3. (*verb*) given a CFT or string theory, to construct a new theory using a symmetry group H. One adds closed strings twisted by any of the elements $h \in H$, and requires all states to be invariant under the transformations in H. This is equivalent to treating H as a world-sheet gauge symmetry. The term *orbifold* is also used as a synonym; 4. (*noun, archaic*) world-sheet parity.

twisted state a closed string with twisted periodicity.

type ... supergravity the low energy supergravity theory of the corresponding string theory. The Roman numeral signifies the number of $d = 10$ supersymmetries, and IIA and IIB distinguish whether the two supersymmetries have opposite or identical chiralities respectively.

type I superstring the theory of open and closed unoriented superstrings, which is consistent only for the gauge group $SO(32)$. The right-movers and left-movers, being related by the open string boundary condition, transform under the same spacetime supersymmetry.

type IIA superstring a theory of closed oriented superstrings. The right-movers and left-movers transform under separate spacetime supersymmetries, which have opposite chiralities.

type IIB superstring a theory of closed oriented superstrings. The right-movers and left-movers transform under separate spacetime supersymmetries, which have the same chirality.

type 0 string a pair of nonsupersymmetric string theories, which have tachyons and no spacetime fermions, with the same world-sheet action as the type II theories but with different projections on the Hilbert space.

***U*-duality** any of the dualities of a string theory, usually of a toroidally compactified type II theory. This includes the *S-dualities* and *T-dualities*, but in contrast to these includes also transformations that mix the radii and couplings.

ultrashort multiplet see *BPS state*.

unit gauge a choice of coordinate and Weyl gauges in two dimensions, such that the metric is the identity.

unitary as applied to a quantum system, the property of having a conserved inner product in a positive-norm Hilbert space.

unoriented string theory a string theory in which world-sheet parity-reversal Ω is a discrete gauge symmetry. The perturbation theory includes unoriented world-sheets, and the spectrum is restricted to states with $\Omega = +1$.

UV divergence a divergence arising from short distances in spacetime, usually signifying a limit to the validity of a theory.

vacuum a stable Poincaré-invariant state. The novel feature of systems with unbroken supersymmetries is the frequent appearance of degenerate but physically inequivalent vacua.
While it is now clear that the different string theories are actually different *vacua* in a single theory, it is still common to use the term *theory* for each. Also, different CFTs within a single string theory are sometimes referred to as different *theories* rather than *vacua*. The term *model* is used to refer to string vacua whose low energy physics resembles the Standard Model.

vector multiplet in $d = 4$, $N = 1$ or $d = 6$, $N = 1$ supersymmetry, the multiplet whose bosonic content is a massless vector field. The $d = 6$ multiplet reduces to a vector field plus two real scalars in $d = 4$, $N = 2$ supersymmetry.

Veneziano amplitude the bosonic string tree-level amplitude for four open string tachyons.

Verma module the set of states obtained by acting on a highest weight state with Virasoro raising generators in all inequivalent ways, with the requirement that all such states be linearly independent (the generators act freely). This gives a representation of the Virasoro algebra that depends on the central charge c and on the weight h of the highest weight state. This is very similar to a *conformal family*, except that in the case of a degenerate representation some states in the latter may vanish; the Verma module would in this case give a reducible representation.

vertex operator a local operator on the string world-sheet, corresponding to a string in the initial or final state.

Virasoro algebra in a CFT, the infinite-dimensional Lie algebra of the Fourier modes of the energy-momentum tensor.

Virasoro–Shapiro amplitude the bosonic string tree-level amplitude for four closed string tachyons.

***W* algebras** a family of extended chiral algebras with currents of spin greater than 2.

***W* string** a proposed generalization of string theory with constraints of spin greater than 2. No complete construction exists.

Ward identity a relation between the divergence of the expectation value of a conserved current and the same expectation value without the current.

weight 1. in CFT, the L_0 or $\tilde{L}_0$ eigenvalue, which determines the behavior of an operator under scale transformations and rotations (= *conformal weight*); 2. in Lie algebra, a vector of the eigenvalues of the maximal set of commuting generators of a Lie algebra, in a state of a given representation.

Wess–Zumino consistency condition the condition that the second derivative of a functional integral with respect to the background fields be symmetric. This strongly constrains the form of possible anomalies in a theory.

Wess–Zumino–Novikov–Witten (WZNW) model a conformally invariant nonlinear sigma model on a group manifold with an antisymmetric tensor background.

Weyl anomaly an anomaly in the Weyl transformation, this determines the critical dimension in the Polyakov formalism. This is sometimes called the *trace anomaly* or the *conformal anomaly*.

Weyl condition in even dimensions, the condition that a fermion have definite chirality. This defines a spinor representation of the Lorentz group, which contains half of the components of the Dirac representation of the gamma matrix algebra.

Weyl transformation a position-dependent rescaling of the metric.

Wilson line 1. a gauge field with vanishing field strength but with nontrivial parallel transport (holonomy) around nontrivial paths in spacetime. 2. the gauge-invariant operator that measures such a field: the trace of the path-ordered product of the line integral of the vector potential.

Wilsonian action the action in a low energy effective field theory, which incorporates the effects of higher energy virtual states.

winding state a closed string whose configuration is a nontrivial path in a non-simply-connected spacetime.

world-sheet 1. the two-dimensional surface in spacetime swept out by the motion of a string. 2. the abstract two-dimensional parameter space used to describe the motion of a string.

world-sheet duality the equivalence between Hamiltonian descriptions obtained by cutting open a string world-sheet along inequivalent circles. Important examples are the associativity condition, from the four-point sphere amplitude, and the equivalence of the open string loop and closed string tree descriptions of the cylinder.

wrapped refers to a p-brane, q of whose dimensions are wound on a nontrivial compact submanifold of spacetime, leaving a $(p-q)$-dimensional extended object.

Yang–Mills field non-Abelian gauge field.

Zamolodchikov metric the expectation value of a pair of local operators on the sphere, giving a natural inner product for the corresponding string states. This is the metric that appears in the kinetic term in the spacetime action.

zero modes 1. of a differential operator: eigenfunctions with zero eigenvalue. In a Gaussian functional integral, these would give an infinite factor in the bosonic case and a zero factor in the fermionic case. In general these have a physical origin and the functional integral has an appropriate integrand to give a finite result; 2. center-of-mass nodes.

zero-point energy the energy due to vacuum fluctuations of quantum fields.

zero-slope limit the limit $\alpha' \to 0$. Only massless string states, described by low energy field theory, remain.

Index

Page numbers in italics refer to exercises

Abelian differential 293
action (D-brane) 270–3
  Born–Infeld form 271, 273
  linearized 275–6
action (general)
  Euclidean 335
  fermionic oscillator 343
  harmonic oscillator 334, 339–41
  in path integral 77, 331, 343
action (point particle) 9–11
  Euclidean 82–3
  gauge-fixed 129
  in curved spacetime 108
  Wilson line background 264
action (spacetime)
  $c = 1$ moduli 246
  closed string tachyon 194
  dimensional reduction 232–5, 249, 265
  higher-derivative terms 188–9, 195
  Hilbert (gravitational) 83, 114, 194
  inheritance principle 258
  local symmetries 196
  massless closed string 114–15
    quadratic 275–6
  open string 188–9, 196–7, 276
  string field 311–14, *327–8*
  tadpole terms 197–8, 225, 309
action (world-sheet)
  bc 50
  $\beta\gamma$ 52
Brink–Di Vecchia–Howe–Deser–Zumino 12
Euclidean 82–3
Euler term (χ) 15–16, *30*, 83–4, 179, 203
Faddeev–Popov ghost 89, 157–8
gauge-fixing 126–7, 131
Hilbert (gravitational) 15–16
linear dilaton 49
Liouville 324–5
massless scalar 32, 34, *346*
Nambu–Goto 10–13, 15, 28, *30*
  and D-brane action 270
noncritical string 325–6
nonlinear sigma model 108–10, 112
  T-duality *280–1*
Polyakov 12–16, 28
  and noncritical string 325
  Euclidean 82–3
  flat world-sheet 32
wrapped torus 237
  antisymmetric tensor background 249–50
affine Lie algebra 243
annulus, *see* cylinder
anomaly
  BRST 132–3, *143*
    from string loops 308–9
  conformal *same as* Weyl
  current conservation *119*, 158–9
  ghost number 158–9, 177
  gravitational (diff) 95, 349

anomaly (*cont.*)
Lorentz 23–4, 94
trace *same as* Weyl
Weyl 90–7, *119*, 349
from string loops 308–9
in nonlinear sigma model 111–12
in vertex operators 102–7
integrated 93–4
anticommuting c-numbers 341
antiperiodic boundary conditions
in traces 212, 336, 344, *345–6*
in twisted states 257–9
antisymmetric tensor field
as axion 29
background 108–15, *120*
constant 249–51, 254–5
coupling to
D-brane 270–2, *282*
winding strings 241–2
dimensional reduction 234–5, 241
field equation 111–12
gauge invariance 27, 109, 196
in BRST quantization 137
in light-cone quantization 27–30
in OCQ 125
spacetime action 114
vertex operator 104–7
axion 1, 6, 29

bc CFT 50–1, *75–6*, 89
expectation values 176–8, 212–13
mode expansion 61–3
state–operator mapping 65–6
$\beta\gamma$ CFT 52, *75*
BRST invariance
of string amplitudes 160–1, 283, 285, 287–90, 351
of vertex operators 164, *165*
BRST quantization 126ff, 350
and string field theory 311–15
nilpotence 128, 130–2, 138, *143*
point particle 129–31
string 131–43
anomaly 132, *143*, 308–9
charge 132
current 131
equivalence to light-cone 137–41
equivalence to OCQ 141–3
no-ghost theorem 137–41
transformation 131
background charge 170, *204*, 208
background fields 108–18
linearized, and vertex operators 108–11
Batalin–Vilkovisky (BV) formalism 128
Beltrami differential 162
Beltrami's equation 162
Berezin integration 341
Jacobian 345
beta function (Euler) 183
beta function (field theory) 91, 111–12, 349
string loop corrections 308–9
world-sheet corrections 112, 195
black holes 6–7, 320
Boltzmann weight 325
Borel summation 316, 352
Born–Infeld action 271, 273
bosonic string theory
artifacts of 23, 219, 223–4, 246, 275, 306
reason for studying 8
bosonization 303
boundary state 226, *230*

CPT symmetry 23
Calabi–Yau compactification 352
Campbell–Baker–Hausdorff formula 202
canceled propagator argument 161, 196, 288, 309
Casimir energy 22
in unitary CFT 73
mnemonic 73
central charge 48, *75–6*
and BRST anomaly 132, *143*
and compactification 113, 117, 137
and density of states 213
and gravitational anomaly 95
and ground state energy 73
and Weyl anomaly 92–6

bc 51
$\beta\gamma$ 52
free scalar 48
in Virasoro algebra 56
linear dilaton 49
positivity 72, *76*
Chan–Paton factor 186–7
Chan–Paton states 30, 184–92
and D-branes 268–70
in cylinder amplitude 223
in Möbius strip 228
non-Abelian gauge interactions 188–9
unoriented string 189–92
Chern–Simons term 234–5
chiral gauge symmetry 6
closed state (BRST) 128
closed string 14
BRST quantization 136–7
light-cone quantization 25–9
OCQ 124–5
one-loop amplitudes 216–22, 226–8
poles in cylinder amplitude 225–6, *230*, 274–6
tree-level amplitudes 192–8
relation to open 195
vertex operators 63–8, 101–7
cocycle 239
coefficient functions (OPE) 37–8
and conformal bootstrap 304
and conformal invariance 47
and matrix elements 200
and three-point amplitudes 284
determine CFT 303
generalized 303
interchange symmetry 40–1
cohomology (BRST) 128–9, 134–43, *144*, 350
and amplitudes 283, 285, 289–90
Coleman–Weinberg formula 221, *229–30*, 350
collective coordinates (D-brane) 269, *281–2*
commutators and anticommutators
canonical 17, 20, 26, 130
Jacobi identity 128, 133, *143*
relation to OPE 54–5
wrong-sign 27–8, 72, 121
commutators and anticommutators (specific)
$[\alpha_m^\mu, \alpha_n^\nu]$ 20, 26–7, 58, 62
$\{b_n, c_n\}$ 61, 130, 133
$[\delta_{\tau_1}, \delta_{\tau_2}]$ 129
$[G_I, G_J]$ (constraints) 133
$[j_m^i, j_n^j]$ 243
$[L_m, L_n]$ 56–7, *75–6*
$[L_m, \mathcal{O}_n]$ 56
$[N^g, b_m]$ 62
$[N^g, c_m]$ 62
$[p_\mu, x^\nu]$ 17, 20, 26, 58, 62, 130
$[Q, \mathcal{A}]$ 55
$[Q_1, Q_2]$ 55
$\{Q_B, b_m\}$ 132
$\{Q_B, L_m\}$ *143*
$\{Q_B, Q_B\}$ 132–3, *143*
$[X_L(z), X_L(z')]$ 239
compact CFT 117, 137
compactification 2, 6, 115–18
abstract CFT 117–18, 137
scale 110, 112–13, 116, 244–5
see also orbifold, orientifold, toroidal
complex coordinates 33–4, 153
divergence theorem 33–4
complex manifold 153
conformal bootstrap 304
conformal field theory (CFT) 43–76, 349, 351
and compactification 115–18
$c = 1$ 261–3, 351
compact 117, 137
coupling to curved metric 95, 97
expectation values
and matrix elements 199–201
Möbius invariance 198–9, *205*
positive genus 169–78, 201–2
zero genus 208–16
free 32–6, 49–52
matter 122, 131
mode expansions 58–63
on Riemann surface 154
periodic scalar 235–40, 242–4, 247–8

conformal field theory (CFT) (*cont.*)
sewing and cutting 300–5
state–operator isomorphism 63–8
unitary 71–3, 117, 137
see also bc, $\beta\gamma$, free scalar, linear dilaton CFTs
conformal gauge 85
and light-cone gauge 83
curvature 93
ghost action 89
OCQ 121
sphere 166
variations around 152
conformal invariance 43–5,
and OPE 45–7
relation to diff×Weyl invariance 85–6
conformal Killing equation 151
conformal Killing group (CKG) or vector (CKV) 149–52, 155, 157, 161
specific Riemann surfaces
positive genus 167–9, 176, 178, 192, 197,
zero genus 206–8, 212, 216–17
conformal transformation 44–5, *75*
and Virasoro algebra 52–8
$d \geq 3$ *75*
of energy-momentum tensor 47–9
of ghost-number current 51
conjugacy classes 260
constraints (in quantization) 28, 130, 133, *143–4*, 350
contraction 39–40, 50, 65, *75*, 172, 175
self- 103, 171, 338
coordinate transformation
large *see* modular invariance
spacetime 28, 109, 196, 285, 310
world-sheet *see* diff invariance
cosmological constant
spacetime 222, 306, 350
world-sheet 95, 323
covariant derivative 89, *118*
covariant gauge *see* conformal gauge
critical dimension
effect of backgrounds 113, *120*
from BRST anomaly 132, *143*
from cylinder amplitude *230*
from Lorentz anomaly 21–4, 27
from modular invariance *230*
from OCQ 124, *143*
from Veneziano amplitude *204*
from Weyl anomaly 94–5
critical point 44
cross-ratio 199
cross-cap 101–2, 207–8, 227–8
state *230*
current algebra 243–4
curved spacetime 108–18
cycle
A- and *B*- 147–8, 292, 296, 304
cylinder (annulus) 78, 81
long *see* handle, degenerating Riemann surface 207
semi-infinite 63–4
string amplitude 222–6, *229–30*
calculation of D-brane tension 274–6
closed string poles 225–6, *230*
divergence 224–5, 229
unoriented 226

D-brane 7, 31, 104, 268–82, 317, 351
action 270–3
gauge symmetry 268–73
half-D-brane 280
tension 273–7
Dedekind eta function 210, 216
Del Guidice–Di Vecchia–Fubini (DDF) operators 143, 240
density of states 213, 321
determinant
functional 182, 338–41, 344–5
in Born–Infeld action 271
of Laplacian 170
variation 12
see also Faddeev–Popov determinant
diff (reparameterization) invariance 12–13
anomaly 91, 95, 349
gauge-fixing 85, 129, 325
of vertex operators 102–3
relation to conformal invariance 44, 85–6
removes unphysical states 27–8, 117
with boundary 90
dilaton 6
and equivalence principle 115
and string coupling 112–15
background 108–15, 233–4
coupling to
D-brane 270–7
open strings 196–8
orientifold plane 280
dimensional reduction 233–4, 241
effect of T-duality on 248
field equation 111–12, 246
in BRST quantization 137
in light-cone quantization 27, 29–30
in OCQ 125
linear *see* linear dilaton CFT
moduli space 254
propagator (field theory) 276
spacetime action 114–15, 196–8, 225, 275–6
vertex operator 104–7
dimensional reduction 235, 249
Dirichlet boundary condition *31*, 266–70, *281*
T-dual to Neumann 266–70
discrete gauge symmetry 260–1, 277–8
disk
conformal to semi-infinite cylinder 63–8
expectation values 174–5, 178
Riemann surface 168
string amplitudes 178–89, 196–8, *204–5*
with D-branes 270, 274, *281–2*
tadpole 197–8, 223–6
cancellation with projective plane 229
divergences
analytic continuation past 181, 287
cancellation for $SO(8192)$ 228–9
from boundary of moduli space 196, 296, 305–6
IR 181, 223–5, 306–10
UV 2–7, 218–20
absence in string theory 4–5, 79, 219, 223–5, 309–10
zero mode 173
doubling trick 57, 63, 66, 178
dual resonance model 348
duality *see* T-duality, world-sheet duality.

effective field theory 110, 114, 117, 188–9, 196–7, 270–3, 309
Einstein metric 115, 194, 275
Einstein's equation 112–13, 222
electroweak scale 222
energy-momentum tensor 13, 91
and derivative with respect to moduli 161, 210–11
as constraint 121
boundary condition 56–7
choice of 49, 95, 303
conformal transformation 47–9
trace anomaly 92–7
traceless and holomorphic 43–4
vanishes for Chan–Paton states 185
enhanced gauge symmetry 242–4, 248, 254
epsilon prescription 182, 287–8, 305–6
equivalence classes
BRST 128
gauge 84, 87
OCQ 123
equivalence principle 115
Euclidean adjoint 202–3, 336
Euclidean continuation 82–3, 170, 210, 335
Euler number 96, 101, *118*, 152
in string action 83–4, 112, 179, 203
exact state (BRST) 128
exceptional Lie algebras 192
expectation values 34, 36
calculations 169–78, 201–2, 208–13
exponential operator 40, 46
exterior derivative 128

Faddeev–Popov determinant 86–90, 154–8
  on disk 176
  on sphere 176
  on torus 212–13, 216–17
  shortcuts 179, 218
  sign 156, 177, 345
Faddeev–Popov ghosts 51–2, 88–90, 126
  boundary condition 90
  central charge 94
  vacuum 61, 141
Faddeev–Popov procedure, general 126
Feynman diagrams
  calculation of D-brane interaction 275–6
  calculation of Weyl anomaly *119*
  decomposition of string amplitudes 297–306
  vs string graphs 3–5, 101
fiducial metric 85
field strength, three-index 109, 112, 234–5
first-quantization 310
Fischler–Susskind mechanism 308–9
fixed point
  orbifold 256–8
  UV 4
flat direction 246, 262–3, 272–3
free energy 321–2, *328*
free parameters, absence of 6, 83–4, 112–13
free scalar CFT 32–41
  expectation values 169–76, 208–11, *229*
  left-, right-moving 238
  mode expansion 58–61
  state–operator mapping 63–5
Fuchsian group 292–3
functional integral *see* path integral
fundamental region 148–50
  in torus amplitude 216, 219–20

gauge boson 1, 4
  open string 28, 123–4, 135–6
    amplitudes 187–9, *205*
    in BRST quantization 135–6
    in light-cone quantization 23
    in OCQ 123–4
    vertex operator 107
  closed string 241–3
    amplitudes 242
gauge coupling 115, 234, 242–5, 276
gauge-fixing
  dependence on gauge choice 90–1, 121–2, 126–8
  diff×Weyl 17–18, 25, 84–6, 131, 154–5
  general 126–8
  see also specific gauges
gauge interaction 2, 28, 78, 184–92
gauge transformation 107
Gaussian-averaged gauges 127
general relativity 1–2, 5, 28, 108
  off-shell amplitudes 104
genus 100–1, 150
  large 315–17
geodesic curvature *30*, 82, 96
geodesic distance 102
geodesic equation 108
ghost number 51, 62, 127
  anomaly *119*, 158–9, 177
  of string field 313–14
  states vs operators 65–6
ghosts *see* Faddeev–Popov ghosts
global symmetry 189
Goldstone boson 233
graded Lie algebra *143*
grand unification 2, 5
Grassmann variables 35, 88, 156–8, 341–5
gravitational coupling 114
  relation to
    gauge coupling 234, 244–5
    string coupling 194, *205*
graviton
  coupling to
    D-brane 270–7
    open strings 196–8
    orientifold plane 280
  in BRST quantization 137
  in light-cone quantization 27, 29–30
  in OCQ 125
  propagator (field theory) 276
  vertex operator 104–7
gravity 1–7, 78
Green's function 169–71, 174–6, 208, *229–30*, 275

Hagedorn temperature 321–2, 352
Hamiltonian
  and path integral 329–32, 335–6, 343
  effect of backgrounds 250, 264
  in partition function 209, 212–13
  point particle 17, 130, 218
  string 19–21, 25–6
    relation to L_0 54, 73, 125
handle 83, 100–1, 290–2
  degenerating 97–8, 101, 224–6, 229, 295–6, 305–9
hard scattering 183–4, 194, *281–2*, 317–19, 352
heterotic string 348
hidden sector 6
Higgs field 1, 222, 245–6
high-energy scattering *see* hard scattering, Regge scattering
high temperature 320–2
highest weight state 70–2, 122, 125, *143*
  and tensor (primary) operators 71
holomorphic
  current 42
  field 34, 56, 72
  quadratic differential 152
  transition functions 153–4
  vector field 152
hyperelliptic surface 294

infrared (IR) divergences 181, 223–5, 306–10
inheritance principle 258, 261
inner product
  bilinear (Zamolodchikov) 199–204
    in sewing 301–2
    relation to Hermitean 202–4
  Fermi oscillator 343
  Hermitean 134–5
  positive 71–2, 116–17, 138
  traceless symmetric tensor *118*
instantons 316
interaction
  irrelevant 3
  string 77–82, 104, 178ff

Jacobi identity 128, 133, *143*
Jacobi's abstruse identity 215
Jacobian, Berezin integral 345

Kac–Moody algebra 243
Kaluza–Klein gauge field 2, 232–4, 241–3, 249, 257, 351
Kaluza–Klein states 232, 245
Klein bottle
  Riemann surface 207
  string amplitude 226–9, *230*, 278, 280

large order behavior 315–17
lattice
  even self-dual 252–3, *281*
  momentum 251–3, *281*
Laurent expansion *see* mode expansion
length, minimum 248, 318
level
  current algebra 243–4
  oscillator 23, 26–7
level-matching 27, 258
light-cone coordinates 16
light-cone gauge 17
  off-shell 104
light-cone quantization
  point particle 17
  string 17–30, 349
    equivalence to BRST and OCQ 137–43
linear dilaton theory 49–50, 61, 75, *204*, 213, 349
  as background 113, *120*, 323–4, *328*
Liouville field theory 325
loop expansion 115
Lorentz invariance 12, 106–7, 116–17, 290
  in light-cone gauge 20–4, 90, 94, 349

Mandelstam variables 180
manifold
  complex 153
  differentiable 153
  Riemannian 153
mass-shell condition
  from vertex operators 101, 106
  in BRST 131, 134, 137–8
  in light-cone quantization 17, 22, 26
matrix elements, relation to path integral 199–201, 332–4
matrix models 326–7, 352
measure on moduli space 154–65
  BRST invariance 160–1
  derivation 154–7
  gauge invariance 159–60
    in terms of
      determinants 157–8
      transition functions 161–4
  see also Faddeev–Popov
metric (spacetime) 9, 108–18
  dimensional reduction 231–3
  Einstein 115, 194, 275
  internal 249ff
  light-cone coordinates 16
  sigma model 115
  string 115
metric (world-line) 10
  gauge-fixing 129, 145–6
metric (world-sheet)
  constant curvature 292
  flat 32–3
  gauge-fixing 17, 84–6, 146–52
  induced 11, 15
  minimal area 300, 351
  noncritical string 94–5, 325–6
  Polyakov 12–13ff
mnemonics, zero-point 73–4, 125
Möbius group 56, 167
Möbius strip
  Riemann surface 208
  string amplitude 228–9, 280
Möbius transformation 178–80, 192, 198–9
mode (or Laurent) expansion
  bc CFT 61
  free scalar (X^μ) 26, 58
    periodic 235–8
    twisted 257
  holomorphic operator 56
  open strings 20, 57, 62–3
    Dirichlet 267
  T_{ab} 53, 70
model, solvable 323–7
modular group 148–9, 151
  and theta functions 214–16
  spacetime 253–5
  surfaces without 207–8, 223, 278
  used to analyze asymptotics 213, 224, 227–8
modular invariance 210, 217–18, *229–30*
  constraints from 213, 304, 350
  in orbifold compactification 257, 259
  in toroidal compactification 237, 251–3
  in twisting construction 260
  needed for consistency 220
  one-point function 304
moduli (spacetime) 233–4, 249, 265
  moduli space 253–5
    branches 262
    $c = 1$ 261–3, 351
    metric on 254
moduli (world-sheet)
  encoded in transition functions 152–4, 161–4
  for metric 147–52
  moduli space 148–52, 290–4
    boundary of 161, 295–6, 300; CFT asymptotics 302–3; divergences from 196, 224–6, 300, 305–9
    graphical decomposition 297–300, 351
    in open string field theory 313–4, 316–7
    measure on 154–65, 350
    total derivative on 161, 196, 220, 288
    vertex operator positions 150–1, 163–4

Nambu–Goto action 10–13, 15, 28, *30*
Narain compactification 251–5, 351
naturalness problem 2
Neumann boundary condition 14, *31*, 62, 174
  T-dual to Dirichlet 265–70
nilpotence *see* BRST
no-ghost theorem 137–43, 350
Noether's theorem 41–3, *74–5*
  anomalous 159
  as local property 86
non-Abelian gauge theory 188–92, 243
  from string field theory 313
  on D-brane 272–3
noncommutative geometry 272
noncritical string theory 325–7
nonlinear sigma model 108–10, *120*, 349
nonperturbative effects 234, 248, 269, 310, 315–7
  $\exp(-1/g_c)$ 317, 327
normal ordering
  boundary 75, 175
  conformal 36–40, 50, 60
  constants 22–3, 27, *76*
    direct calculation 60
    from state–operator mapping 72
    from Virasoro algebra 59, 61
    mnemonics 73, 125
  creation–annihilation 60
  diff-invariant 102–3
  relation between different forms 59–60, *76*
null state
  BRST 128, 136, 311
  decoupling *119*, 161, 196, 220, 285–95
  OCQ 122–5, *143*

$O(k,k,\mathbf{R})$ 252–3
$O(k,k,\mathbf{Z})$ 253
off-shell amplitudes 97, 103–4
old covariant quantization (OCQ) 121–5
  equivalence to BRST and light-cone 141–3
  vertex operators 125, 143, 160–1, 164, *165*
one-loop amplitudes 150, 206–30
  finiteness 220
  vs field theory 218–20
open string theory 14
  BRST quantization 131–6
  Dirichlet 266–70
  light-cone quantization 17–25
  must include closed strings 81–2, 225
  OCQ 122–5
  one-loop amplitudes 222–6, 228–9
  tree-level amplitudes 178–92
    relation to closed 195
  vertex operators 66, 107
operator calculations 201–2
operator equation 35–6, 333–4
operator product expansion (OPE) 36–41, *74–6*
  and conformal invariance 45–9
  and Ward identity 42–3
  associativity 70, 201, 304
  convergence 37–8
  derivation
    CFT 68–70
    free field theory 37–8
  generalized 302–3
  interchange symmetry 40–1
  Laurent expansion 70–1
  nonzero separation 48
  relation to commutator 54–7
  z-, $\bar{z}$-dependence 47
operator product expansions (specific)
  $\mathscr{A}_i \mathscr{A}_j$ 47
  $b\,b$, $b\,c$, $c\,c$ 50
  $\beta\,\beta$, $\beta\,\gamma$, $\gamma\,\gamma$ 52
  DDF operators 240
  $\partial X^\mu\, \partial X_\mu$ 211
  $:e^{ik\cdot X}::e^{ik'\cdot X}:$ 40, 251
  $:\mathscr{F}::\mathscr{G}:$ 39
  $j_B\, b$, $j_B\, c$ 132
  $j_B\, \mathcal{O}^m$ 132
  $j_B\, j_B$ 132

operator product expansions (specific) (*cont.*)
$j_B \mathscr{V}$ *165*
$j^i j^j$ 243
$j^\mu : e^{ik\cdot X}$: 43
$\mathcal{O}_i \mathcal{O}_j$ 200
$T \mathscr{A}$ 45–6
$T e^{ik\cdot X}$ 46
$T j$ 51
$T j_B$ 133
$T \mathcal{O}$ 46
$T T$ 47–8
$T \tilde{T}$ 47–8
$\mathscr{V} \mathscr{V}'$ 284, 289, *327*
$X_L X_L$, $X_L X_R$, $X_R X_R$ 238
$X^\mu X^\nu$ 38
orbifold 256–63, 351
fixed points 256–8
inheritance principle 258, 261
modular invariance 259–60
twisted states *31*, 256–61
vertex operators 258, 351
twisting construction 259–63
oriented string theory 29
see complement of unoriented string theory
orientifold 277–80
fixed planes 278–80
local physics oriented 278
outline of both volumes 8–9

$PSL(2, \mathbf{C})$ 168, 192, 197–9, *205*
$PSL(2, \mathbf{R})$ 168, 178, 199, *205*, 255, 292–3
$PSL(2, \mathbf{Z})$ 148, 255
parity
spacetime 6, 255, 277
world-sheet 29, 200, 255, 345
action on Chan–Paton factors 189–92
gauging 277
in one-loop amplitudes 207–8, 226–8
partition function 209
bc 212
general CFT 213
orbifold 258–61
scalar CFT 209–11
thermal 320–2, *328*
toroidal compactification 237, 250–2, *281*
path (functional) integrals 32–6
and expectation values 34
canonical commutators from 54–5, 58–61
Euclidean rotation 82–3, 170, 210, 335
evaluation
by integrating Weyl anomaly 93–4
by operator methods 201–2, 209–10, 212
direct 157–9, 169–72, 176–7, 337–41, 344–5
using holomorphicity 172–4, 177, 210–11
existence 332
fermionic 341–5
for traces 335–6, 344
Hamiltonian 331
introduction 329–46
Lagrangian 331
of total derivative 34, 333, 341
operator equations in 35, 333
Polyakov 82–4
gauge-fixing 84–90, 154–9
in curved spacetime 108–12
Weyl anomaly in 90–7
relation to operator expressions via cutting 55, 199–201, 300–5, 332–4
representation of string amplitudes 77–82, 97–101
twisted 260, 336
used to derive state–operator mapping 66–8
Pauli–Villars regulator 91, 340
period matrix 293
periodic boundary condition 14, *31*, 52
perturbation theory, string 5, 101, 283–310
physical state
BRST 127–9
b_0 condition 131, 134, 136, 289–90
vertex operators 164
OCQ 121–3
L_0 condition 125
vertex operators 125, 143, 160–1, 164
Planck length 3, 116
Planck mass 3, 245
plumbing fixture 294–305
Poincaré invariance 10, 12–16, 91, 185
and tadpoles 225
four-dimensional 16
of boundary conditions 14–15
point particle
action 9–11
and Wilson line 264
BRST quantization 129–31
in curved spacetime 108
light-cone quantization 17
path integral 77, 82–3, *164–5*
vacuum loop 145–6, *164*, 218, 220–2
Poisson resummation formula 237, 250, 252
Polyakov action *see* action, Polyakov
Polyakov path integral *see* path integral, Polyakov
potential energy
$c = 1$ moduli 246
from dilaton tadpoles 197–8, 225, 309
from quartic Yang–Mills interaction 265, 272
primary field *see* tensor operator
projective plane
expectation values 176, 178
Riemann surface 168–9
string amplitudes 196–8
tadpole 197–8, 227–8
cancellation with disk 229
propagator
dilaton 275–6
graviton 275–6
photon 290
point particle 115, *164–5*
pole 225–6, 307–8
string 297, 305–6, 313–14
proton decay 1
puncture 296–300

QCD string 25
quantization
canonical 17, 19–20, 26, 130, 249–50, 164
path integral 32–6, 77–82, 329–46
see also BRST, light-cone, OCQ, path integrals
quantum field theory 1–7, 104, 113, 310, 331–2
UV divergence 218–20
world-sheet 13–15, 32ff, 108–10, 118
quantum gravity 2–7, 25
quantum mechanics 1, 116, 329–46

Regge behavior 183–4, 193, 320, 352
Regge slope 11, 24, 112
renormalization 38, 102–3, 189, 332, 340
renormalization group equation 4, 91, 111
Ricci curvature
spacetime 111–12, 114
dimensional reduction 232–3
world-sheet 15, 85, 92–3
Riemann–Roch theorem 152, 158–9, 293
Riemann surface 100–1, 152–4, 161–4, 351
sewing 294–300, 351
specific 166–9, 206–8, 290–4
Riemann tensor 85

$SL(2, \mathbf{C})$ (algebra) 71
invariant state 71, 203
$SL(2, \mathbf{R})$ (algebra) 56, 71
$SO(D-2)$ 23–4, 27, *30*
$SO(D-1)$ 23–4, 27, *30–1*
$SO(n)$ 30, 191–2, 228
on D-brane 279–80
$SO(8192)$ 228–9, 280
$Sp(k)$ 30, 191–2, 228, 279
$SU(n)$ 189

$SU(2) \times SU(2)$ radius 242–8, 261–3
on D-brane 264–5, 269–72, 279
$SU(3) \times SU(2) \times U(1)$ model 1
see also Standard Model
S-matrix 97–101, 103–4
calculations 178–205
expression for 99, 155–7, 164
unitarity 181–2, 186–7, 193–4, 196, 285–90, 306
vs Green's function 104
s, *t*, *u* (Mandelstam variables) 180
saturating geodesic 300
scale transformation 44–8, 90–1
Schottky group 290–1, 351
Schwarzian 48
Schwinger–Dyson equation 334
self-contractions 103, 171, 338
self-dual radius 242–8, 261–3
sewing 332, 351
circles 298–9
construction of general surface from three-punctured spheres 296–7
of CFTs 300–5
and conformal bootstrap 303–5
of Riemann surfaces 294–300
parameter 294–5
used to analyze string amplitudes 305–7
Siegel gauge 313
sigma model metric 115
soliton 104, 235
sphere
expectation values 174–8, 198–204
Riemann surface 166–8
string amplitudes 192–6, *204–5*, 283–90
thrice-punctured 296–7, 303
cancellation with projective plane 229
spin 23–4, 184, *204*
in CFT 46, 213
in Coleman–Weinberg formula 221
spurious state 122
Standard Model 1–7, 25, 112, 310, 348–9
state–operator isomorphism 63–73, 101–2
explicit form 65
path integral derivation 66–8
vertex operators 101–2, 164, 199–201
Stirling's approximation 184, 315
string coupling 83–4
and dilaton 112–13
as overall factor in action 115
in vertex operators 101–2, 107, 179
position-dependent 323–4
relation between open and closed 83–4, 276
relation to gauge coupling 187, 234
relation to gravitational coupling 194, *205*
string field theory 310–17, 352
string length scale 25, 110, 244, 248
string metric 115
string tension 11, *30*
subtraction 39, 47, 60
supersymmetry 2, 6–7, 222, 283, 315, 348
supersymmetry breaking 222, 233–4

T-duality 247–8, *280–1*
and D-branes 268–73
as gauge symmetry 248
$O(k,k,\mathbf{Z})$ 253–5
open string theory 265–8
reverses winding and momentum 247, 268
unoriented string theory 277–80
tachyon
amplitudes 178–84, 187, 192–5, *204–5*, 217, *229–30*
artifact of bosonic string theory 23, 219, 223–4, 246, 275, 306
in BRST quantization 135
in light-cone quantization 23, 27, 29–30
in linear dilaton background 323–4, *328*
in OCQ 123
present in all bosonic string theories 117
vertex operator 101, 107
tadpole 197, 309
cancellation for $SO(8192)$ 229
disk 197, 223–6, 228–9
projective plane 197, 227–9
torus 307
target space 109
Teichmüller parameters 148
temperature 320–2, *328*
tensor operator (primary field) 46–7, *76*
expectation values 199
maps to highest weight state 71
OPEs 47, 200, 303
tetrad
spacetime 250–1
world-line 10, 129, 145–6
world-sheet 83
theta functions 214–16, *229*, 350
in expectation values 208–9, 211, 228, 259
toroidal compactification 231–55, 236–80, *280–1*, 351
abstract (Narain) description 251–4
in CFT 235–40
in field theory 231–5
in string theory 241–8
of several dimensions 249–55
of two dimensions 255
open string 263–8
torus
and partition function 209
expectation values 208–13
one-point function 304
in twisted CFT 260
Riemann surface 146–50, 154, 206–7
string amplitude 216–22, *229–30*, 237, 250
finiteness 218–20
trace
of energy-momentum tensor 13, 43, 92–7, 111
partition function 209, 335–6, 344, *345–6*
transition functions 153–4, 161–4
unoriented 277
transverse
Hilbert space 138
oscillators 20–1, 26, 125, 217–18
tree-level amplitudes 178–98, 283–90, 350–1
inheritance principle 258, 261
open–closed relation 195, 350
poles 181–2, 193–4, 196, 287–8
twisted fields *31*, 257, 259, 336, *345–6*
twisted state *31*, 257–61
twisting (orbifold) construction 259–63

$U(n)$ 30, 185, 188–9, *230*
ultraviolet (UV) divergences *see under* divergences
unit gauge 85
unitarity *see under* S-matrix
unitary CFT 71–3, 117, 137, 213
unoriented string theory 29–30, 80–1
Chan–Paton states 189–92
one-loop amplitudes 226–9
orientation-reversal as gauge symmetry 149, 153, 277–9
Riemann surfaces 80–1, 100–1
T-duality 277–80, 351

vacuum amplitude
field theory 218
physics of 220–2
string 198, 216–30, 274–6
vacuum degeneracy 7, 118, 233–4, 261, 348
Veneziano amplitude 178–84, *204*, 348
vertex operators 63–8, *74*, 97–102, 349
general construction 63–6
must be on mass-shell 101–3, 113
normalization 101–2, 107, 187, 194, 276, 284
OCQ 125, 143, 160–1, 164, *165*
Polyakov approach 102–7, *119*
positions 99–100
coincident limit 195–6, 306–9
in specific amplitudes 178–80, 192–3, *204*, 216–17

vertex operators (*cont.*)
positions (*cont.*)
relation to moduli and CKVs 149–52, 154–7, 163–4
specific 101–2, 107
twisted state 258, 351
winding state 238–9, 243–5, 251
Virasoro algebra 52–7, 75–6
Virasoro generators
as constraints 121–2, 133
centered on operators 70
$L_{-1}\cdot$ as derivative 71
$L_0\cdot$ as weight 71
Hermiticity 71
oscillator expressions 59, 61–2
winding state 238
Virasoro–Shapiro amplitude 192–5, *204–5*

Ward identity 41–3, 55
as local property 86
for conformal symmetry 45–6
weight (conformal) 46–7, 71–3
of vertex operators 101, 125, 143
positivity 72
Wess–Zumino consistency condition 96
Weyl anomaly *see under* anomaly
Weyl transformation 13, 15–16, *30*
as redundancy 13, 84, 117, 326
gauge-fixing of 17–18, 85
ghosts trivial 88, 131
of curvature 15, 85, 114
of nonlinear sigma model 110–12
of vertex operators 102–7
relation to conformal transformation 86
spacetime 114, 233, 254
Wick's theorem 60
Wilson line 263–8, 279–80
breaks gauge symmetry 265, 270
winding state *31*, 235–9
Dirichlet open string 266–7
massless as $R \to 0$ 247
vertex operators 238–9, 243–5, 251
world-sheet 13–14
world-sheet duality 287, 332
cylinder 226, 274
wrapped
D-brane 273
world-sheet 237, 250

Yukawa couplings 1, 79

Zamolodchikov inner product 200
zero modes
ghost 157–9, 161, 176, 212
X^μ 169–70, 173
zero-norm states *see* null states
zero-point energy 59–61, 72
heuristic 21–2, 26
mnemonics 73, 125
zero-slope limit 188

10

Type I and type II superstrings

Having spent volume one on a thorough development of the bosonic string, we now come to our real interest, the supersymmetric string theories. This requires a generalization of the earlier framework, enlarging the world-sheet constraint algebra. This idea arises naturally if we try to include spacetime fermions in the spectrum, and by guesswork we are led to *superconformal symmetry*. In this chapter we discuss the (1,1) superconformal algebra and the associated type I and II superstrings. Much of the structure is directly parallel to that of the bosonic string so we can proceed rather quickly, focusing on the new features.

10.1 The superconformal algebra

In bosonic string theory, the mass-shell condition

$$p_\mu p^\mu + m^2 = 0 \tag{10.1.1}$$

came from the physical state condition

$$L_0|\psi\rangle = 0\,, \tag{10.1.2}$$

and also from $\tilde{L}_0|\psi\rangle = 0$ in the closed string. The mass-shell condition is the Klein–Gordon equation in momentum space. To get spacetime fermions, it seems that we need the Dirac equation

$$ip_\mu \Gamma^\mu + m = 0 \tag{10.1.3}$$

instead. This is one way to motivate the following generalization, and it will lead us to all the known consistent string theories.

Let us try to follow the pattern of the bosonic string, where L_0 and $\tilde{L}_0$ are the center-of-mass modes of the world-sheet energy-momentum tensor $(T_B, \tilde{T}_B)$. A subscript B for 'bosonic' has been added to distinguish these from the fermionic currents now to be introduced. It seems then that we

need new conserved quantities T_F and $\tilde{T}_F$, whose center-of-mass modes give the Dirac equation, and which play the same role as T_B and $\tilde{T}_B$ in the bosonic theory. Noting further that the spacetime momenta p^μ are the center-of-mass modes of the world-sheet current $(\partial X^\mu, \bar{\partial} X^\mu)$, it is natural to guess that the gamma matrices, with algebra

$$\{\Gamma^\mu, \Gamma^\nu\} = 2\eta^{\mu\nu} , \tag{10.1.4}$$

are the center-of-mass modes of an anticommuting world-sheet field ψ^μ.

With this in mind, we consider the world-sheet action

$$S = \frac{1}{4\pi} \int d^2z \left(\frac{2}{\alpha'} \partial X^\mu \bar{\partial} X_\mu + \psi^\mu \bar{\partial} \psi_\mu + \tilde{\psi}^\mu \partial \tilde{\psi}_\mu \right) . \tag{10.1.5}$$

For reference we recall from chapter 2 the XX operator product expansion (OPE)

$$X^\mu(z,\bar{z}) X^\nu(0,0) \sim -\frac{\alpha'}{2} \eta^{\mu\nu} \ln |z|^2 . \tag{10.1.6}$$

The ψ conformal field theory (CFT) was described in section 2.5. The fields ψ^μ and $\tilde{\psi}^\mu$ are respectively holomorphic and antiholomorphic, and the operator products are

$$\psi^\mu(z)\psi^\nu(0) \sim \frac{\eta^{\mu\nu}}{z} , \quad \tilde{\psi}^\mu(\bar{z})\tilde{\psi}^\nu(0) \sim \frac{\eta^{\mu\nu}}{\bar{z}} . \tag{10.1.7}$$

The *world-sheet supercurrents*

$$T_F(z) = i(2/\alpha')^{1/2} \psi^\mu(z) \partial X_\mu(z) , \quad \tilde{T}_F(\bar{z}) = i(2/\alpha')^{1/2} \tilde{\psi}^\mu(\bar{z}) \bar{\partial} X_\mu(\bar{z}) \tag{10.1.8}$$

are also respectively holomorphic and antiholomorphic, since they are just the products of (anti)holomorphic fields. The annoying factors of $(2/\alpha')^{1/2}$ could be eliminated by working in units where $\alpha' = 2$, and then be restored if needed by dimensional analysis. Also, throughout this volume the : : normal ordering of coincident operators will be implicit.

This gives the desired result: the modes ψ_0^μ and $\tilde{\psi}_0^\mu$ will satisfy the gamma matrix algebra, and the centers-of-mass of T_F and $\tilde{T}_F$ will have the form of Dirac operators. We will see that the resulting string theory has spacetime fermions as well as bosons, and that the tachyon is gone.

From the OPE and the Ward identity it follows (exercise 10.1) that the currents

$$j^\eta(z) = \eta(z) T_F(z) , \quad \tilde{j}^\eta(\bar{z}) = \bar{\eta}(\bar{z}) \tilde{T}_F(\bar{z}) \tag{10.1.9}$$

generate the *superconformal transformation*

$$\epsilon^{-1}(2/\alpha')^{1/2} \delta X^\mu(z,\bar{z}) = -\eta(z)\psi^\mu(z) - \eta(z)^* \tilde{\psi}^\mu(\bar{z}) , \tag{10.1.10a}$$

$$\epsilon^{-1}(\alpha'/2)^{1/2} \delta \psi^\mu(z) = \eta(z) \partial X^\mu(z) , \tag{10.1.10b}$$

$$\epsilon^{-1}(\alpha'/2)^{1/2} \delta \tilde{\psi}^\mu(\bar{z}) = \eta(z)^* \bar{\partial} X^\mu(\bar{z}) . \tag{10.1.10c}$$

This transformation mixes the commuting field X^μ with the anticommuting fields ψ^μ and $\tilde{\psi}^\mu$, so the parameter $\eta(z)$ must be anticommuting. As with conformal symmetry, the parameters are arbitrary holomorphic or antiholomorphic functions. That this is a symmetry of the action (10.1.5) follows at once because the current is (anti)holomorphic, and so conserved.

The commutator of two superconformal transformations is a conformal transformation,

$$\delta_{\eta_1} \delta_{\eta_2} - \delta_{\eta_2} \delta_{\eta_1} = \delta_v , \quad v(z) = -2\eta_1(z)\eta_2(z) , \tag{10.1.11}$$

as the reader can check by acting on the various fields. Similarly, the commutator of a conformal and superconformal transformation is a superconformal transformation. The conformal and superconformal transformations thus close to form the *superconformal algebra*. In terms of the currents, this means that the OPEs of T_F with itself and with

$$T_B = -\frac{1}{\alpha'} \partial X^\mu \partial X_\mu - \frac{1}{2} \psi^\mu \partial \psi_\mu \tag{10.1.12}$$

close. That is, only T_B and T_F appear in the singular terms:

$$T_B(z)T_B(0) \sim \frac{3D}{4z^4} + \frac{2}{z^2} T_B(0) + \frac{1}{z} \partial T_B(0) , \tag{10.1.13a}$$

$$T_B(z)T_F(0) \sim \frac{3}{2z^2} T_F(0) + \frac{1}{z} \partial T_F(0) , \tag{10.1.13b}$$

$$T_F(z)T_F(0) \sim \frac{D}{z^3} + \frac{2}{z} T_B(0) , \tag{10.1.13c}$$

and similarly for the antiholomorphic currents. The $T_B T_F$ OPE implies that T_F is a tensor of weight $(\frac{3}{2}, 0)$. Each scalar contributes 1 to the central charge and each fermion $\frac{1}{2}$, for a total

$$c = (1 + \tfrac{1}{2})D = \tfrac{3}{2}D . \tag{10.1.14}$$

This enlarged algebra with T_F and $\tilde{T}_F$ as well as T_B and $\tilde{T}_B$ will play the same role that the conformal algebra did in the bosonic string. That is, we will impose it on the states as a constraint algebra — it must annihilate physical states, either in the sense of old covariant quantization (OCQ) or of Becchi–Rouet–Stora–Tyutin (BRST) quantization. Because of the Minkowski signature of spacetime the timelike ψ^0 and $\tilde{\psi}^0$, like X^0, have opposite sign commutators and lead to negative norm states. The fermionic constraints T_F and $\tilde{T}_F$ will remove these states from the spectrum.

More generally, the $N = 1$ superconformal algebra in operator product

form is

$$T_B(z)T_B(0) \sim \frac{c}{2z^4} + \frac{2}{z^2}T_B(0) + \frac{1}{z}\partial T_B(0)\,, \tag{10.1.15a}$$

$$T_B(z)T_F(0) \sim \frac{3}{2z^2}T_F(0) + \frac{1}{z}\partial T_F(0)\,, \tag{10.1.15b}$$

$$T_F(z)T_F(0) \sim \frac{2c}{3z^3} + \frac{2}{z}T_B(0)\,. \tag{10.1.15c}$$

The Jacobi identity requires the same constant c in the T_BT_B and T_FT_F products (exercise 10.5). Here, $N = 1$ refers to the number of $(\frac{3}{2},0)$ currents. In the present case there is also an antiholomorphic copy of the same algebra, so we have an $(N,\tilde{N}) = (1,1)$ *superconformal field theory (SCFT)*. We will consider more general algebras in section 11.1.

Free SCFTs

The various free CFTs described in chapter 2 have superconformal generalizations. One free SCFT combines an anticommuting bc theory with a commuting $\beta\gamma$ system, with weights

$$h_b = \lambda\,, \quad h_c = 1-\lambda\,, \tag{10.1.16a}$$

$$h_\beta = \lambda - \tfrac{1}{2}\,, \quad h_\gamma = \tfrac{3}{2} - \lambda\,. \tag{10.1.16b}$$

The action is

$$S_{BC} = \frac{1}{2\pi}\int d^2z\,(b\bar{\partial}c + \beta\bar{\partial}\gamma)\,, \tag{10.1.17}$$

and

$$T_B = (\partial b)c - \lambda\partial(bc) + (\partial\beta)\gamma - \frac{1}{2}(2\lambda-1)\partial(\beta\gamma)\,, \tag{10.1.18a}$$

$$T_F = -\frac{1}{2}(\partial\beta)c + \frac{2\lambda-1}{2}\partial(\beta c) - 2b\gamma\,. \tag{10.1.18b}$$

The central charge is

$$[-3(2\lambda-1)^2+1] + [3(2\lambda-2)^2-1] = 9 - 12\lambda\,. \tag{10.1.19}$$

Of course there is a corresponding antiholomorphic theory.

We can anticipate that the superconformal ghosts will be of this form with $\lambda = 2$, the anticommuting $(2,0)$ ghost b being associated with the commuting $(2,0)$ constraint T_B as in the bosonic theory, and the commuting $(\frac{3}{2},0)$ ghost β being associated with the anticommuting $(\frac{3}{2},0)$ constraint T_F. The ghost central charge is then $-26+11 = -15$, and the condition that the total central charge vanish gives the critical dimension

$$0 = \frac{3}{2}D - 15 \Rightarrow D = 10\,. \tag{10.1.20}$$

For $\lambda = 2$,

$$T_B = -(\partial b)c - 2b\partial c - \frac{1}{2}(\partial\beta)\gamma - \frac{3}{2}\beta\partial\gamma\,, \tag{10.1.21a}$$

$$T_F = -(\partial\beta)c - \frac{3}{2}\beta\partial c - 2b\gamma\,. \tag{10.1.21b}$$

Another free SCFT is the superconformal version of the linear dilaton theory. This has again the action (10.1.5), while

$$T_B(z) = -\frac{1}{\alpha'}\partial X^\mu\partial X_\mu + V_\mu\partial^2X^\mu - \frac{1}{2}\psi^\mu\partial\psi_\mu\,, \tag{10.1.22a}$$

$$T_F(z) = i(2/\alpha')^{1/2}\psi^\mu\partial X_\mu - i(2\alpha')^{1/2}V_\mu\partial\psi^\mu\,, \tag{10.1.22b}$$

each having an extra term as in the bosonic case. The reader can verify that these satisfy the $N = 1$ algebra with

$$c = \frac{3}{2}D + 6\alpha' V^\mu V_\mu\,. \tag{10.1.23}$$

10.2 Ramond and Neveu–Schwarz sectors

We now study the spectrum of the $X^\mu\psi^\mu$ SCFT on a circle. Much of this is as in chapter 2, but the new ingredient is a more general periodicity condition. It is clearest to start with the cylindrical coordinate $w = \sigma^1 + i\sigma^2$. The matter fermion action

$$\frac{1}{4\pi}\int d^2w\left(\psi^\mu\partial_{\bar{w}}\psi_\mu + \tilde{\psi}^\mu\partial_w\tilde{\psi}\right) \tag{10.2.1}$$

must be invariant under the periodic identification of the cylinder, $w \cong w + 2\pi$. This condition plus Lorentz invariance still allows two possible periodicity conditions for ψ^μ,

$$\text{Ramond (R):}\quad \psi^\mu(w+2\pi) = +\psi^\mu(w)\,, \tag{10.2.2a}$$

$$\text{Neveu–Schwarz (NS):}\quad \psi^\mu(w+2\pi) = -\psi^\mu(w)\,, \tag{10.2.2b}$$

where the sign must be the same for all μ. Similarly there are two possible periodicities for $\tilde{\psi}^\mu$. Summarizing, we will write

$$\psi^\mu(w+2\pi) = \exp(2\pi i\nu)\,\psi^\mu(w)\,, \tag{10.2.3a}$$

$$\tilde{\psi}^\mu(\bar{w}+2\pi) = \exp(-2\pi i\tilde{\nu})\,\tilde{\psi}^\mu(\bar{w})\,, \tag{10.2.3b}$$

where ν and $\tilde{\nu}$ take the values 0 and $\frac{1}{2}$.

Since we are initially interested in theories with the maximum Poincaré invariance, X^μ must be periodic. (Antiperiodicity of X^μ is interesting, and we have already encountered it for the twisted strings on an orbifold, but it would break some of the translation invariance.) The supercurrent then

has the same periodicity as the corresponding ψ,

$$T_F(w+2\pi) = \exp(2\pi i v)\, T_F(w)\,, \tag{10.2.4a}$$
$$\tilde{T}_F(\bar{w}+2\pi) = \exp(-2\pi i \tilde{v})\, \tilde{T}_F(\bar{w})\,. \tag{10.2.4b}$$

Thus there are four different ways to put the theory on a circle, each of which will lead to a different Hilbert space — essentially there are four different kinds of closed superstring. We will denote these by $(v,\tilde{v})$ or by NS–NS, NS–R, R–NS, and R–R. They are analogous to the twisted and untwisted sectors of the $\mathbf{Z}_2$ orbifold. Later in the chapter we will see that consistency requires that the full string spectrum contain certain combinations of states from each sector.

To study the spectrum in a given sector expand in Fourier modes,

$$\psi^\mu(w) = i^{-1/2}\sum_{r\in\mathbf{Z}+v} \psi_r^\mu \exp(irw)\,, \quad \tilde{\psi}^\mu(\bar{w}) = i^{1/2}\sum_{r\in\mathbf{Z}+\tilde{v}} \tilde{\psi}_r^\mu \exp(-ir\bar{w})\,, \tag{10.2.5}$$

the phase factors being inserted to conform to convention later. On each side the sum runs over integers in the R sector and over (integers $+\frac{1}{2}$) in the NS sector. Let us also write these as Laurent expansions. Besides replacing $\exp(-iw)\to z$ we must transform the fields,

$$\psi^\mu_{z^{1/2}}(z) = (\partial_z w)^{1/2}\psi^\mu_{w^{1/2}}(w) = i^{1/2}z^{-1/2}\psi^\mu_{w^{1/2}}(w)\,. \tag{10.2.6}$$

The clumsy subscripts are a reminder that these transform with half the weight of a vector. Henceforth the frame will be indicated implicitly by the argument of the field. The Laurent expansions are then

$$\psi^\mu(z) = \sum_{r\in\mathbf{Z}+v} \frac{\psi_r^\mu}{z^{r+1/2}}\,, \quad \tilde{\psi}^\mu(\bar{z}) = \sum_{r\in\mathbf{Z}+\tilde{v}} \frac{\tilde{\psi}_r^\mu}{\bar{z}^{r+1/2}}\,. \tag{10.2.7}$$

Notice that in the NS sector, the branch cut in $z^{-1/2}$ offsets the original antiperiodicity, while in the R sector it introduces a branch cut. Let us also recall the corresponding bosonic expansions

$$\partial X^\mu(z) = -i\left(\frac{\alpha'}{2}\right)^{1/2}\sum_{m=-\infty}^{\infty}\frac{\alpha_m^\mu}{z^{m+1}}\,, \quad \bar{\partial}X^\mu(\bar{z}) = -i\left(\frac{\alpha'}{2}\right)^{1/2}\sum_{m=-\infty}^{\infty}\frac{\tilde{\alpha}_m^\mu}{\bar{z}^{m+1}}\,, \tag{10.2.8}$$

where $\alpha_0^\mu = \tilde{\alpha}_0^\mu = (\alpha'/2)^{1/2}p^\mu$ in the closed string and $\alpha_0^\mu = (2\alpha')^{1/2}p^\mu$ in the open string.

The OPE and the Laurent expansions (or canonical quantization) give the anticommutators

$$\{\psi_r^\mu, \psi_s^\nu\} = \{\tilde{\psi}_r^\mu, \tilde{\psi}_s^\nu\} = \eta^{\mu\nu}\delta_{r,-s}\,, \tag{10.2.9a}$$
$$[\alpha_m^\mu, \alpha_n^\nu] = [\tilde{\alpha}_m^\mu, \tilde{\alpha}_n^\nu] = m\eta^{\mu\nu}\delta_{m,-n}\,. \tag{10.2.9b}$$

For T_F and T_B the Laurent expansions are

$$T_F(z) = \sum_{r\in\mathbf{Z}+v}\frac{G_r}{z^{r+3/2}}\,, \quad \tilde{T}_F(\bar{z}) = \sum_{r\in\mathbf{Z}+\tilde{v}}\frac{\tilde{G}_r}{\bar{z}^{r+3/2}}\,, \tag{10.2.10a}$$

$$T_B(z) = \sum_{m=-\infty}^{\infty}\frac{L_m}{z^{m+2}}\,, \quad \tilde{T}_B(\bar{z}) = \sum_{m=-\infty}^{\infty}\frac{\tilde{L}_m}{\bar{z}^{m+2}}\,. \tag{10.2.10b}$$

The usual CFT contour calculation gives the mode algebra

$$[L_m, L_n] = (m-n)L_{m+n} + \frac{c}{12}(m^3-m)\delta_{m,-n}\,, \tag{10.2.11a}$$
$$\{G_r, G_s\} = 2L_{r+s} + \frac{c}{12}(4r^2-1)\delta_{r,-s}\,, \tag{10.2.11b}$$
$$[L_m, G_r] = \frac{m-2r}{2}G_{m+r}\,. \tag{10.2.11c}$$

This is known as the *Ramond algebra* for r,s integer and the *Neveu–Schwarz algebra* for r,s half-integer. The antiholomorphic fields give a second copy of these algebras.

The superconformal generators in either sector are

$$L_m = \frac{1}{2}\sum_{n\in\mathbf{Z}}{:}\alpha_{m-n}^\mu\alpha_{\mu n}{:} + \frac{1}{4}\sum_{r\in\mathbf{Z}+v}(2r-m){:}\psi_{m-r}^\mu\psi_{\mu r}{:} + a^{\mathrm{m}}\delta_{m,0}\,, \tag{10.2.12a}$$

$$G_r = \sum_{n\in\mathbf{Z}}\alpha_n^\mu\psi_{\mu r-n}\,. \tag{10.2.12b}$$

Again $:\ :$ denotes creation–annihilation normal ordering. The normal ordering constant can be obtained by any of the methods from chapter 2; we will use here the mnemonic from the end of section 2.9. Each periodic boson contributes $-\frac{1}{24}$. Each periodic fermion contributes $+\frac{1}{24}$ and each antiperiodic fermion $-\frac{1}{48}$. Including the shift $+\frac{1}{24}c = \frac{1}{16}D$ gives

$$\mathrm{R}:\ a^{\mathrm{m}} = \frac{1}{16}D\,, \quad \mathrm{NS}:\ a^{\mathrm{m}} = 0\,. \tag{10.2.13}$$

For the open string, the condition that the surface term in the equation of motion vanish allows the possibilities

$$\psi^\mu(0,\sigma^2) = \exp(2\pi i v)\,\tilde{\psi}^\mu(0,\sigma^2)\,, \quad \psi^\mu(\pi,\sigma^2) = \exp(2\pi i v')\,\tilde{\psi}^\mu(\pi,\sigma^2)\,. \tag{10.2.14}$$

By the redefinition $\tilde{\psi}^\mu\to\exp(-2\pi i v')\tilde{\psi}^\mu$, we can set $v'=0$. There are therefore two sectors, R and NS, as compared to the four of the closed string. To write the mode expansion it is convenient to combine ψ^μ and $\tilde{\psi}^\mu$ into a single field with the extended range $0\le\sigma^1\le 2\pi$. Define

$$\psi^\mu(\sigma^1,\sigma^2) = \tilde{\psi}^\mu(2\pi-\sigma^1,\sigma^2) \tag{10.2.15}$$

for $\pi\le\sigma^1\le 2\pi$. The boundary condition $v'=0$ is automatic, and the antiholomorphicity of $\tilde{\psi}^\mu$ implies the holomorphicity of the extended ψ^μ.

Finally, the boundary condition (10.2.14) at $\sigma^1 = 0$ becomes a periodicity condition on the extended ψ^μ, giving one set of R or NS oscillators and the corresponding algebra.

NS and R spectra

We now consider the spectrum generated by a single set of NS or R modes, corresponding to the open string or to one side of the closed string. The NS spectrum is simple. There is no $r = 0$ mode, so we define the ground state to be annihilated by all $r > 0$ modes,

$$\psi_r^\mu |0\rangle_{\rm NS} = 0 \,, \quad r > 0 \,. \tag{10.2.16}$$

The modes with $r < 0$ then act as raising operators; since these are anticommuting, each mode can only be excited once.

The main point of interest is the R ground state, which is degenerate due to the ψ_0^μs. Define the ground states to be those that are annihilated by all $r > 0$ modes. The ψ_0^μ satisfy the Dirac gamma matrix algebra (10.1.4) with

$$\Gamma^\mu \cong 2^{1/2}\psi_0^\mu \,. \tag{10.2.17}$$

Since $\{\psi_r^\mu, \psi_0^\nu\} = 0$ for $r > 0$, the ψ_0^μ take ground states into ground states. The ground states thus form a representation of the gamma matrix algebra. This representation is worked out in section B.1; in $D = 10$ it has dimension 32. The reader who is not familiar with properties of spinors in various dimensions should read section B.1 at this point. We can take a basis of eigenstates of the Lorentz generators S_a, eq. (B.1.10):

$$|s_0, s_1, \ldots, s_4\rangle_{\rm R} \equiv |\mathbf{s}\rangle_{\rm R} \,, \quad s_a = \pm\tfrac{1}{2} \,. \tag{10.2.18}$$

The half-integral values show that these are indeed spacetime spinors. A more general basis for the spinors would be denoted $|\alpha\rangle_{\rm R}$. In the R sector of the open string not only the ground state but all states have half-integer spacetime spins, because the raising operators are vectors and change the S_a by integers. In the NS sector, the ground state is annihilated by $S^{\mu\nu}$ and is a Lorentz singlet, and all other states then have integer spin.

The Dirac representation **32** is reducible to two Weyl representations $\mathbf{16} + \mathbf{16}'$, distinguished by their eigenvalue under Γ as in eq. (B.1.11). This has a natural extension to the full string spectrum. The distinguishing property of Γ is that it anticommutes with all Γ^μ. Since the Dirac matrices are now the center-of-mass modes of ψ^μ, we need an operator that anticommutes with the full ψ^μ. We will call this operator

$$\exp(\pi i F) \,, \tag{10.2.19}$$

where F, the *world-sheet fermion number*, is defined only mod 2. Since ψ^μ changes F by one it anticommutes with the exponential. It is convenient

to write F in terms of spacetime Lorentz generators, which in either sector of the ψ CFT are

$$\Sigma^{\mu\lambda} = -\frac{i}{2} \sum_{r \in \mathbf{Z}+\nu} [\psi_r^\mu, \psi_{-r}^\lambda] \,. \tag{10.2.20}$$

This is the natural extension of the zero-mode part (B.1.8). Define now

$$S_a = i^{-\delta_{a,0}} \Sigma^{2a,2a+1} \,, \tag{10.2.21}$$

the i being included to make S_0 Hermitean, and let

$$F = \sum_{a=0}^{4} S_a \,. \tag{10.2.22}$$

This has the desired property. For example,

$$S_1(\psi_r^2 \pm i\psi_r^3) = (\psi_r^2 \pm i\psi_r^3)(S_1 \pm 1) \,, \tag{10.2.23}$$

so these oscillators change F by ± 1. The definition (10.2.22) makes it obvious that F is conserved by the OPE of the vertex operators, as a consequence of Lorentz invariance.[1] When we include the ghost part of the vertex operator in section 10.4, we will see that it contributes to the total F, so that on the total matter plus ghost ground state one has

$$\exp(\pi i F)|0\rangle_{\rm NS} = -|0\rangle_{\rm NS} \,, \tag{10.2.24a}$$

$$\exp(\pi i F)|\mathbf{s}\rangle_{\rm R} = |\mathbf{s}'\rangle_{\rm R}\Gamma_{\mathbf{s}'\mathbf{s}} \,. \tag{10.2.24b}$$

The ghost ground state contributes a factor -1 in the NS sector and $-i$ in the R sector.

Closed string spectra

In the closed string, the NS–NS states have integer spin. Because the spins S_a are additive, the half-integers from the two sides of the R–R sector also combine to give integer spin. The NS–R and R–NS states, on the other hand, have half-integer spin.

Let us look in more detail at the R–R sector, where the ground states $|\mathbf{s}, \mathbf{s}'\rangle_{\rm R}$ are degenerate on both the right and left. They transform as the product of two Dirac representations, which is worked out in section B.1:

$$\begin{aligned} \mathbf{32}_{\rm Dirac} \times \mathbf{32}_{\rm Dirac} &= [0] + [1] + [2] + \ldots + [10] \\ &= [0]^2 + [1]^2 + \ldots + [4]^2 + [5] \,, \end{aligned} \tag{10.2.25}$$

[1] Lorentz invariance of the OPE holds separately for the ψ and X CFTs (and the $\tilde{\psi}$ CFT in the closed string) because they are decoupled from one another. However, the world-sheet supercurrent is only invariant under the overall Lorentz transformation.

Table 10.1. *$SO(9,1)$ representations of massless R–R states.*

$(\exp(\pi i F), \exp(\pi i \tilde{F}))$			$SO(9,1)$ rep.
$(+1,+1)$:	$\mathbf{16} \times \mathbf{16}$	=	$[1] + [3] + [5]_+$
$(+1,-1)$:	$\mathbf{16} \times \mathbf{16}'$	=	$[0] + [2] + [4]$
$(-1,+1)$:	$\mathbf{16}' \times \mathbf{16}$	=	$[0] + [2] + [4]$
$(-1,-1)$:	$\mathbf{16}' \times \mathbf{16}'$	=	$[1] + [3] + [5]_-$

where $[n]$ denotes an antisymmetric rank n tensor. For the closed string there are separate world-sheet fermion numbers F and $\tilde{F}$, which on the ground states reduce to the *chirality* matrices Γ and $\tilde{\Gamma}$ acting on the two sides. The ground states thus decompose as in table 10.1.

10.3 Vertex operators and bosonization

Consider first the unit operator. Fields remain holomorphic at the origin, and in particular they are single-valued. From the Laurent expansion (10.2.7), the single-valuedness means that the unit operator must be in the NS sector; the conformal transformation that takes the incoming string to the point $z = 0$ cancels the branch cut from the antiperiodicity. The holomorphicity of ψ at the origin implies, via the contour argument, that the state corresponding to the unit operator satisfies

$$\psi_r^\mu |1\rangle = 0 , \quad r = \frac{1}{2}, \frac{3}{2}, \dots , \tag{10.3.1}$$

and therefore

$$|1\rangle = |0\rangle . \tag{10.3.2}$$

Since the $\psi\psi$ OPE is single-valued, all products of ψ and its derivatives must be in the NS sector. The contour argument gives the map

$$\psi_{-r}^\mu \to \frac{1}{(r-1/2)!} \partial^{r-1/2} \psi^\mu(0) , \tag{10.3.3}$$

so that there is a one-to-one map between such products and NS states. The analog of the Noether relation (2.9.6) between the superconformal variation of an NS operator and the OPE is

$$\delta_\eta \mathscr{A}(z,\bar{z}) = -\epsilon \sum_{n=0}^{\infty} \frac{1}{n!} \left[\partial^n \eta(z) G_{n-1/2} + (\partial^n \eta(z))^* \tilde{G}_{n-1/2} \right] \cdot \mathscr{A}(z,\bar{z}) . \tag{10.3.4}$$

The R sector vertex operators must be more complicated because the Laurent expansion (10.2.7) has a branch cut. We have encountered this before, for the winding state vertex operators in section 8.2 and the orbifold twisted state vertex operators in section 8.5. Each of these introduces a branch cut (the first a log and the second a square root) into X^μ. For the winding state vertex operators there was a simple expression as the exponential of a free field. For the twisted state vertex operators there was no simple expression and their amplitudes are determined only with more effort. Happily, through a remarkable property of two-dimensional field theory, the R sector vertex operators can be related directly to the bosonic winding state vertex operators.

Let $H(z)$ be the holomorphic part of a scalar field,

$$H(z)H(0) \sim -\ln z . \tag{10.3.5}$$

For world-sheet scalars not associated directly with the embedding of the string in spacetime this is the normalization we will always use, corresponding to $\alpha' = 2$ for the embedding coordinates. As in the case of the winding state vertex operators we can be cavalier about the location of the branch cut as long as the final expressions are single-valued. We will give a precise oscillator definition below. Consider the basic operators $e^{\pm iH(z)}$. These have the OPE

$$e^{iH(z)} e^{-iH(0)} \sim \frac{1}{z} , \tag{10.3.6a}$$

$$e^{iH(z)} e^{iH(0)} = O(z) , \tag{10.3.6b}$$

$$e^{-iH(z)} e^{-iH(0)} = O(z) . \tag{10.3.6c}$$

The poles and zeros in the OPE together with smoothness at infinity determine the expectation values of these operators on the sphere, up to an overall normalization which can be set to a convenient value:

$$\left\langle \prod_i e^{i\epsilon_i H(z_i)} \right\rangle_{S_2} = \prod_{i<j} z_{ij}^{\epsilon_i \epsilon_j} , \quad \sum_i \epsilon_i = 0 . \tag{10.3.7}$$

The ϵ_i are ± 1 here, but this result holds more generally.

Now consider the CFT of two Majorana–Weyl fermions $\psi^{1,2}(z)$, and form the complex combinations

$$\psi = 2^{-1/2}(\psi^1 + i\psi^2) , \quad \bar{\psi} = 2^{-1/2}(\psi^1 - i\psi^2) . \tag{10.3.8}$$

These have the properties

$$\psi(z)\bar{\psi}(0) \sim \frac{1}{z} , \tag{10.3.9a}$$

$$\psi(z)\psi(0) = O(z) , \tag{10.3.9b}$$

$$\bar{\psi}(z)\bar{\psi}(0) = O(z) . \tag{10.3.9c}$$

Eqs. (10.3.6) and (10.3.9) are identical in form, and so the expectation values of $\psi(z)$ on the sphere are identical to those of $e^{iH(z)}$. We will write

$$\psi(z) \cong e^{iH(z)} , \quad \bar{\psi}(z) \cong e^{-iH(z)} \tag{10.3.10}$$

to indicate this. Of course, all of this extends to the antiholomorphic case,

$$\tilde{\psi}(\bar{z}) \cong e^{i\tilde{H}(\bar{z})} , \quad \tilde{\bar{\psi}}(z) \cong e^{-i\tilde{H}(\bar{z})} . \tag{10.3.11}$$

Since arbitrary local operators with integral k_R and k_L can be formed by repeated operator products of $e^{\pm iH(z)}$ and $e^{\pm i\tilde{H}(\bar{z})}$, and arbitrary local operators built out of the fermions and their derivatives can be formed by repeated operator products of $\psi(z)$, $\bar{\psi}(z)$, $\tilde{\psi}(\bar{z})$, and $\tilde{\bar{\psi}}(\bar{z})$, the equivalence of the theories can be extended to all local operators. Finally, in order for these theories to be the same as CFTs, the energy-momentum tensors must be equivalent. The easiest way to show this is via the operator products

$$e^{iH(z)}e^{-iH(-z)} = \frac{1}{2z} + i\partial H(0) + 2zT_B^H(0) + O(z^2) , \tag{10.3.12a}$$

$$\psi(z)\bar{\psi}(-z) = \frac{1}{2z} + \psi\bar{\psi}(0) + 2zT_B^{\psi}(0) + O(z^2) . \tag{10.3.12b}$$

With the result (10.3.10), this implies equivalence of the H momentum current with the ψ number current, and of the two energy-momentum tensors,

$$\psi\bar{\psi} \cong i\partial H , \quad T_B^{\psi} \cong T_B^H . \tag{10.3.13}$$

As a check, e^{iH} and ψ are both $(\frac{1}{2}, 0)$ tensors.

In the operator description of the theory, define

$$\psi(z) \cong {:}e^{iH(z)}{:} . \tag{10.3.14}$$

From the Campbell–Baker–Hausdorff (CBH) formula (6.7.23) we have for equal times $|z| = |z'|$

$$\begin{aligned} {:}e^{iH(z)}{:}\,{:}e^{iH(z')}{:} &= \exp\{-[H(z), H(z')]\}\,{:}e^{iH(z')}{:}\,{:}e^{iH(z)}{:} \\ &= -{:}e^{iH(z')}{:}\,{:}e^{iH(z)}{:} , \end{aligned} \tag{10.3.15}$$

where we have used the fact (8.2.21) that at equal times $[H(z), H(z')] = \pm i\pi$. Thus the bosonized operators do anticommute. This is possible for operators constructed purely out of bosons because they are nonlocal. In particular, note that the CBH formula gives the equal time commutator

$$\begin{aligned} H(z){:}e^{iH(z')}{:} &= {:}e^{iH(z')}{:}\Big(H(z) + i[H(z), H(z')]\Big) \\ &= {:}e^{iH(z')}{:}\Big(H(z) - \pi\,\mathrm{sign}(\sigma_1 - \sigma_1')\Big) , \end{aligned} \tag{10.3.16}$$

so that the fermion field operator produces a kink, a discontinuity, in the bosonic field.

This rather surprising equivalence is known as *bosonization*. Equivalence between field theories with very different actions and fields occurs frequently in two dimensions, especially in CFTs because holomorphicity puts strong constraints on the theory. (The great recent surprise is that it is also quite common in higher-dimensional field and string theories.) Many interesting CFTs can be constructed in several different ways. One form or another will often be more useful for specific purposes. Notice that there is no simple correspondence between one-boson and one-fermion states. The current, for example, is linear in the boson field but quadratic in the fermion field. A single boson is the same as one ψ fermion and one $\bar{\psi}$ fermion at the same point. On a Minkowski world-sheet, where holomorphic becomes left-moving, the fermions both move left at the speed of light and remain coincident, indistinguishable from a free boson. A single fermion, on the other hand, is created by an operator exponential in the boson field and so is a coherent state, which as we have seen is in the shape of a kink (10.3.16).

The complicated relationship between the bosonic and fermionic spectra shows up also in the partition function. Operator products of $e^{\pm iH(z)}$ generate all operators with integer k_L. The bosonic momentum and oscillator sums then give

$$\mathrm{Tr}\,(q^{L_0}) = \left(\sum_{k_L \in \mathbf{Z}} q^{k_L^2/2}\right) \prod_{n=1}^{\infty} (1 - q^n)^{-1} . \tag{10.3.17}$$

In the NS sector of the fermionic theory, the oscillator sum gives

$$\mathrm{Tr}\,(q^{L_0}) = \prod_{n=1}^{\infty} (1 + q^{n-1/2})^2 . \tag{10.3.18}$$

We know indirectly that these must be equal, since we can use the OPE to construct an analog in the fermionic theory for any local operator of the bosonic theory and vice versa. Expanding the products gives

$$1 + 2q^{1/2} + q + 2q^{3/2} + 4q^2 + 4q^{5/2} + \ldots \tag{10.3.19}$$

for each, and in fact the equality of (10.3.17) and (10.3.18) follows from the equality of the product and sum expressions for theta functions, section 7.2. Note that while bosonization was derived for the sphere, the sewing construction from chapter 9 guarantees that it holds on all Riemann surfaces, provided that we make equivalent projections on the spectra. In particular, we have seen that summing over integer k_L corresponds to summing over all local fermionic operators, the NS sector.

Bosonization extends readily to the R sector. In fact, once we combine two fermions into a complex pair we can consider the more general periodicity condition

$$\psi(w + 2\pi) = \exp(2\pi i v)\,\psi(w) \tag{10.3.20}$$

for any real v. In ten dimensions only $v = 0, \frac{1}{2}$ arose, but these more general periodicities are important in less symmetric situations. The Laurent

expansion has the same form (10.2.7) as before,

$$\psi(z) = \sum_{r\in\mathbf{Z}+\nu} \frac{\psi_r}{z^{r+1/2}} , \quad \bar\psi(z) = \sum_{s\in\mathbf{Z}-\nu} \frac{\bar\psi_s}{z^{s+1/2}} , \tag{10.3.21}$$

with indices displaced from integers by $\pm\nu$. The algebra is

$$\{\psi_r, \bar\psi_s\} = \delta_{r,-s} . \tag{10.3.22}$$

Define a reference state $|0\rangle_\nu$ by

$$\psi_{n+\nu}|0\rangle_\nu = \bar\psi_{n+1-\nu}|0\rangle_\nu = 0 , \quad n = 0, 1, \ldots . \tag{10.3.23}$$

The first nonzero terms in the Laurent expansions are then $r = -1+\nu$ and $s = -\nu$, so for the corresponding local operator $\mathscr{A}_\nu$ the OPE is

$$\psi(z)\mathscr{A}_\nu(0) = O(z^{-\nu+1/2}) , \quad \bar\psi(z)\mathscr{A}_\nu(0) = O(z^{\nu-1/2}) . \tag{10.3.24}$$

The conditions (10.3.23) uniquely identify the state $|0\rangle_\nu$, and so the corresponding OPEs (10.3.24) determine the bosonic equivalent

$$\exp[i(-\nu+1/2)H] \cong \mathscr{A}_\nu . \tag{10.3.25}$$

One can check the identification (10.3.25) by verifying that the weight is $h = \frac{1}{2}(\nu - \frac{1}{2})^2$. In the bosonic form this comes from the term $\frac{1}{2}p^2$ in L_0. In the fermionic form it follows from the usual commutator method (2.7.8) or the zero-point mnemonic.

The boundary condition (10.3.20) is the same for ν and $\nu+1$, but the reference state that we have defined is not. It is a ground state only for $0 \le \nu \le 1$. As we vary ν, the state $|0\rangle_\nu$ changes continuously, and when we get back to the original theory at $\nu+1$, by the definition (10.3.23) it has become the excited state

$$|0\rangle_{\nu+1} = \bar\psi_{-\nu}|0\rangle_\nu . \tag{10.3.26}$$

This is known as *spectral flow*. For the R case $\nu = 0$ there are the two degenerate ground states

$$|s\rangle \cong e^{isH}, \quad s = \pm\tfrac{1}{2} . \tag{10.3.27}$$

For the superstring in ten dimensions we need five bosons, H^a for $a = 0, \ldots, 4$. Then[2]

$$2^{-1/2}(\pm\psi^0 + \psi^1) \cong e^{\pm iH^0} \tag{10.3.28a}$$

$$2^{-1/2}(\psi^{2a} \pm i\psi^{2a+1}) \cong e^{\pm iH^a} , \quad a = 1, \ldots, 4 . \tag{10.3.28b}$$

[2] The precise operator definition has a subtlety when there are several species of fermion. The H^a for different a are independent and so the exponentials commute rather than anticommute. A cocycle is needed, as in eq. (8.2.22). A general expression will be given in the next section.

The vertex operator $\Theta_\mathbf{s}$ for an R state $|\mathbf{s}\rangle$ is

$$\Theta_\mathbf{s} \cong \exp\Big[i\sum_a s_a H^a\Big] . \tag{10.3.29}$$

This operator, which produces a branch cut in ψ^μ, is sometimes called a *spin field*. For closed string states, this is combined with the appropriate antiholomorphic vertex operator, built from $\tilde H^a$.

The general bc CFT, renaming $\psi \to b$ and $\bar\psi \to c$, is obtained by modifying the energy-momentum tensor of the $\lambda = \frac{1}{2}$ theory to

$$T_B^{(\lambda)} = T_B^{(1/2)} - (\lambda - \tfrac{1}{2})\partial(bc) . \tag{10.3.30}$$

The equivalences (10.3.13) give the corresponding bosonic operator

$$T_B^{(\lambda)} \cong T_B^H - i(\lambda - \tfrac{1}{2})\partial^2 H . \tag{10.3.31}$$

This is the same as the linear dilaton CFT, with $V = -i(\lambda - \frac{1}{2})$. With this correspondence between V and λ, the linear dilaton and bc theories are equivalent,

$$b \cong e^{iH} , \quad c \cong e^{-iH} . \tag{10.3.32}$$

As a check, the central charges agree,

$$c = 1 - 3(2\lambda - 1)^2 = 1 + 12V^2 . \tag{10.3.33}$$

So do the dimensions of the fields (10.3.32), λ for b and $1-\lambda$ for c, agreeing with $k^2/2 + ikV$ for e^{ikH}. The nontensor behaviors of the currents bc and $i\partial H$ are also the same. Since the inner product for the reparameterization ghosts makes b and c Hermitean, the bosonic field H must be anti-Hermitean in this application. The bosonization of the ghosts is usually written in terms of a Hermitean field with the opposite sign OPE,

$$H \to i\rho ; \quad c \cong e^{\rho} , \quad b \cong e^{-\rho} . \tag{10.3.34}$$

10.4 The superconformal ghosts

To build the BRST current we will need, in addition to the anticommuting b and c ghosts of the bosonic string, commuting ghost fields β and γ of weight $(\frac{3}{2}, 0)$ and $(-\frac{1}{2}, 0)$, and the corresponding antiholomorphic fields. The action for this SCFT was given in eq. (10.1.17) and the currents T_B and T_F in eq. (10.1.21). The ghosts β and γ must have the same periodicity (10.2.4) as the generator T_F with which they are associated. This is necessary to make the BRST current periodic, so that it can be

integrated to give the BRST charge. Thus,

$$\beta(z) = \sum_{r\in\mathbf{Z}+\nu} \frac{\beta_r}{z^{r+3/2}}\,, \quad \gamma(z) = \sum_{r\in\mathbf{Z}+\nu} \frac{\gamma_r}{z^{r-1/2}}\,, \tag{10.4.1a}$$

$$b(z) = \sum_{m=-\infty}^{\infty} \frac{b_m}{z^{m+2}}\,, \quad c(z) = \sum_{m=-\infty}^{\infty} \frac{c_m}{z^{m-1}}\,, \tag{10.4.1b}$$

and similarly for the antiholomorphic fields. The (anti)commutators are

$$[\gamma_r, \beta_s] = \delta_{r,-s}\,, \quad \{b_m, c_n\} = \delta_{n,-m}\,. \tag{10.4.2}$$

Define the ground states $|0\rangle_{\mathrm{NS,R}}$ by

$$\beta_r|0\rangle_{\mathrm{NS}} = 0\,,\ r \geq \tfrac{1}{2}\,, \quad \gamma_r|0\rangle_{\mathrm{NS}} = 0\,,\ r \geq \tfrac{1}{2} \tag{10.4.3a}$$

$$\beta_r|0\rangle_{\mathrm{R}} = 0\,,\ r \geq 0\,, \quad \gamma_r|0\rangle_{\mathrm{R}} = 0\,,\ r \geq 1\,, \tag{10.4.3b}$$

$$b_m|0\rangle_{\mathrm{NS,R}} = 0\,,\ m \geq 0\,, \quad c_m|0\rangle_{\mathrm{NS,R}} = 0\,,\ m \geq 1\,. \tag{10.4.3c}$$

We have grouped β_0 with the lowering operators and γ_0 with the raising ones, in parallel with the bosonic case. The spectrum is built as usual by acting on the ground states with the raising operators. The generators are

$$L_m^{\mathrm{g}} = \sum_{n\in\mathbf{Z}} (m+n)\,{:}b_{m-n}c_n{:} + \sum_{r\in\mathbf{Z}+\nu} \frac{1}{2}(m+2r)\,{:}\beta_{m-r}\gamma_r{:} + a^{\mathrm{g}}\delta_{m,0}\,, \tag{10.4.4a}$$

$$G_r^{\mathrm{g}} = -\sum_{n\in\mathbf{Z}} \left[\frac{1}{2}(2r+n)\beta_{r-n}c_n + 2b_n\gamma_{r-n}\right]. \tag{10.4.4b}$$

The normal ordering constant is determined by the usual methods to be

$$\mathrm{R}\colon\ a^{\mathrm{g}} = -\frac{5}{8}\,, \quad \mathrm{NS}\colon\ a^{\mathrm{g}} = -\frac{1}{2}\,. \tag{10.4.5}$$

Vertex operators

We focus here on the $\beta\gamma$ CFT, as the bc parts of the vertex operators are already understood. Let us start by considering the state corresponding to the unit operator. From the Laurent expansions (10.4.1) it is in the NS sector and satisfies

$$\beta_r|1\rangle = 0\,,\ r \geq -\frac{1}{2}\,, \quad \gamma_r|1\rangle = 0\,,\ r \geq \frac{3}{2}\,. \tag{10.4.6}$$

This is not the same as the ground state $|0\rangle_{\mathrm{NS}}$: the mode $\gamma_{1/2}$ annihilates $|0\rangle_{\mathrm{NS}}$ while its conjugate $\beta_{-1/2}$ annihilates $|1\rangle$. We found this also for the bc ghosts with c_1 and b_{-1}. Since anticommuting modes generate just two states, we had the simple relation $|0\rangle = c_1|1\rangle$ (focusing on the holomorphic side). For commuting oscillators things are not so simple: there is no state

in the Fock space built on $|1\rangle$ by acting with $\gamma_{1/2}$ that has the properties of $|0\rangle_{\mathrm{NS}}$. The definition of the state $|0\rangle_{\mathrm{NS}}$ translates into

$$\gamma(z)\delta(\gamma(0)) = O(z)\,, \quad \beta(z)\delta(\gamma(0)) = O(z^{-1})\,, \tag{10.4.7}$$

for the corresponding operator $\delta(\gamma)$. The notation $\delta(\gamma)$ reflects the fact that the field γ has a simple zero at the vertex operator. Recall that for the bc ghosts the NS ground state maps to the operator c, which is the anticommuting analog of a delta function. One can show that an insertion of $\delta(\gamma)$ in the path integral has the property (10.4.7).

To give an explicit description of this operator it is again convenient to bosonize. Of course β and γ are already bosonic, but bosonization here refers to a rewriting of the theory in a way that is similar to, but a bit more intricate than, the bosonization of the anticommuting bc theory.

Start with the current $\beta\gamma$. The operator product

$$\beta\gamma(z)\,\beta\gamma(0) \sim -\frac{1}{z^2} \tag{10.4.8}$$

is the same as that of $\partial\phi$, where $\phi(z)\phi(0) \sim -\ln z$ is a holomorphic scalar. Holomorphicity then implies that this equivalence extends to all correlation functions,

$$\beta\gamma(z) \cong \partial\phi(z)\,. \tag{10.4.9}$$

The OPE of the current with β and γ then suggests

$$\beta(z) \stackrel{?}{\cong} e^{-\phi(z)}\,, \quad \gamma(z) \stackrel{?}{\cong} e^{\phi(z)}\,. \tag{10.4.10}$$

For the bc system we would be finished: this approach leads to the same bosonization as before. For the $\beta\gamma$ system, however, the sign of the current–current OPE and therefore of the $\phi\phi$ OPE is changed. The would-be bosonization (10.4.10) gives the wrong OPEs: it would imply

$$\beta(z)\beta(0) \stackrel{?}{=} O(z^{-1})\,, \quad \beta(z)\gamma(0) \stackrel{?}{=} O(z^{1})\,, \quad \gamma(z)\gamma(0) \stackrel{?}{=} O(z^{-1})\,, \tag{10.4.11}$$

whereas the correct OPE is

$$\beta(z)\beta(0) = O(z^{0})\,, \quad \beta(z)\gamma(0) = O(z^{-1})\,, \quad \gamma(z)\gamma(0) = O(z^{0})\,. \tag{10.4.12}$$

To repair this, additional factors are added,

$$\beta(z) \cong e^{-\phi(z)}\partial\xi(z)\,, \quad \gamma \cong e^{\phi(z)}\eta(z)\,. \tag{10.4.13}$$

In order not to spoil the OPE with the current (10.4.9), the new fields $\eta(z)$ and $\xi(z)$ must be nonsingular with respect to ϕ, which means that the $\eta\xi$ theory is a new CFT, decoupled from the ϕ CFT. Further, the equivalence (10.4.13) will hold — all OPEs will be correct — if η and ξ satisfy

$$\eta(z)\xi(0) \sim \frac{1}{z}\,, \quad \eta(z)\eta(0) = O(z)\,, \quad \partial\xi(z)\partial\xi(0) = O(z)\,. \tag{10.4.14}$$

This identifies the $\eta\xi$ theory as a holomorphic CFT of the bc type: the OPE of like fields has a zero due to the anticommutativity.

It remains to study the energy-momentum tensor. We temporarily consider the general $\beta\gamma$ system, with β having weight λ'. The OPE

$$T(z)\beta\gamma(0) = \frac{1-2\lambda'}{2z^3} + \dots \tag{10.4.15}$$

determines the ϕ energy-momentum tensor,

$$T_B^\phi = -\frac{1}{2}\partial\phi\partial\phi + \frac{1}{2}(1-2\lambda')\partial^2\phi \,. \tag{10.4.16}$$

The exponentials in the bosonization (10.4.13) thus have weights $\lambda'-1$ and $-\lambda'$ respectively, as compared with the weights λ' and $1-\lambda'$ of β and γ. This fixes the weights of η and ξ as 1 and 0: this is a $\lambda = 1$ bc system, with

$$T_B^{\eta\xi} = -\eta\partial\xi \tag{10.4.17}$$

and

$$T_B^{\beta\gamma} \cong T_B^\phi + T_B^{\eta\xi} \,. \tag{10.4.18}$$

As a check, the central charges are $3(2\lambda'-1)^2+1$ for T_B^ϕ and -2 for $T_B^{\eta\xi}$, adding to the $3(2\lambda'-1)^2-1$ of the $\beta\gamma$ CFT. The need for extra degrees of freedom is not surprising. The $\beta\gamma$ theory has a greater density of states than the bc theory because the modes of a commuting field can be excited any number of times. One can check that the total partition functions agree, in the appropriate sectors.

If need be one can go further and represent the $\eta\xi$ theory in terms of a free boson, conventionally χ with $\chi(z)\chi(0) \sim \ln z$, as in the previous section. Thus

$$\eta \cong e^{-\chi} \,, \quad \xi \cong e^{\chi} \,, \tag{10.4.19a}$$

$$\beta \cong e^{-\phi+\chi}\partial\chi \,, \quad \gamma \cong e^{\phi-\chi} \,. \tag{10.4.19b}$$

The energy-momentum tensor is then

$$T_B = -\frac{1}{2}\partial\phi\partial\phi + \frac{1}{2}\partial\chi\partial\chi + \frac{1}{2}(1-2\lambda')\partial^2\phi + \frac{1}{2}\partial^2\chi \,. \tag{10.4.20}$$

For the string, the relevant value is $\lambda' = \frac{3}{2}$. The properties (10.4.7) of $\delta(\gamma)$ determine the bosonization,

$$\delta(\gamma) \cong e^{-\phi} \,, \quad h = \frac{1}{2} \,. \tag{10.4.21}$$

The fermionic parts of the tachyon and massless NS vertex operators are then

$$e^{-\phi} \,, \quad e^{-\phi}e^{\pm iH^a} \tag{10.4.22}$$

respectively. For $\lambda' = \frac{3}{2}$, the exponential $e^{l\phi}$ has weight $-\frac{1}{2}l^2 - l$.

The operator Σ corresponding to $|0\rangle_R$ satisfies

$$\beta(z)\Sigma(0) = O(z^{-1/2}) \,, \quad \gamma(z)\Sigma(0) = O(z^{1/2}) \,. \tag{10.4.23}$$

This determines

$$\Sigma = e^{-\phi/2} \,, \quad h = \frac{3}{8} \,. \tag{10.4.24}$$

Adding the contribution -1 of the bc ghosts, the weight of $e^{-\phi/2}$ and of $e^{-\phi}$ agree with the values (10.4.5). The R ground state vertex operators are then

$$\mathscr{V}_s = e^{-\phi/2}\Theta_s \,, \tag{10.4.25}$$

with the spin field Θ_s having been defined in eq. (10.3.29).

We need to extend the definition of world-sheet fermion number F to be odd for β and γ. The ultimate reason is that it anticommutes with the supercurrent T_F and we will need it to commute with the BRST operator, which contains terms such as γT_F. The natural definition for F is then that it be the charge associated with the current (10.4.9), which is l for $e^{l\phi}$. Again, it is conserved by the OPE. This accounts for the ghost contributions in eq. (10.2.24). Note that this definition is based on spin rather than statistics, since the ghosts have the wrong spin-statistics relation; it would therefore be more appropriate to call F the *world-sheet spinor number*.

For completeness we give a general expression for the cocycle for exponentials of free fields, though we emphasize that for most purposes the details are not necessary. In general one has operators

$$\exp(ik_L \cdot H_L + ik_R \cdot H_R) \,, \tag{10.4.26}$$

with the holomorphic and antiholomorphic scalars not necessarily equal in number. The momenta k take values in some lattice Γ. The naive operator product is proportional to $z^{-k\circ k'}$, and for all pairs in Γ, $k \circ k'$ must be an integer. The notation is as in section 8.4, $k\circ k' = k_L\cdot k'_L - k_R\cdot k'_R$. When $k \circ k'$ is an odd integer the vertex operators anticommute rather than commute. A correctly defined vertex operator is

$$C_k(\alpha_0)\,{:}\exp(ik_L \cdot H_L + ik_R \cdot H_R){:} \tag{10.4.27}$$

with the cocycle C_k defined as follows. Take a set of basis vectors k_α for Γ; that is, Γ consists of the integer linear combinations $n_\alpha k_\alpha$. Similarly write the vector of zero-mode operators in this basis, $\alpha_0 = \alpha_{0\alpha}k_\alpha$, Then for $k = n_\alpha k_\alpha$,

$$C_k(\alpha_0) = \exp\Big(\pi i \sum_{\alpha>\beta} n_\alpha \alpha_{0\beta} k_\alpha \circ k_\beta\Big) \,. \tag{10.4.28}$$

This generalizes the simple case (8.2.22). The reader can check that vertex operators with even $k \circ k$ now commute with all vertex operators, and those with odd $k \circ k$ anticommute among themselves. Note that a cocycle has no effect on the commutativity of a vertex operator with itself, so an exponential must be bosonic if $k \circ k$ is even and fermionic if $k \circ k$ is odd.

10.5 Physical states

In the bosonic string we started with a (diff×Weyl)-invariant theory. After fixing to conformal gauge we had to impose the vanishing of the conformal algebra as a constraint on the states. In the present case there is an analogous gauge-invariant form, and the superconformal algebra emerges as a constraint in the gauge-fixed theory. However, it is not necessary to proceed in this way, and it would require us to develop some machinery that in the end we do not need. Rather we can generalize directly in the gauge-fixed form, defining the superconformal symmetry to be a constraint and proceeding in parallel to the bosonic case to construct a consistent theory. We will first impose the constraint in the old covariant formalism, and then in the BRST formalism.

OCQ

In this formalism, developed for the bosonic string in section 4.1, one ignores the ghost excitations. We begin with the open string, imposing the physical state conditions

$$L_n^{\mathrm{m}}|\psi\rangle = 0\,,\ n>0\,,\quad G_r^{\mathrm{m}}|\psi\rangle = 0\,,\ r\geq 0\,. \tag{10.5.1}$$

Only the matter part of any state is nontrivial — the ghosts are in their ground state — and the superscript 'm' denotes the matter part of each generator. There are also the equivalence relations

$$L_n^{\mathrm{m}}|\chi\rangle \cong 0\,,\ n<0\,,\quad G_r^{\mathrm{m}}|\chi\rangle \cong 0\,,\ r<0\,. \tag{10.5.2}$$

The mass-shell condition can always be written in terms of the total matter plus ghost Virasoro generator, which is the same as the world-sheet Hamiltonian H because the total central charge is zero:

$$L_0|\psi\rangle = H|\psi\rangle = 0\,. \tag{10.5.3}$$

In ten flat dimensions this is

$$H = \begin{cases} \alpha' p^2 + N - \dfrac{1}{2} & (\mathrm{NS}) \\ \alpha' p^2 + N & (\mathrm{R}) \end{cases}\,. \tag{10.5.4}$$

The zero-point constants from the ghosts and longitudinal oscillators have canceled as usual, leaving the contribution of the transverse modes,

$$\mathrm{NS}:\ 8\left(-\frac{1}{24}-\frac{1}{48}\right) = -\frac{1}{2}\,,\quad \mathrm{R}:\ 8\left(-\frac{1}{24}+\frac{1}{24}\right) = 0\,. \tag{10.5.5}$$

For the tachyonic and massless levels we need only the terms

$$G_0^{\mathrm{m}} = (2\alpha')^{1/2} p_\mu \psi_0^\mu + \dots\,, \tag{10.5.6a}$$

$$G_{\pm 1/2}^{\mathrm{m}} = (2\alpha')^{1/2} p_\mu \psi_{\pm 1/2}^\mu + \dots\,. \tag{10.5.6b}$$

The NS sector works out much as in the bosonic string. The lowest state is $|0;k\rangle_{\mathrm{NS}}$, labeled by the matter state and momentum. The only nontrivial condition is from L_0, giving

$$m^2 = -k^2 = -\frac{1}{2\alpha'}\,. \tag{10.5.7}$$

This state is a tachyon. It has $\exp(\pi i F) = -1$, where F was given in eq. (10.2.24). The first excited state is

$$|e;k\rangle_{\mathrm{NS}} = e\cdot\psi_{-1/2}|0;k\rangle_{\mathrm{NS}}\,. \tag{10.5.8}$$

The nontrivial physical state conditions are

$$0 = L_0|e;k\rangle_{\mathrm{NS}} = \alpha' k^2 |e;k\rangle_{\mathrm{NS}}\,, \tag{10.5.9a}$$

$$0 = G_{1/2}^{\mathrm{m}}|e;k\rangle_{\mathrm{NS}} = (2\alpha')^{1/2} k\cdot e|0;k\rangle_{\mathrm{NS}}\,, \tag{10.5.9b}$$

while

$$G_{-1/2}^{\mathrm{m}}|0;k\rangle_{\mathrm{NS}} = (2\alpha')^{1/2} k\cdot\psi_{-1/2}|0;k\rangle_{\mathrm{NS}} \tag{10.5.10}$$

is null. Thus

$$k^2 = 0\,,\quad e\cdot k = 0\,,\quad e^\mu \cong e^\mu + k^\mu\,. \tag{10.5.11}$$

This state is massless, the half-unit of excitation canceling the zero-point energy, and has $\exp(\pi i F) = +1$. Like the first excited state of the bosonic string it is a massless vector, with $D-2$ spacelike polarizations. The constraints have removed the unphysical polarizations of ψ^μ, just as for X^μ in the bosonic case.

In the R sector the lowest states are

$$|u;k\rangle_{\mathrm{R}} = |\mathbf{s};k\rangle_{\mathrm{R}} u_{\mathbf{s}}\,. \tag{10.5.12}$$

Here $u_{\mathbf{s}}$ is the polarization, and the sum on $\mathbf{s}$ is implicit. The nontrivial physical state conditions are

$$0 = L_0|u;k\rangle_{\mathrm{R}} = \alpha' k^2 |u;k\rangle_{\mathrm{R}}\,, \tag{10.5.13a}$$

$$0 = G_0^{\mathrm{m}}|u;k\rangle_{\mathrm{R}} = \alpha'^{1/2}|\mathbf{s}';k\rangle_{\mathrm{R}} k\cdot\Gamma_{\mathbf{s}'\mathbf{s}} u_{\mathbf{s}}\,. \tag{10.5.13b}$$

Table 10.2. *Massless and tachyonic open string states.*

sector	$SO(8)$ spin	m^2
NS+	$\mathbf{8}_v$	0
NS−	$\mathbf{1}$	$-1/2\alpha'$
R+	$\mathbf{8}$	0
R−	$\mathbf{8}'$	0

The ground states are massless because the zero-point energy vanishes in the R sector. The $G_0^{\rm m}$ condition gives the massless Dirac equation

$$k \cdot \Gamma_{\mathrm{s's}} u_{\mathrm{s}} = 0 \,, \tag{10.5.14}$$

which was our original goal in introducing the superconformal algebra. The $G_0^{\rm m}$ condition implies the L_0 condition, because $G_0^2 = L_0$ in the critical dimension and the ghost parts of G_0 annihilate the ghost vacuum.

In ten dimensions, massless particle states are classified by their behavior under the $SO(8)$ rotations that leave the momentum invariant. Take a frame with $k^0 = k^1$. In the NS sector, the massless physical states are the eight transverse polarizations forming the vector representation $\mathbf{8}_v$ of $SO(8)$. In the R sector, the massless Dirac operator becomes

$$k_0\Gamma^0 + k_1\Gamma^1 = -k_1\Gamma^0(\Gamma^0\Gamma^1 + 1) = 2k_1\Gamma^0(S_0 - \tfrac{1}{2}) \,. \tag{10.5.15}$$

The physical state condition is then

$$(S_0 - \tfrac{1}{2})|\mathbf{s}, 0; k\rangle_{\rm R} u_{\mathbf{s}} = 0 \,, \tag{10.5.16}$$

so precisely the states with $s_0 = +\frac{1}{2}$ survive. As discussed in section B.1, we have under $SO(9,1) \to SO(1,1) \times SO(8)$ the decompositions

$$\mathbf{16} \to (+\tfrac{1}{2}, \mathbf{8}) + (-\tfrac{1}{2}, \mathbf{8}') \,, \tag{10.5.17a}$$

$$\mathbf{16}' \to (+\tfrac{1}{2}, \mathbf{8}') + (-\tfrac{1}{2}, \mathbf{8}) \,. \tag{10.5.17b}$$

Thus the Dirac equation leaves an $\mathbf{8}$ with $\exp(\pi i F) = +1$ and an $\mathbf{8}'$ with $\exp(\pi i F) = -1$.

The tachyonic and massless states are summarized in table 10.2. The open string spectrum has four sectors, according to the periodicity ν and the world-sheet fermion number $\exp(\pi i F)$. We will use the notation NS$\pm$ and R$\pm$ to label these sectors. We will see in the next section that consistency requires us to keep only certain subsets of sectors, and that there are consistent string theories without the tachyon.

Table 10.3. *Products of $SO(8)$ representations appearing at the massless level of the closed string. The R–NS sector has the same content as the NS–R sector.*

sector	$SO(8)$ spin		tensors		dimensions
(NS+,NS+)	$\mathbf{8}_v \times \mathbf{8}_v$	=	$[0] + [2] + (2)$	=	$\mathbf{1} + \mathbf{28} + \mathbf{35}$
(R+,R+)	$\mathbf{8} \times \mathbf{8}$	=	$[0] + [2] + [4]_+$	=	$\mathbf{1} + \mathbf{28} + \mathbf{35}_+$
(R+,R−)	$\mathbf{8} \times \mathbf{8}'$	=	$[1] + [3]$	=	$\mathbf{8}_v + \mathbf{56}_t$
(R−,R−)	$\mathbf{8}' \times \mathbf{8}'$	=	$[0] + [2] + [4]_-$	=	$\mathbf{1} + \mathbf{28} + \mathbf{35}_-$
(NS+,R+)	$\mathbf{8}_v \times \mathbf{8}$			=	$\mathbf{8}' + \mathbf{56}$
(NS+,R−)	$\mathbf{8}_v \times \mathbf{8}'$			=	$\mathbf{8} + \mathbf{56}'$

Closed string spectrum

The closed string is two copies of the open string, with the momentum rescaled $k \to \frac{1}{2}k$ in the generators. With $\nu, \tilde{\nu}$ taking the values 0 and $\frac{1}{2}$, the mass-shell condition can be summarized as

$$\frac{\alpha'}{4} m^2 = N - \nu = \tilde{N} - \tilde{\nu} \,. \tag{10.5.18}$$

The tachyonic and massless closed string spectrum is obtained by combining one left-moving and one right-moving state, subject to the equality (10.5.18).

The (NS−,NS−) sector contains a closed string tachyon with $m^2 = -2/\alpha'$. At the massless level, combining the various massless left- and right-moving states from table 10.2 leads to the $SO(8)$ representations shown in table 10.3. Note that level matching prevents pairing of the NS− sector with any of the other three. As in the bosonic string, vector times vector decomposes into scalar, antisymmetric tensor, and traceless symmetric tensor denoted (2). The products of spinors are discussed in section B.1.

The 64 states in $\mathbf{8}_v \times \mathbf{8}$ and $\mathbf{8}_v \times \mathbf{8}'$ each separate into two irreducible representations. Denoting a state in $\mathbf{8}_v \times \mathbf{8}$ by $|i, \mathbf{s}\rangle$, we can form the eight linear combinations

$$|i, \mathbf{s}\rangle \Gamma^i_{\mathbf{ss}'} \,. \tag{10.5.19}$$

These states transform among themselves under $SO(8)$, and they are in the $\mathbf{8}'$ representation because the chirality of the loose index $\mathbf{s}'$ is opposite to that of $\mathbf{s}$. The other 56 states form an irreducible representation $\mathbf{56}$. The product $\mathbf{8}_v \times \mathbf{8}'$ works in the same way. Note that there are several cases of distinct representations with identical dimensions: at dimension 8 a vector and two spinors, at dimension 56 an antisymmetric rank 3 tensor and two vector-spinors, at dimension 35 a traceless symmetric rank 2 tensor and self-dual and anti-self-dual rank 4 tensors.

BRST quantization

From the general structure discussed in chapter 4, in particular the expression (4.3.14) for the BRST operator for a general constraint algebra, the BRST operator can be constructed as a simple extension of the bosonic one:

$$Q_{\rm B} = \frac{1}{2\pi i}\oint (dz\, j_{\rm B} - d\bar z\, \tilde j_{\rm B})\,, \tag{10.5.20}$$

where

$$\begin{aligned} j_{\rm B} &= cT_B^{\rm m} + \gamma T_F^{\rm m} + \frac{1}{2}\left(cT_B^{\rm g} + \gamma T_F^{\rm g}\right) \\ &= cT_B^{\rm m} + \gamma T_F^{\rm m} + bc\partial c + \frac{3}{4}(\partial c)\beta\gamma + \frac{1}{4}c(\partial\beta)\gamma - \frac{3}{4}c\beta\partial\gamma - b\gamma^2\,, \end{aligned} \tag{10.5.21}$$

and the same on the antiholomorphic side. As in the bosonic case, this is a tensor up to an unimportant total derivative term.

The BRST current has the essential property

$$j_{\rm B}(z)b(0) \sim \ldots + \frac{1}{z}T_B(0)\,, \quad j_{\rm B}(z)\beta(0) \sim \ldots + \frac{1}{z}T_F(0)\,, \tag{10.5.22}$$

so that the commutators of $Q_{\rm B}$ with the b, β ghosts give the corresponding constraints.[3] In modes,

$$\{Q_{\rm B}, b_n\} = L_n\,, \quad [Q_{\rm B}, \beta_r] = G_r\,. \tag{10.5.23}$$

From these one can verify nilpotence by the same steps as in the bosonic case (exercise 4.3) whenever the total central charge vanishes. Thus, we can replace some of the spacelike $X^\mu\psi^\mu$ SCFTs with any positive-norm SCFT such that the total matter central charge is $c^{\rm m} = \tilde c^{\rm m} = 15$. The BRST current must be periodic for the BRST charge to be well defined. The supercurrent of the SCFT must therefore have the same periodicity, R or NS, as the ψ^μ, β, and γ. The expansion of the BRST operator is

$$\begin{aligned} Q_{\rm B} = &\sum_m c_{-m}L_m^{\rm m} + \sum_r \gamma_{-r}G_r^{\rm m} - \sum_{m,n}\frac{1}{2}(n-m)\,{:}b_{-m-n}c_m c_n{:}\,, \\ &+ \sum_{m,r}\left[\frac{1}{2}(2r-m)\,{:}\beta_{-m-r}c_m\gamma_r{:} - {:}b_{-m}\gamma_{m-r}\gamma_r{:}\right] + a^{\rm g}c_0\,, \end{aligned} \tag{10.5.24}$$

where m and n run over integers and r over (integers + ν). The ghost normal ordering constant is as in eq. (10.4.5).

[3] The $bc\beta\gamma$ theory actually has a one-parameter family of superconformal symmetries, related by rescaling $\beta \to x\beta$ and $\gamma \to x^{-1}\gamma$. The general BRST construction (4.3.14) singles out the symmetry (10.1.21); this is most easily verified by noting that it correctly leads to the OPEs (10.5.22).

The observable spectrum is the space of BRST cohomology classes. As in the bosonic theory, we impose the additional conditions

$$b_0|\psi\rangle = L_0|\psi\rangle = 0\,. \tag{10.5.25}$$

In addition, in the R sector we impose

$$\beta_0|\psi\rangle = G_0|\psi\rangle = 0\,, \tag{10.5.26}$$

the logic being the same as for (10.5.25). The reader can again work out the first few levels by hand, the result being exactly the same as for OCQ. The no-ghost theorem is as in the bosonic case. The BRST cohomology has a positive definite inner product and is isomorphic to OCQ and to the transverse Hilbert space $\mathscr{H}^\perp$, which is defined to have no $\alpha^{0,1}$, $\psi^{0,1}$, b, c, β, or γ excitations. The proof is a direct imitation of the bosonic argument of chapter 4.

We have defined $\exp(\pi i F)$ to commute with $Q_{\rm B}$. We can therefore consider subspaces with definite eigenvalues of $\exp(\pi i F)$ and the no-ghost theorem holds separately in each.

10.6 Superstring theories in ten dimensions

We now focus on the theory in ten flat dimensions. For the four sectors of the open string spectrum we will use in addition to the earlier notation NS±, R± the notation

$$(\alpha, F)\,, \tag{10.6.1}$$

where the combination

$$\alpha = 1 - 2\nu \tag{10.6.2}$$

is 1 in the R sector and 0 in the NS sector. Both α and F are defined only mod 2. The closed string has independent periodicities and fermion numbers on both sides, and so has 16 sectors labeled by

$$(\alpha, F, \tilde\alpha, \tilde F)\,. \tag{10.6.3}$$

Actually, six of these sectors are empty: in the NS− sector the level $L_0 - \alpha' p^2/4$ is half-integer, while in the sectors NS+, R+, and R− it is an integer. It is therefore impossible to satisfy the level-matching condition $L_0 = \tilde L_0$ if NS− is paired with one of the other three.

Not all of these states can be present together in a consistent string theory. Consider first the closed string spectrum. We have seen that the spinor fields have branch cuts in the presence of R sector vertex operators. Various pairs of vertex operators will then have branch cuts in their operator products — they are not *mutually local*. The operator F

counts the number of spinor fields in a vertex operator, so the net phase when one vertex operator circles another is

$$\exp \pi i\left(F_1\alpha_2 - F_2\alpha_1 - \tilde{F}_1\tilde{\alpha}_2 + \tilde{F}_2\tilde{\alpha}_1\right) . \tag{10.6.4}$$

If this phase is not unity, the amplitude with both operators cannot be consistently defined.

A consistent closed string theory will then contain only some subset of the ten sectors. Thus there are potentially 2^{10} combinations of sectors, but only a few of these lead to consistent string theories. We impose three consistency conditions:

(a) From the above discussion, all pairs of vertex operators must be mutually local: if both $(\alpha_1, F_1, \tilde{\alpha}_1, \tilde{F}_1)$ and $(\alpha_2, F_2, \tilde{\alpha}_2, \tilde{F}_2)$ are in the spectrum then

$$F_1\alpha_2 - F_2\alpha_1 - \tilde{F}_1\tilde{\alpha}_2 + \tilde{F}_2\tilde{\alpha}_1 \in 2\mathbf{Z} . \tag{10.6.5}$$

(b) The OPE must close. The parameter α is conserved mod 2 under operator products (for example, R $\times$ R = NS), as is F. Thus if $(\alpha_1, F_1, \tilde{\alpha}_1, \tilde{F}_1)$ and $(\alpha_2, F_2, \tilde{\alpha}_2, \tilde{F}_2)$ are in the spectrum then so is

$$(\alpha_1 + \alpha_2, F_1 + F_2, \tilde{\alpha}_1 + \tilde{\alpha}_2, \tilde{F}_1 + \tilde{F}_2) . \tag{10.6.6}$$

(c) For an arbitrary choice of sectors, the one-loop amplitude will not be modular-invariant. We will study modular invariance in the next section, but in order to reduce the number of possibilities it is useful to extract one simple necessary condition:

There must be at least one left-moving R sector ($\alpha = 1$) and at least one right-moving R sector ($\tilde{\alpha} = 1$).

We now solve these constraints. Assume first that there is at least one R–NS sector, $(\alpha, \tilde{\alpha}) = (1, 0)$. By the level-matching argument, it must either be (R+,NS+) or (R−,NS+). Further, by (a) only one of these can appear, because the product of the corresponding vertex operators is not single-valued. By (c), there must also be at least one NS–R or R–R sector, and because R–NS $\times$ R–R = NS–R, there must in any case be an NS–R sector. Again, this must be either (NS+,R+) or (NS+,R−), but not both. So we have four possibilities, (R+,NS+) or (R−,NS+) with (NS+,R+) or (NS+,R−). Applying closure and single-valuedness leads to precisely two additional sectors in each case, namely (NS+,NS+) and one R–R sector. The spectra which solve (a), (b), and (c) with at least one R–NS sector are

IIB: (NS+,NS+) (R+,NS+) (NS+,R+) (R+,R+) ,
IIA: (NS+,NS+) (R+,NS+) (NS+,R−) (R+,R−) ,
IIA′: (NS+,NS+) (R−,NS+) (NS+,R+) (R−,R+) ,
IIB′: (NS+,NS+) (R−,NS+) (NS+,R−) (R−,R−) .

Notice that none of these theories contains the tachyon, which lives in the sector (NS−,NS−).

These four solutions represent just two physically distinct theories. In the IIA and IIA′ theories the R–R states have the opposite chirality on the left and the right, and in the IIB and IIB′ theories they have the same chirality. A spacetime reflection on a single axis, say

$$X^2 \to -X^2 , \quad \psi^2 \to -\psi^2 , \quad \tilde{\psi}^2 \to -\tilde{\psi}^2 , \tag{10.6.7}$$

leaves the action and the constraints unchanged but reverses the sign of $\exp(\pi i F)$ in the left-moving R sectors and the sign of $\exp(\pi i \tilde{F})$ in the right-moving R sectors. At the massless level this switches the Weyl representations, $\mathbf{16} \leftrightarrow \mathbf{16}'$. It therefore turns the IIA′ theory into IIA, and IIB′ into IIB.

Now suppose that there is no R–NS sector. By (c), there must be at least one R–R sector. In fact the combination of (NS+,NS+) with any single R–R sector solves (a), (b), and (c), but these turn out not to be modular-invariant. Proceeding further, one readily finds the only other solutions,

0A: (NS+,NS+) (NS−,NS−) (R+,R−) (R−,R+) ,
0B: (NS+,NS+) (NS−,NS−) (R+,R+) (R−,R−) .

These are modular-invariant, but both have a tachyon and there are no spacetime fermions.

In conclusion, we have found two potentially interesting string theories, the *type IIA* and *IIB superstring theories*. Referring back to table 10.3, one finds the massless spectra

$$\text{IIA:} \quad [0] + [1] + [2] + [3] + (2) + \mathbf{8} + \mathbf{8}' + \mathbf{56} + \mathbf{56}' , \tag{10.6.8a}$$

$$\text{IIB:} \quad [0]^2 + [2]^2 + [4]_+ + (2) + \mathbf{8}^2 + \mathbf{56}^2 . \tag{10.6.8b}$$

The IIB theory is defined by keeping all sectors with

$$\exp(\pi i F) = \exp(\pi i \tilde{F}) = +1 , \tag{10.6.9}$$

and the IIA theory by keeping all sectors with

$$\exp(\pi i F) = +1 , \quad \exp(\pi i \tilde{F}) = (-1)^{\tilde{\alpha}} . \tag{10.6.10}$$

This projection of the full spectrum down to eigenspaces of $\exp(\pi i F)$ and $\exp(\pi i \tilde{F})$ is known as the *Gliozzi–Scherk–Olive (GSO) projection*. In the IIA theory the opposite GSO projections are taken in the NS–R and R–NS sectors, so the spectrum is nonchiral. That is, the spectrum is invariant under spacetime parity, which interchanges $\mathbf{8} \leftrightarrow \mathbf{8}'$ and $\mathbf{56} \leftrightarrow \mathbf{56}'$. On the world-sheet, this symmetry is the product of spacetime parity and world-sheet parity. In the IIB theory the same GSO projection is taken in each sector and the spectrum is chiral.

The type 0 theories are formed by a different method: for example, 0B is defined by keeping all sectors with

$$\alpha = \tilde{\alpha} , \quad \exp(\pi i F) = \exp(\pi i \tilde{F}) . \tag{10.6.11}$$

The projections that define the type II theories act separately on the left- and right-moving spinors, while the projections that define the type 0 theory tie the two together. The latter are sometimes called *diagonal GSO projections.*

The most striking features of the type II theories are the massless vector–spinor *gravitinos* in the NS–R and R–NS sectors. The terminology type II refers to the fact that these theories each have two gravitinos. In the IIA theory the gravitinos have opposite chiralities (Γ eigenvalues), and in the IIB theory they have the same chirality. The NS–R gravitino state is

$$\psi^{\mu}_{-1/2}|0;\mathbf{s};k\rangle_{\text{NS-R}} u_{\mu\mathbf{s}} . \tag{10.6.12}$$

The physical state conditions are

$$k^2 = k^{\mu} u_{\mu\mathbf{s}} = k \cdot \Gamma_{\mathbf{s}\mathbf{s}'} u_{\mu\mathbf{s}'} = 0 , \tag{10.6.13}$$

as well as the equivalence relation

$$u_{\mu\mathbf{s}} \cong u_{\mu\mathbf{s}} + k_{\mu} \zeta_{\mathbf{s}} . \tag{10.6.14}$$

We have learned that such equivalence relations are the signature of a local spacetime symmetry. Here the symmetry parameter $\zeta_{\mathbf{s}}$ is a spacetime spinor so we have local *spacetime supersymmetry.* In flat spacetime there will be a conserved spacetime supercharge $Q^A_{\mathbf{s}}$, where A distinguishes the symmetries associated with the two gravitinos, and $\mathbf{s}$ is a spinor index of the same chirality as the corresponding gravitino. Thus the IIA theory has one supercharge transforming as the **16** of $SO(9,1)$ and one transforming as the **16′**, and the IIB theory has two transforming as the **16**.

The gravitino vertex operators are

$$\mathscr{V}_{\mathbf{s}} e^{-\tilde{\phi}} \tilde{\psi}^{\mu} e^{ik\cdot X} , \quad e^{-\phi} \psi^{\mu} \tilde{\mathscr{V}}_{\mathbf{s}} e^{ik\cdot X} . \tag{10.6.15}$$

The operators $\mathscr{V}_{\mathbf{s}}$ and $\tilde{\mathscr{V}}_{\mathbf{s}}$, defined in eq. (10.4.25), have weights $(1,0)$ and $(0,1)$ and so are world-sheet currents associated with the spacetime supersymmetries.

This is our first encounter with spacetime supersymmetry, and the reader should now study the appropriate sections of appendix B. Section B.2 gives an introduction to spacetime supersymmetry. Section B.4 discusses antisymmetric tensor fields, which we have in the massless IIA and IIB spectra. Section B.5 briefly discusses the IIA and IIB supergravity theories which describe the low energy physics of the IIA and IIB superstrings. In each of the type II theories, there is a unique massless representation, which has $2^8 = 256$ states. The massless superstring spectra are the

massless representations of IIA and IIB $d = 10$ spacetime supersymmetry respectively. This is to be expected: if all requirements for a consistent string theory are met (and they are) then the existence of the gravitinos implies that the corresponding supersymmetries must be present.

The reader may feel that the construction in this section, which is the *Ramond–Neveu–Schwarz (RNS)* form of the superstring, is somewhat *ad hoc.* In particular one might expect that the spacetime supersymmetry should be manifest from the start. There is certainly truth to this, but the existing supersymmetric formulation (the *Green–Schwarz superstring*) seems to be even more unwieldy.

Note that the world-sheet and spacetime supersymmetries are distinct, and that the connection between them is indirect. The world-sheet supersymmetry parameter $\eta(z)$ is a spacetime scalar and world-sheet spinor, while the spacetime supersymmetry parameter $\zeta_{\mathbf{s}}$ is a spacetime spinor and world-sheet scalar. The world-sheet supersymmetry is a constraint in the world-sheet theory, annihilating physical states. The spacetime supersymmetry is a global symmetry of the world-sheet theory, giving relations between masses and amplitudes, though it becomes a local symmetry in spacetime.

Let us note one more feature of the GSO projection. In bosonized form, all the R sector vertex operators have odd length-squared and all the NS sector vertex operators have even length-squared, in terms of the $\circ$ product defined in section 10.4. This can be seen at the lowest levels for the operators (10.4.22) and (10.4.25), the tachyon having been removed by the GSO projection. By the remark at the end of section 10.4, the spacetime spin is then correlated with the *world-sheet* statistics. In fact, this is the same as the *space-time* statistics. The world-sheet statistics governs the behavior of the world-sheet amplitude under simultaneous exchange of world-sheet position, spacetime momentum, and other quantum numbers. After integrating over position, this determines the symmetry of the spacetime S-matrix. The result is the expected spacetime spin-statistics connection. Note that operators with the wrong spin-statistics connection, such as ψ^{μ} and $e^{-\phi}$, appear at intermediate stages but the projections that produce a consistent theory also give the spin-statistics connection. This is certainly a rather technical way for the spin-statistics theorem to arise, but it is worth noting that all string theories seem to obey the usual spin-statistics relation.

Unoriented and open superstrings

The IIB superstring, with the same chiralities on both sides, has a world-sheet parity symmetry Ω. We can gauge this symmetry to obtain an

unoriented closed string theory.[4] In the NS–NS sector, this eliminates the [2], leaving [0] + (2), just as it does in the unoriented bosonic theory. The fermionic NS–R and R–NS sectors of the IIB theory have the same spectra, so the Ω projection picks out the linear combination (NS–R) + (R–NS), with massless states $\mathbf{8}' + \mathbf{56}$. In particular, one gravitino survives the projection. Finally, the existence of the gravitino means that there must be equal numbers of massless bosons and fermions, so a consistent definition of the world-sheet parity operator must select the [2] from the R–R sector to give 64 of each. One can understand this as follows. The R–R vertex operators

$$\mathscr{V}_{\mathbf{s}}\tilde{\mathscr{V}}_{\mathbf{s}'} \tag{10.6.16}$$

transform as $\mathbf{8} \times \mathbf{8} = [0] + [2] + [4]_+$. The [0] and $[4]_+$ are symmetric under interchange of $\mathbf{s}$ and $\mathbf{s}'$ and the [2] antisymmetric (one can see this by counting states, 36 versus 28, or in more detail by considering the S_a eigenvalues of the representations). World-sheet parity adds or subtracts a tilde to give

$$\tilde{\mathscr{V}}_{\mathbf{s}}\mathscr{V}_{\mathbf{s}'} = -\mathscr{V}_{\mathbf{s}'}\tilde{\mathscr{V}}_{\mathbf{s}}\,, \tag{10.6.17}$$

where the final sign comes from the fermionic nature of the R vertex operators. Thus, projecting onto $\Omega = +1$ picks out the antisymmetric [2]. The result is the *type I closed unoriented theory*, with spectrum

$$[0] + [2] + (2) + \mathbf{8}' + \mathbf{56} \;=\; \mathbf{1} + \mathbf{28} + \mathbf{35} + \mathbf{8}' + \mathbf{56}\,. \tag{10.6.18}$$

However, this theory by itself is inconsistent, as we will explain further below.

Now consider open string theories. Closure of the OPE in open + open $\to$ closed scattering implies that any open string that couples consistently to type I or type II closed superstrings must have a GSO projection in the open string sector. The two possibilities and their massless spectra are

$$\begin{aligned} &\text{I:} \quad \text{NS+, R+} = \mathbf{8}_v + \mathbf{8}\,, \\ &\tilde{\text{I}}\text{:} \quad \text{NS+, R}- = \mathbf{8}_v + \mathbf{8}'\,. \end{aligned}$$

Adding Chan–Paton factors, the gauge group will again be $U(n)$ in the oriented case and $SO(n)$ or $Sp(k)$ in the unoriented case. The $\mathbf{8}$ or $\mathbf{8}'$ spinors are known as *gauginos* because they are related to the gauge bosons by supersymmetry. They must be in the adjoint representation of the gauge group, like the gauge bosons, because supersymmetry commutes with the gauge symmetry.

[4] The analogous operation in the IIA theory would be to gauge the symmetry which is the product of world-sheet and spacetime parity, but this breaks some of the Poincaré invariance. We will encounter this in section 13.2.

We can already anticipate that not all of these theories will be consistent. The open string multiplets, with 16 states, are representations of $d = 10$, $N = 1$ supersymmetry but not of $N = 2$ supersymmetry. Thus the open superstring cannot couple to the oriented closed superstring theories, which have two gravitinos.[5] It can only couple to the unoriented closed string theory (10.6.18) and so the open string theory must also be unoriented for consistent interactions. With the chirality (10.6.18), the massless open string states must be $\mathbf{8}_v + \mathbf{8}$. This is required by spacetime supersymmetry, or by conservation of $\exp(\pi i F)$ on the world-sheet. The result is the unoriented *type I open plus closed superstring theory*, with massless content

$$[0] + [2] + (2) + \mathbf{8}' + \mathbf{56} + (\mathbf{8}_v + \mathbf{8})_{SO(n) \text{ or } Sp(k)}\,. \tag{10.6.19}$$

There is a further inconsistency in all but the $SO(32)$ theory. We will see in section 10.8 that for all other groups, as well as the purely closed unoriented theory, there is a one-loop divergence and superconformal anomaly. We will also see, in chapter 12, that the spacetime gauge and coordinate symmetries have an anomaly at one loop for all but the $SO(32)$ theory.

Thus we have found precisely three tachyon-free and nonanomalous string theories in this chapter: type IIA, type IIB, and type I $SO(32)$.

10.7 Modular invariance

Superstring interactions are the subject of the next chapter, but there is one important amplitude that involves no interactions, only the string spectrum. This is the one-loop vacuum amplitude, studied for the bosonic string in chapter 7. We study the vacuum amplitude for the closed superstring in this section and for the open string in the next.

We make the guess, correctly it will turn out, that the torus amplitude is again given by the Coleman–Weinberg formula (7.3.24) with the region of integration replaced by the fundamental region for the moduli space of the torus:

$$Z_{T_2} = V_{10}\int_F \frac{d^2\tau}{4\tau_2}\int \frac{d^{10}k}{(2\pi)^{10}}\sum_{i\in\mathscr{H}^\perp}(-1)^{\mathbf{F}_i}q^{\alpha'(k^2+m_i^2)/4}\bar{q}^{\alpha'(k^2+\bar{m}_i^2)/4}\,, \tag{10.7.1}$$

with $q = \exp(2\pi i\tau)$. We have included the minus sign for spacetime

[5] At the world-sheet level the problem is that the total derivative null gravitino vertex operators give rise to nonzero world-sheet boundary terms. Only one linear combination of the two null gravitinos decouples, so we must make the world-sheet parity projection in order to eliminate the other.

fermions from the Coleman–Weinberg formula, distinguishing the spacetime fermion number $\mathbf{F}$ from the world-sheet fermion number F. The masses are given in terms of the left- and right-moving parts of the transverse Hamiltonian by

$$m^2 = 4H^\perp/\alpha' \ , \quad \tilde{m}^2 = 4\tilde{H}^\perp/\alpha' \ . \tag{10.7.2}$$

The trace includes a sum over the different $(\alpha, F; \tilde{\alpha}, \tilde{F})$ sectors of the superstring Hilbert space. In each sector it breaks up into a product of independent sums over the transverse X, ψ, and $\tilde{\psi}$ oscillators, and the transverse Hamiltonian similarly breaks up into a sum. Each transverse X contributes as in the bosonic string, the total contribution of the oscillator sum and momentum integration being as in eq. (7.2.9),

$$\begin{aligned} Z_X(\tau) &= (4\pi^2\alpha'\tau_2)^{-1/2}(q\bar{q})^{-1/24}\prod_{n=1}^{\infty}\Bigg(\sum_{N_n,\tilde{N}_n=1}^{\infty} q^{nN_n}\bar{q}^{n\tilde{N}_n}\Bigg) \\ &= (4\pi^2\alpha'\tau_2)^{-1/2}|\eta(q)|^{-2} \ , \end{aligned} \tag{10.7.3}$$

where $\eta(\tau) = q^{1/24}\prod_{n=1}^{\infty}(1-q^n)$. In addition there is a factor $i(4\pi^2\alpha'\tau_2)^{-1}$ from the $k^{0,1}$ integrations.

For the ψs, the mode sum in each sector depends on the spatial periodicity α and includes a projection operator $\frac{1}{2}[1 \pm \exp(\pi i F)]$. Although for the present we are interested only in R and NS periodicities, let us work out the partition functions for the more general periodicity (10.3.20),

$$\psi(w + 2\pi) = \exp[\pi i(1-\alpha)]\,\psi(w) \tag{10.7.4}$$

where again $\alpha = 1 - 2\nu$. By the definition (10.3.23) of the ground state, the raising operators are

$$\psi_{-m+(1-\alpha)/2} \ , \quad \bar{\psi}_{-m+(1+\alpha)/2} \ , \quad m = 1, 2, \ldots \ . \tag{10.7.5}$$

The ground state weight was found to be $\alpha^2/8$. Then

$$\mathrm{Tr}_\alpha\left(q^H\right) = q^{(3\alpha^2-1)/24}\prod_{m=1}^{\infty}\left[1+q^{m-(1-\alpha)/2}\right]\left[1+q^{m-(1+\alpha)/2}\right] \ . \tag{10.7.6}$$

To define the general boundary conditions we have joined the fermions into complex pairs. Thus we can define a fermion number Q which is $+1$ for ψ and -1 for $\bar{\psi}$. To be precise, define Q to be the H-momentum in the bosonization (10.3.10) so that it is conserved by the OPE. The bosonization (10.3.25) then gives the charge of the ground state as $\alpha/2$.

Thus we can define the more general trace

$$Z^\alpha{}_\beta(\tau) = \mathrm{Tr}_\alpha\left[q^H \exp(\pi i\beta Q)\right] \tag{10.7.7a}$$

$$= q^{(3\alpha^2-1)/24}\exp(\pi i\alpha\beta/2) \times \prod_{m=1}^{\infty}\left[1+\exp(\pi i\beta)q^{m-(1-\alpha)/2}\right]\left[1+\exp(-\pi i\beta)q^{m-(1+\alpha)/2}\right] \tag{10.7.7b}$$

$$= \frac{1}{\eta(\tau)}\vartheta\begin{bmatrix}\alpha/2\\ \beta/2\end{bmatrix}(0,\tau) \ . \tag{10.7.7c}$$

The notation in the final line was introduced in section 7.2, but our discussion of these functions in the present volume will be self-contained.

The charge Q modulo 2 is the fermion number F that appears in the GSO projection. Thus the traces that are relevant for the ten-dimensional superstring are

$$Z^0{}_0(\tau) = \mathrm{Tr}_{\mathrm{NS}}\left[q^H\right] \ , \tag{10.7.8a}$$

$$Z^0{}_1(\tau) = \mathrm{Tr}_{\mathrm{NS}}\left[\exp(\pi i F)\,q^H\right] \ , \tag{10.7.8b}$$

$$Z^1{}_0(\tau) = \mathrm{Tr}_{\mathrm{R}}\left[q^H\right] \ , \tag{10.7.8c}$$

$$Z^1{}_1(\tau) = \mathrm{Tr}_{\mathrm{R}}\left[\exp(\pi i F)\,q^H\right] \ . \tag{10.7.8d}$$

We should emphasize that these traces are for a *pair* of dimensions.

Tracing over all eight fermions, the GSO projection keeps states with $\exp(\pi i F) = +1$. This is $Z^+_\psi(\tau)$, where

$$Z^\pm_\psi(\tau) = \frac{1}{2}\left[Z^0{}_0(\tau)^4 - Z^0{}_1(\tau)^4 - Z^1{}_0(\tau)^4 \mp Z^1{}_1(\tau)^4\right] \ . \tag{10.7.9}$$

The half is from the projection operator, the minus sign in the second term is from the ghost contribution to $\exp(\pi i F)$, and the minus signs in the third and fourth (R sector) terms are from spacetime spin-statistics. For $\tilde{\psi}$ in the IIB theory one obtains the conjugate $Z^+_\psi(\tau)^*$. In the IIA theory, $\tilde{F} = -1$ in the R sector so the result is $Z^-_\psi(\tau)^*$. In all,

$$Z_{T_2} = iV_{10}\int_F \frac{d^2\tau}{16\pi^2\alpha'\tau_2^2}\,Z_X^8 Z^+_\psi(\tau)Z^\pm_\psi(\tau)^* \ . \tag{10.7.10}$$

We know from the discussion of bosonic amplitudes that modular invariance is necessary for the consistency of string theory. In the superstring this works out in an interesting way. The combination $d^2\tau/\tau_2^2$ is modular-invariant, as is Z_X. To understand the modular transformations of the fermionic traces, note that $Z^\alpha{}_\beta$ is given by a path integral on the torus

over fermionic fields ψ with periodicities

$$\psi(w+2\pi) = -\exp(\pi i\alpha)\,\psi(w)\,, \tag{10.7.11a}$$

$$\psi(w+2\pi\tau) = -\exp(\pi i\beta)\,\psi(w)\,. \tag{10.7.11b}$$

This gives

$$\psi[w+2\pi(\tau+1)] = \exp[\pi i(\alpha+\beta)]\,\psi(w)\,. \tag{10.7.12}$$

Naively then, $Z^{\alpha}{}_{\beta}(\tau) = Z^{\alpha}{}_{\alpha+\beta-1}(\tau+1)$, since both sides are given by the same path integral. Also, defining $w' = w/\tau$ and $\psi'(w') = \psi(w)$,

$$\psi'(w'+2\pi) = -\exp(\pi i\beta)\,\psi'(w') \tag{10.7.13a}$$

$$\psi'(w'-2\pi/\tau) = -\exp(-\pi i\alpha)\,\psi'(w')\,, \tag{10.7.13b}$$

so that naively $Z^{\alpha}{}_{\beta}(\tau) = Z^{\beta}{}_{-\alpha}(-1/\tau)$. It is easy to see that by these two transformations one can always reach a path integral with $\alpha = 1$, accounting for rule (c) from the previous section.

The reason these modular transformations are naive is that there is no diff-invariant way to define the phase of the path integral for purely left-moving fermions. For left- plus right-moving fermions with matching boundary conditions, the path integral can be defined by Pauli–Villars or other regulators. This is the same as the absolute square of the left-moving path integral, but leaves a potential phase ambiguity in that path integral separately.[6] The naive result is correct for $\tau \to -1/\tau$, but under $\tau \to \tau+1$ there is an additional phase,

$$\begin{aligned} Z^{\alpha}{}_{\beta}(\tau) &= Z^{\beta}{}_{-\alpha}(-1/\tau) \\ &= \exp[-\pi i(3\alpha^2-1)/12]\, Z^{\alpha}{}_{\alpha+\beta-1}(\tau+1)\,. \end{aligned} \tag{10.7.14}$$

The $\tau \to \tau+1$ transformation follows from the explicit form (10.7.7b), the phase coming from the zero-point energy with the given boundary conditions. The absence of a phase in $\tau \to -1/\tau$ can be seen at once for $\tau = i$. Note that $Z^{1}{}_{1}$ actually vanishes due to cancellation between the two R sector ground states, but we have assigned a formal transformation law for a reason to be explained below.

The phase represents a *global gravitational anomaly*, an inability to define the phase of the path integral such that it is invariant under large coordinate transformations. Of course, a single left-moving fermion has $c \neq \tilde{c}$ and so has an anomaly even under infinitesimal coordinate transformations, but the global anomaly remains even when a left- and right-moving fermion are combined. For example, the product $Z^{1}{}_{0}(\tau)^{*}Z^{0}{}_{0}(\tau)$ has no infinitesimal anomaly and should come back to itself under $\tau \to \tau+2$,

[6] The phase factor is a holomorphic function of τ, because the $Z^{\alpha}{}_{\beta}$ are. Since it has magnitude 1, this implies that it is actually independent of τ.

but in fact picks up a phase $\exp(-\pi i/2)$. This phase arises from the *level mismatch*, the difference of zero-point energies in the NS and R sectors.

The reader can verify that with the transformations (10.7.14), the combinations $Z^{\pm}_{\psi}$ are invariant under $\tau \to -1/\tau$ and are multiplied by $\exp(2\pi i/3)$ under $\tau \to \tau+1$. Combined with the conjugates from the right-movers, the result is modular-invariant and the torus amplitude consistent. It is necessary for the construction of this invariant that there be a multiple of *eight* transverse fermions. Recall from section 7.2 that invariance under $\tau \to \tau+1$ requires that $L_0 - \tilde{L}_0$ be an integer for all states. For a single real fermion in the R–NS sector the difference in ground state energies is $\frac{1}{16}$. For eight fermions this becomes $\frac{1}{2}$, so that states with an odd number of NS excitations (as required by the GSO projection) are level-matched. Note also that modular invariance forces the minus signs in the combination (10.7.9), in particular the relative sign of $(Z^{0}{}_{0})^4$ and $(Z^{1}{}_{0})^4$ which corresponds to Fermi statistics for the R sector states.

In the type 0 superstrings the fermionic trace is

$$\frac{1}{2}\left[|Z^{0}{}_{0}(\tau)|^N + |Z^{0}{}_{1}(\tau)|^N + |Z^{1}{}_{0}(\tau)|^N \mp |Z^{1}{}_{1}(\tau)|^N\right] \tag{10.7.15}$$

with $N = 8$. This is known as the *diagonal modular invariant*, and it is invariant for any N because the phases cancel in the absolute values.

The type II theories have spacetime supersymmetry. This implies equal numbers of bosons and fermions at each mass level, and so Z_{T_2} should *vanish* in these theories by cancellation between bosons and fermions. Indeed it does, as a consequence of $Z^{1}{}_{1} = 0$ and the 'abstruse identity' of Jacobi,

$$Z^{0}{}_{0}(\tau)^4 - Z^{0}{}_{1}(\tau)^4 - Z^{1}{}_{0}(\tau)^4 = 0\,. \tag{10.7.16}$$

The same cancellation occurs in the open and unoriented theories.

Although we have focused on the path integral without vertex operators, amplitudes with vertex operators must also be modular-invariant. In the present case the essential issue is the path integral measure, and one can show by explicit calculation (or by indirect arguments) that the modular properties are the same with or without vertex operators. However, with a general vertex operator insertion the $\alpha = \beta = 1$ path integral will no longer vanish, nor will the sum of the other three. The general amplitude will then be modular-invariant provided that the vacuum is modular-invariant *without* using the vanishing of $Z^{1}{}_{1}$ or the abstruse identity (10.7.16) — as we have required.

More on $c = 1$ CFT

The equality of the bosonic and fermionic partition functions (10.3.17) and (10.3.18) was one consequence of bosonization. These partition func-

tions are not modular-invariant and so do not define a sensible string background. The fermionic spectrum consists of all NS–NS states. The bosonic spectrum consists of all states with integer k_R and k_L; this is not the spectrum of toroidal compactification at any radius. The simplest modular-invariant fermionic partition function is the diagonal invariant, taking common periodicities for the left- and right-movers. In terms of the states, this amounts to projecting

$$\alpha = \tilde{\alpha}\,, \quad \exp(\pi i F) = \exp(\pi i \tilde{F})\,. \tag{10.7.17}$$

The NS–NS sector consists of the local operators we have been considering, and the chirality projection $\exp(\pi i F) = \exp(\pi i \tilde{F})$ means that on the bosonic side $k_R = k_L$ mod 2. The bosonic equivalents for the R–R sector states have half-integral k_R and k_L with again $k_R = k_L$ mod 2. In all,

$$(k_R, k_L) = (n_1, n_2) \text{ or } (n_1 + \tfrac{1}{2}, n_2 + \tfrac{1}{2}) \tag{10.7.18}$$

for integers n_1 and n_2 such that $n_1 - n_2 \in 2\mathbf{Z}$. This is the spectrum of a boson on a circle of radius 2, or 1 by T-duality, which we see is equivalent to a complex fermion with the diagonal modular-invariant projection. (The dimensionless radius r for the H scalar corresponds to the radius $R = r(\alpha'/2)^{1/2}$ for X^μ, so $r = 2^{1/2}$ is self-dual.)

To obtain an equivalent fermionic theory at arbitrary radius, add

$$\partial H \bar{\partial} H \cong -\bar{\psi}\psi\tilde{\bar{\psi}}\tilde{\psi} \tag{10.7.19}$$

to the world-sheet Lagrangian density. The H theory is still free, but the equivalent fermionic theory is now an interacting field theory known as the *Thirring model*. The Thirring model has a nontrivial perturbation series but is solvable precisely because of its equivalence to a free boson. Actually, for any *rational* r, the bosonic theory is also equivalent to a *free* fermion theory with a more complicated twist (exercise 10.15).

Another interesting CFT consists of the set of vertex operators with

$$k_R = m/3^{1/2}\,, \quad k_L = n/3^{1/2}\,, \quad m - n \in 3\mathbf{Z}\,. \tag{10.7.20}$$

(This discussion should actually be read after section 11.1.) It is easy to check that this has the same properties as the set of vertex operators with integer $k_{R,L}$. That is, it is a single-valued operator algebra, but does not correspond to the spectrum of the string for any value of r, and does not have a modular-invariant partition function. Its special property is the existence of the operators

$$\exp\left[\pm i 3^{1/2} H(z)\right]\,, \quad \exp\left[\pm i 3^{1/2} \tilde{H}(\bar{z})\right]\,. \tag{10.7.21}$$

These have weights $(\frac{3}{2}, 0)$ and $(0, \frac{3}{2})$: they are world-sheet supercurrents!

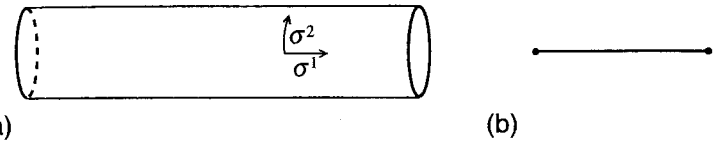


Fig. 10.1. (a) Cylinder in the limit of small t. (b) Analogous field theory graph.

This CFT has (2,2) world-sheet supersymmetry. The standard representation, in which the supercurrent is quadratic in free fields, has two free X and two free ψ fields for central charge 3. This is rather more economical, with one free scalar and central charge 1. The reader can readily check that with appropriate normalization the supercurrents generate the $N = 2$ OPE (11.1.4).

This theory becomes modular-invariant if one twists by the symmetry

$$(H, \tilde{H}) \to (H, \tilde{H}) + \frac{1}{2 \times 3^{1/2}}(1, -1)\,. \tag{10.7.22}$$

This projects the spectrum onto states with $m - n \in 6\mathbf{Z}$ and adds in a twisted sector with $m, n \in \mathbf{Z} + \frac{1}{2}$. The resulting spectrum is the string theory at $r = 2 \times 3^{1/2}$. This twist is a diagonal GSO projection, in that the supercurrent is odd under the symmetry.

10.8 Divergences of type I theory

The cylinder, Möbius strip, and Klein bottle have no direct analog of the modular group, but the condition that the tadpole divergences cancel among these three graphs plays a similar role in restricting the possible consistent theories. The cancellation is very similar to what we have already seen in the bosonic theory in chapter 7. The main new issue is the inclusion of the various sectors in the fermionic path integral, and in particular the separate contributions of closed string NS–NS and R–R tadpoles.

The cylinder

Consider first the cylinder, shown in figure 10.1(a). One can immediately write down the amplitude by combining the bosonic result (7.4.1), converted to ten dimensions, with the fermionic trace (10.7.9) from one side of the type II string. We write it as a sum of two terms,

$$Z_{C_2} = Z_{C_2,0} + Z_{C_2,1}\,, \tag{10.8.1}$$

where

$$Z_{C_2,0} = iV_{10}n^2 \int_0^\infty \frac{dt}{8t}(8\pi^2\alpha' t)^{-5}\eta(it)^{-8}\left[Z^0{}_0(it)^4 - Z^1{}_0(it)^4\right], \tag{10.8.2a}$$

$$Z_{C_2,1} = iV_{10}n^2 \int_0^\infty \frac{dt}{8t}(8\pi^2\alpha' t)^{-5}\eta(it)^{-8}\left[-Z^0{}_1(it)^4 - Z^1{}_1(it)^4\right]. \tag{10.8.2b}$$

Note that the GSO and Ω projection operators each contribute a factor of $\frac{1}{2}$. We have separated the terms according to whether $\exp(\pi i F)$ appears in the trace. In $Z_{C_2,0}$ it does not, and so the ψ^μ are antiperiodic in the σ^2 direction. In $Z_{C_2,1}$ it does appear and the ψ^μ are periodic. We can also regard the cylinder as a closed string appearing from and returning to the vacuum as in figure 10.1(b); we have used this idea in chapters 7 and 8. The periodicities of the ψ^μ mean that in terms of the closed string exchange, the part $Z_{C_2,0}$ comes from NS–NS strings and the part $Z_{C_2,1}$ from R–R strings.

We know from the previous section that the total fermionic partition function vanishes by supersymmetry, so that $Z_{C_2,1} = -Z_{C_2,0}$; we concentrate then on $Z_{C_2,0}$. Using the modular transformations

$$\eta(it) = t^{-1/2}\eta(i/t), \quad Z^\alpha{}_\beta(it) = Z^\beta{}_{-\alpha}(i/t) \tag{10.8.3}$$

and defining $s = \pi/t$, this becomes

$$\begin{aligned} Z_{C_2,0} &= i\frac{V_{10}n^2}{8\pi(8\pi^2\alpha')^5}\int_0^\infty ds\,\eta(is/\pi)^{-8}\left[Z^0{}_0(is/\pi)^4 - Z^0{}_1(is/\pi)^4\right] \\ &= i\frac{V_{10}n^2}{8\pi(8\pi^2\alpha')^5}\int_0^\infty ds\,[16 + O(\exp(-2s))]\,. \end{aligned} \tag{10.8.4}$$

The divergence as $s \to \infty$ is due to a massless closed string tadpole, which as noted must be an NS–NS state. Thus we identify this as a dilaton plus graviton interaction $(-G)^{1/2}e^{-\Phi}$ coming from the disk, as in the bosonic string.

However, there is a paradox here: the $d = 10$, $N = 1$ supersymmetry algebra does not allow such a term. Even more puzzling, $Z_{C_2,1}$ has an equal and opposite divergence which must be from a tadpole of an R–R state, but the only massless R–R state is the rank 2 tensor which cannot have a Lorentz-invariant tadpole.

One can guess the resolution of this as follows. The type IIB string has rank n potentials for all even n, with n and $8 - n$ equivalent by Poincaré duality. The Ω projection removes $n = 0$ and its equivalent $n = 8$, as well as $n = 4$: all the multiples of four. This leaves $n = 2$, its equivalent $n = 6$ — and $n = 10$. A 10-form potential C_{10} can exist in ten dimensions but its 11-form field strength dC_{10} is identically zero. The integral of the

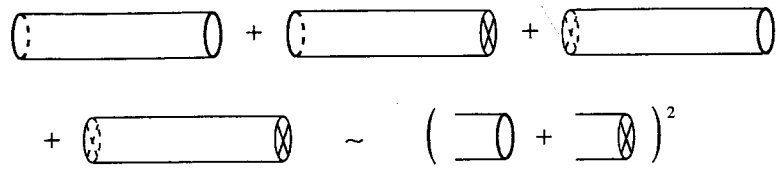


Fig. 10.2. Schematic illustration of cancellation of tadpoles.

potential over spacetime

$$\mu_{10}\int C_{10} \tag{10.8.5}$$

is invariant under $\delta C_{10} = d\chi_9$ and so can appear in the action. Since there is no kinetic term the propagator for this field is $1/0$, and the effect of the tadpole is a divergence

$$\frac{\mu_{10}^2}{0}\,. \tag{10.8.6}$$

This must be the origin of the divergence in $Z_{C_2,1}$, as indeed a more detailed analysis does show. The equation of motion from varying C_{10} is just $\mu_{10} = 0$, so unlike the divergences encountered previously this one cannot be removed by a correction to the background fields. It represents an actual inconsistency.

The Klein bottle

We know from the study of the bosonic string divergences that there is still the possibility of canceling this tadpole as shown in figure 10.2. The cylinder, Möbius strip, and Klein bottle each have divergences from the massless closed string states, the total being proportional to square of the sum of the disk and RP_2 tadpoles. The relative size of the two tadpoles depends on the Chan–Paton factors, and cancels for a particular gauge group.[7]

The relation of the Möbius strip and Klein bottle as depicted in figure 10.2 to the twisted-strip and twisted-cylinder pictures was developed in section 7.4, and is shown in figure 10.3. In order to sum as in figure 10.2, one must rescale the surfaces so that the circumference in the σ^2 direction

[7] In the vacuum amplitude the sum of the NS–NS and R–R divergences is zero for each topology separately because the trace vanishes by supersymmetry. This is not sufficient, because they will no longer cancel when vertex operators are added near one end of each surface. The NS–NS and R–R tadpoles must vanish separately when summed over topologies.

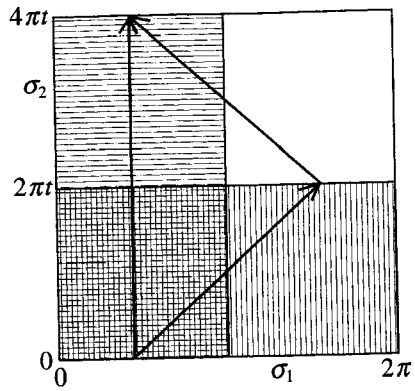


Fig. 10.3. Two fundamental regions for the Klein bottle. The right- and left-hand edges are periodically identified, as are the upper and lower edges. In addition the diagonal arrow shows an orientation-reversing identification. The vertically hatched region is a fundamental region for the twisted-cylinder picture, as is the horizontally hatched region for the decription with two crosscaps. As shown by the arrows, the periodicity of fields in the σ_2-direction of the latter description can be obtained by applying the orientation-reversing periodicity twice. The same picture applies to the Möbius strip, with the right- and left-hand edges boundaries, and with the range of σ_1 changed to π.

and length in the σ^1 directions are uniform; we have taken these to be 2π and s respectively. From figures 10.1 and 10.3 it follows that s is related to the usual modulus t for these surfaces by $s = \pi/t$, $4\pi/t$, and $2\pi/t$ for the cylinder, Möbius strip, and Klein bottle respectively.

Each amplitude is obtained as a sum of traces, from summing over the various periodicity conditions and from expanding out the projection operators. We need to determine which terms contribute to the NS–NS exchange and which to the R–R exchange by examining the boundary conditions on the fermions in the world-sheet path integral. On the Klein bottle the GSO projection operator is

$$\frac{1+\exp(\pi i F)}{2}\cdot\frac{1+\exp(\pi i\tilde{F})}{2}\,. \tag{10.8.7}$$

With $R = \Omega\exp(\pi i\beta F + \pi i\tilde{\beta}\tilde{F})$ in the trace, the path integral boundary conditions are

$$\psi(w+2\pi i t) = -R\psi(w)R^{-1} = -\exp(\pi i\beta)\,\tilde{\psi}(\bar{w})\,, \tag{10.8.8a}$$

$$\tilde{\psi}(\bar{w}+2\pi i t) = -R\tilde{\psi}(\bar{w})R^{-1} = -\exp(\pi i\tilde{\beta})\,\psi(w)\,, \tag{10.8.8b}$$

with the usual extra sign for fermionic fields. As indicated by the arrows in figure 10.3, these imply that

$$\psi(w+4\pi i t) = \exp[\pi i(\beta+\tilde{\beta})]\,\psi(w)\,. \tag{10.8.9}$$

The NS–NS exchange, from the sectors antiperiodic under $\sigma^2 \to \sigma^2 + 4\pi t$, then comes from traces weighted by $\Omega\exp(\pi i F)$ or $\Omega\exp(\pi i\tilde{F})$; further, these two traces are equal. Both NS–NS and R–R states contribute to the traces, making the separate contributions[8]

$$\text{NS–NS:}\quad q^{-1/3}\prod_{m=1}^{\infty}(1+q^{2m-1})^8 = Z^0{}_0(2it)^4\,, \tag{10.8.10a}$$

$$\text{R–R:}\quad -16q^{2/3}\prod_{m=1}^{\infty}(1+q^{2m})^8 = -Z^1{}_0(2it)^4\,, \tag{10.8.10b}$$

where $q = \exp(-2\pi t)$.

The full Klein bottle contribution to the NS–NS exchange is then

$$\begin{aligned} Z_{K_2,0} &= iV_{10}\int_0^\infty \frac{dt}{8t}\,(4\pi^2\alpha' t)^{-5}\eta(2it)^{-8}\left[Z^0{}_0(2it)^4 - Z^1{}_0(2it)^4\right] \\ &= i\frac{2^{10}V_{10}}{8\pi(8\pi^2\alpha')^5}\int_0^\infty ds\,\eta(is/\pi)^{-8}\left[Z^0{}_0(is/\pi)^4 - Z^0{}_1(is/\pi)^4\right] \\ &= i\frac{2^{10}V_{10}}{8\pi(8\pi^2\alpha')^5}\int_0^\infty ds\,[16+O(\exp(-2s))]\,, \end{aligned} \tag{10.8.11}$$

and $Z_{K_2,1} = -Z_{K_2,0}$. The bosonic part is (7.4.15) converted to $D = 10$.

The Möbius strip

In the open string Ω acts as

$$\Omega\psi^\mu(w)\Omega^{-1} = \tilde{\psi}^\mu(\pi-\bar{w}) = \psi^\mu(w-\pi)\,, \tag{10.8.12}$$

using the doubling trick (10.2.15). In terms of the modes this is

$$\Omega\psi_r^\mu\Omega^{-1} = \exp(-\pi i r)\,\psi_r^\mu\,. \tag{10.8.13}$$

The phase is imaginary in the NS sector and squares to -1. Thus

$$\Omega^2 = \exp(\pi i F)\,. \tag{10.8.14}$$

Since $\exp(\pi i F) = 1$ by the GSO projection, this is physically the same as squaring to the identity, but the combined Ω and GSO projections require

[8] In evaluating these, note that only states with identical ψ and $\tilde{\psi}$ excitations contribute to traces containing Ω. The signs from $\exp(\pi i F)$ or $\exp(\pi i\tilde{F})$ just cancel the signs from anticommuting ψs past $\tilde{\psi}$s, so that all terms in each trace have the same sign. The overall sign in the NS–NS trace (positive) can be determined from the graviton, and the overall sign in the R–R trace (negative) from the argument (10.6.17).

a single projection operator

$$\frac{1+\Omega+\Omega^2+\Omega^3}{4} . \tag{10.8.15}$$

With $R = \Omega\exp(\pi i\beta F)$ in the trace, the fields have the periodicities

$$\psi^\mu(w+4\pi i t) = -\exp(\pi i\beta)\psi^\mu(w+2\pi i t-\pi) = \psi^\mu(w-2\pi) . \tag{10.8.16}$$

It follows that in the R sector of the trace the fields are periodic in the σ^2-direction, corresponding to the R–R exchange, while the NS sector of the trace gives the NS–NS exchange.

It is slightly easier to focus on the R–R exchange, where the traces with Ω and $\Omega\exp(\pi i F)$ sum to

$$\begin{aligned} &-16q^{1/3}\prod_{m=1}^{\infty}[1+(-1)^m q^m]^8 - (1-1)^4 q^{1/3}\prod_{m=1}^{\infty}[1-(-1)^m q^m]^8 \\ &\quad = Z^0{}_1(2it)^4 Z^1{}_0(2it)^4 . \end{aligned} \tag{10.8.17}$$

The full Möbius amplitude, rewriting the bosonic part slightly, is

$$\begin{aligned} Z_{M_2,1} &= \pm inV_{10}\int_0^\infty \frac{dt}{8t}(8\pi^2\alpha' t)^{-5}\frac{Z^0{}_1(2it)^4 Z^1{}_0(2it)^4}{\eta(2it)^8 Z^0{}_0(2it)^4} \\ &= \pm 2in\frac{2^5 V_{10}}{8\pi(8\pi^2\alpha')^5}\int_0^\infty ds\,\frac{Z^0{}_1(2is/\pi)^4 Z^1{}_0(2is/\pi)^4}{\eta(2is/\pi)^8 Z^0{}_0(2is/\pi)^4} \\ &= \pm 2in\frac{2^5 V_{10}}{8\pi(8\pi^2\alpha')^5}\int_0^\infty ds\,[16+O(\exp(-2s))] , \end{aligned} \tag{10.8.18}$$

where the upper sign is for $SO(n)$. We have used (7.4.22) in $D = 10$.

The total divergence from R–R exchange is

$$Z_1 = -i(n\mp 32)^2\frac{V_{10}}{8\pi(8\pi^2\alpha')^5}\int_0^\infty ds\,[16+O(\exp(-2s))] . \tag{10.8.19}$$

The R–R tadpole vanishes only for the gauge group $SO(32)$. For each world-sheet topology the NS–NS divergence is the negative of the R–R divergence, so the dilaton–graviton tadpole also vanishes for $SO(32)$. This calculation does not determine the sign of the tadpole, but it should be $n\mp 32$. That is, changing from a symplectic to orthogonal projection changes the sign of RP_2, not of the disk. This is necessary for unitarity: the number of cross-caps is conserved mod 2 when a surface is cut open, so the sign is not determined by unitarity; this is not the case for the number of boundaries.

Exercises

10.1 (a) Find the OPE of T_F with X^μ and ψ^μ.
(b) Show that the residues of the OPEs of the currents (10.1.9) are proportional to the superconformal variations (10.1.10).

10.2 (a) Verify the commutator (10.1.11), up to terms proportional to the equations of motion.
(b) Verify that the commutator of a conformal and a superconformal transformation is a superconformal transformation.

10.3 (a) Verify the OPE (10.1.13).
(b) Extend this to the linear dilaton SCFT (10.1.22).

10.4 Obtain the R and NS algebras (10.2.11) from the OPE.

10.5 From the Jacobi identity for the R–NS algebra, show that the coefficients of the central charge terms in $T_B T_B$ and $T_F T_F$ are related.

10.6 Express $\exp(\pi i F)$ explicitly in terms of mode operators in the R and NS sectors of the ψ^μ CFT.

10.7 Verify that the expectation value (10.3.7) has the appropriate behavior as $z_i \to \infty$, and show that together with the OPE this determines the result up to normalization.

10.8 Verify the weight of the fermionic ground state $\mathscr{A}_\nu$ for general real ν:
(a) from the commutator (2.7.8);
(b) from the mnemonic of section 2.9.
The most direct, but most time-consuming, method would be to find the relation between conformal and creation–annihilation normal ordering as in eq. (2.7.11).

10.9 By any of the above methods, determine the ghost normal ordering constants (10.4.5).

10.10 Enumerate the states corresponding to each term in the expansion (10.3.19), in both fermionic and bosonic form.

10.11 Find the fermionic operator F_n equivalent to $e^{\pm inH(z)}$. Here are two possible methods: build F_n iteratively in n by taking repeated operator products with $e^{\pm iH(z)}$; or deduce $\psi_m^\pm \cdot F_n$ directly from the OPE. Check your answer by comparing dimensions and fermion numbers.

10.12 By looking at the eigenvalues of S_a, verify the spinor decompositions (10.5.17).

10.13 (a) Verify the operator products (10.5.22).
(b) Using the Jacobi identity as in exercise 4.3, verify nilpotence of Q_B.

10.14 Work out the massless level of the open superstring in BRST quantization.

10.15 Consider a single complex fermion, with the spectrum summed over all sectors such that $v = \tilde{v}$ is a multiple of $1/2p$ for integer p. Impose the projection that the numbers of left- and right-moving excitations differ by a multiple of $2p$. Show that the spectrum is the same as that of a periodic scalar at radius $r = 1/p$. Show that this can be understood as a $\mathbf{Z}_p$ twist of the $r = 1$ theory. A further $\mathbf{Z}_{2q}$ twist of the T-dual $r = 2p$ theory produces an arbitrary rational value.

11
The heterotic string

11.1 World-sheet supersymmetries

In the last chapter we were led by guesswork to the idea of enlarging the world-sheet constraint algebra, adding the supercurrents $T_F(z)$ and $\tilde{T}_F(\bar{z})$. Now let us see how much further we can generalize this idea. We are looking for sets of holomorphic and antiholomorphic currents whose Laurent coefficients form a closed algebra.

Let us start by emphasizing the distinction between global symmetries and constraints. Global symmetries on the world-sheet are just like global symmetries in spacetime, implying relations between masses and between amplitudes. However, we are also singling out part of the symmetry to impose as a *constraint*, meaning that physical states must be annihilated by it, either in the OCQ or BRST sense. In the bosonic string, the spacetime Poincaré invariance was a global symmetry of the world-sheet theory, while the conformal symmetry was a constraint. Our present interest is in constraint algebras. In fact we will find only a very small set of possibilities, but some of the additional algebras we encounter will appear later as global symmetries.

To begin we should note that the set of candidate world-sheet symmetry algebras is very large. In the bosonic string, for example, any product of factors $\partial^n X^\mu$ is a holomorphic current. In most cases the OPE of such currents will generate an infinite number of new currents, which is probably too big an algebra to be useful. However, even restricting to algebras with finite numbers of currents leaves an infinite number of possibilities.

Let us focus first on the holomorphic currents. We have seen in section 2.9 that in a unitary CFT an operator is holomorphic if and only if it is of weight $(h, 0)$ with $h \geq 0$. Although the complete world-sheet theory with ghosts and timelike oscillators does not have a positive norm, the spatial part does and so is a unitary representation of the symmetry.

Because $\tilde{h} = 0$, the spin of the current is also equal to h. Also, by taking real and imaginary parts we can assume the currents to be Hermitian. Now let us consider some possibilities:

Spin $h \geq 2$. Algebras with spin > 2 currents are often referred to collectively as *W algebras*. Many are known, including several infinite families, but there is no complete classification. We will encounter one example in chapter 15, as a global symmetry of a CFT. There have been attempts to use some of these as constraint algebras. One complication is that the commutator of generators is in general a nonlinear function of the generators, making the construction of the BRST operator nontrivial. The few examples that have been constructed appear, upon gauge fixing, to be special cases of bosonic strings. Further, the geometric interpretation, analogous to the Riemann surface construction used to formulate bosonic string perturbation theory, is not clear. So we will restrict our attention to constraint algebras with $h \leq 2$. Also, CFTs can have multiple $(2,0)$ currents as global symmetries. The bosonic string has at least 27, namely the ghost energy-momentum tensor and the energy-momentum tensor for each X^μ field. However, only the sum of these has a geometric interpretation, in terms of conformal invariance, and so we will assume that there is precisely one $(2,0)$ constraint current which is the overall energy-momentum tensor.

Spin h not a multiple of $\frac{1}{2}$. For a current j of spin h,

$$j(z)j(0) \sim z^{-2h} \tag{11.1.1}$$

with a coefficient that can be shown by a positivity argument not to vanish. This is multi-valued if $2h$ is not an integer. Although there are again many known CFTs with such currents, the nonlocality of these currents leads to substantial complications if one tries to impose them as constraints. Attempts to construct such *fractional strings* have led only to partial results and it is not clear if such theories exist. So we will restrict our attention to h a multiple of $\frac{1}{2}$.

With these assumptions the possible algebras are very limited, with spins $0, \frac{1}{2}, 1, \frac{3}{2}$, and 2. Solution of the Jacobi identities allows only the algebras shown in table 11.1. The first two entries are of course the conformal and $N = 1$ superconformal algebras that we have already studied. The three $N = 4$ algebras are related. The second algebra is a special case of the first where the $U(1)$ current becomes the gradient of a scalar. The third is a subalgebra of the second.

The ghost central charge is determined by the number of currents of each spin. The central charge for the ghosts associated with a current of spin h is

$$c_h = (-1)^{2h+1}[3(2h-1)^2 - 1] , \tag{11.1.2a}$$

$$c_2 = -26 , \quad c_{3/2} = +11 , \quad c_1 = -2 , \quad c_{1/2} = -1 , \quad c_0 = -2 . \tag{11.1.2b}$$

Table 11.1. *World-sheet superconformal algebras. The number of currents of each spin and the total ghost central charge are listed, as are the global symmetry generated by the spin-1 currents and the transformation of the supercharges under these.*

$n_{3/2} \equiv N$	n_1	$n_{1/2}$	n_0	c^g	symmetry	T_F rep.
0	0	0	0	−26		
1	0	0	0	−15		
2	1	0	0	−6	$U(1)$	± 1
3	3	1	0	0	$SU(2)$	**3**
4	7	4	0	0	$SU(2) \times SU(2) \times U(1)$	**(2,2,0)**
4	6	4	1	0	$SU(2) \times SU(2)$	**(2,2)**
4	3	0	0	12	$SU(2)$	**2**

The sign $(-1)^{2h+1}$ takes into account the statistics of the ghosts, anticommuting for integer spin and commuting for half-integer spin. Since the matter central charge c^m is $-c^g$, there is only one new algebra, $N = 2$, that can have a positive critical dimension.

Actually, for $N = 0$ and $N = 1$ there can also be additional spin-1 and spin-$\frac{1}{2}$ constraints, provided the supercurrent is neutral under the corresponding symmetry. However, these larger algebras are not essentially different. The negative central charges of the ghosts allow additional matter, but the additional constraints precisely remove the added states so that these reduce to the old $N = 0$ and $N = 1$ theories. Nevertheless this construction is sometimes useful, as we will see in section 15.5.

For $N = 2$ it is convenient to join the two real supercurrents into one complex supercurrent

$$T_F^{\pm} = 2^{-1/2}(T_{F1} \pm iT_{F2}) . \tag{11.1.3}$$

The $N = 2$ algebra in operator product form is then

$$T_B(z)T_F^{\pm}(0) \sim \frac{3}{2z^2}T_F^{\pm}(0) + \frac{1}{z}\partial T_F^{\pm}(0) , \tag{11.1.4a}$$

$$T_B(z)j(0) \sim \frac{1}{z^2}j(0) + \frac{1}{z}\partial j(0) , \tag{11.1.4b}$$

$$T_F^{+}(z)T_F^{-}(0) \sim \frac{2c}{3z^3} + \frac{2}{z^2}j(0) + \frac{2}{z}T_B(0) + \frac{1}{z}\partial j(0) , \tag{11.1.4c}$$

$$T_F^{+}(z)T_F^{+}(0) \sim T_F^{-}(z)T_F^{-}(0) \sim 0 , \tag{11.1.4d}$$

$$j(z)T_F^{\pm}(0) \sim \pm\frac{1}{z}T_F^{\pm}(0) , \tag{11.1.4e}$$

$$j(z)j(0) \sim \frac{c}{3z^2} . \tag{11.1.4f}$$

In particular this implies that $T_F^{\pm}$ and j are primary fields and that $T_F^{\pm}$ has charge ± 1 under the $U(1)$ generated by j. The constant c in $T_F^{+}T_F^{-}$

and jj must be the central charge. This follows from the Jacobi identity for the modes, but we will not write out the mode expansion in full until chapter 19, where we will have more need of it.

The smallest linear representation of the $N = 2$ algebra has two real scalars and two real fermions, which we join into a complex scalar Z and complex fermion ψ. The action is

$$S = \frac{1}{2\pi}\int d^2z \left(\partial\bar{Z}\bar{\partial}Z + \bar{\psi}\bar{\partial}\psi + \tilde{\psi}\partial\tilde{\psi}\right) . \tag{11.1.5}$$

The currents are

$$T_B = -\partial\bar{Z}\partial Z - \frac{1}{2}(\bar{\psi}\partial\psi + \psi\partial\bar{\psi}) , \quad j = -\bar{\psi}\psi , \tag{11.1.6a}$$

$$T_F^+ = 2^{1/2} i\psi\partial\bar{Z} , \quad T_F^- = 2^{1/2} i\bar{\psi}\partial Z . \tag{11.1.6b}$$

There is also a set of antiholomorphic currents, so this $Z\psi\tilde{\psi}$ CFT has $(2,2)$ superconformal symmetry.

The central charge of the $Z\psi\tilde{\psi}$ CFT is 3, so two copies will cancel the ghost central charge. Since there are two real scalars in each CFT the critical dimension is 4. However, these dimensions come in complex pairs, so that the spacetime signature can be purely Euclidean, or $(2,2)$, but not the Minkowski $(3,1)$. Further, while the theory has four-dimensional translational invariance it does not have four-dimensional Lorentz invariance — the dimensions are paired together in a definite way in the supercharges. Instead the symmetry is $U(2)$ or $U(1,1)$, complex rotations on the two Zs. Finally, the spectrum is quite small. The constraints fix two full sets of $Z\psi\tilde{\psi}$ oscillators (the analog of the light-cone gauge), leaving none. Thus there is just the center-of-mass motion of a single state. This has some mathematical interest, but whether it has physical applications is more conjectural.

Thus we have reduced what began as a rather large set of possible algebras down to the original $N = 0$ and $N = 1$. There is, however, another generalization, which is to have different algebras on the left- and right-moving sides of the closed string. The holomorphic and antiholomorphic algebras commute and there is no reason that they should be the same. In the open string, the boundary conditions relate the holomorphic and antiholomorphic currents so there is no analogous construction.

This allows the one new possibility, the $(N, \tilde{N}) = (0,1)$ *heterotic string*; $(N, \tilde{N}) = (1,0)$ would be the same on redefining $z \to \bar{z}$. We study this new algebra in detail in the remainder of the chapter. In addition the $(0,2)$ and $(1,2)$ heterotic string theories are mathematically interesting and may have a less direct physical relevance.

It should be emphasized that the analysis in this section had many explicit and implicit assumptions, and one should be cautious in assuming that all string theories have been found. Indeed, there are some string theories that do not fall into this classification. One is the Green–Schwarz form of the superstring. This has no simple covariant gauge-fixing, but in the light-cone gauge it is in fact equivalent to the RNS superstring, via bosonization. We will not have space to develop this in detail, but will see a hint of it in chapter 12. Another exception is *topological string theory*, where in a covariant gauge the constraints do not satisfy spin-statistics as we have assumed. This string theory has no physical degrees of freedom, but is of mathematical interest in that its observables are topological.

In fact, we will find the same set of physical string theories from an entirely different and nonperturbative point of view in chapter 14, suggesting that all have been found. To be precise, there are other theories with stringlike excitations, but the theories found in this and the previous chapter seem to be the only ones which have a limit where they become weakly coupled, so that a string perturbation theory exists.

11.2 The $SO(32)$ and $E_8 \times E_8$ heterotic strings

The $(0,1)$ heterotic string combines the constraints and ghosts from the left-moving side of the bosonic string with those from the right-moving side of the type II string. We could try to go further and combine the whole left-moving side of the bosonic string, with 26 flat dimensions, with the ten-dimensional right-moving side of the type II string. In fact this can be done, but since its physical meaning is not so clear we will for now keep the same number of dimensions on both sides. The maximum is then ten, from the superconformal side. We begin with the fields

$$X^\mu(z,\bar{z}) , \quad \tilde{\psi}^\mu(\bar{z}) , \quad \mu = 0,\ldots,9 , \tag{11.2.1}$$

with total central charge $(c,\tilde{c}) = (10,15)$. The ghost central charges add up to $(c^g,\tilde{c}^g) = (-26,-15)$, so the remaining matter has $(c,\tilde{c}) = (16,0)$. The simplest possibility is to take 32 left-moving spin-$\frac{1}{2}$ fields

$$\lambda^A(z) , \quad A = 1,\ldots,32 . \tag{11.2.2}$$

The total matter action is

$$S = \frac{1}{4\pi}\int d^2z \left(\frac{2}{\alpha'}\partial X^\mu\bar{\partial}X_\mu + \lambda^A\bar{\partial}\lambda^A + \tilde{\psi}^\mu\partial\tilde{\psi}_\mu\right) . \tag{11.2.3}$$

The operator products are

$$X^\mu(z,\bar{z})X^\nu(0,0) \sim -\eta^{\mu\nu}\frac{\alpha'}{2}\ln|z|^2 , \tag{11.2.4a}$$

$$\lambda^A(z)\lambda^B(0) \sim \delta^{AB}\frac{1}{z} , \tag{11.2.4b}$$

$$\tilde{\psi}^\mu(\bar{z})\tilde{\psi}^\nu(0) \sim \eta^{\mu\nu}\frac{1}{\bar{z}} . \tag{11.2.4c}$$

The matter energy-momentum tensor and supercurrent are

$$T_B = -\frac{1}{\alpha'}\partial X^\mu \partial X_\mu - \frac{1}{2}\lambda^A \partial \lambda^A \,, \tag{11.2.5a}$$

$$\tilde{T}_B = -\frac{1}{\alpha'}\bar{\partial} X^\mu \bar{\partial} X_\mu - \frac{1}{2}\tilde{\psi}^\mu \bar{\partial}\tilde{\psi}_\mu \,, \tag{11.2.5b}$$

$$\tilde{T}_F = i(2/\alpha')^{1/2}\tilde{\psi}^\mu \partial X_\mu \,. \tag{11.2.5c}$$

The world-sheet theory has symmetry $SO(9,1) \times SO(32)$. The $SO(32)$, acting on the λ^A, is an internal symmetry. In particular, none of the λ^A can have a timelike signature because there is no fermionic constraint on the left-moving side to remove states of negative norm. So while the action for the λ^A is the same as for the $\tilde{\psi}^\mu$ of the RNS superstring, the resulting theory is very different because of the constraints.

The right-moving ghosts are the same as in the RNS superstring, and the left-movers the same as in the bosonic string. It is straightforward to construct the nilpotent BRST charge and show the no-ghost theorem, with any BRST-invariant periodicity conditions. As usual this still holds if we replace any of the spatial $(X^\mu, \tilde{\psi}^\mu)$ and the λ^A with a unitary (1,0) SCFT of the equivalent central charge.

To finish the description of the theory, we need to give the boundary conditions on the fields and specify which sectors are in the spectrum. This is more complicated than in the type II strings, because now neither Poincaré nor BRST invariance require common boundary conditions on all the λ^A. Periodicity of T_B only requires that the λ^A be periodic up to an arbitrary $O(32)$ rotation,

$$\lambda^A(w + 2\pi) = O^{AB}\lambda^B(w) \,. \tag{11.2.6}$$

We will not carry out a systematic search for consistent theories as we did for the RNS string, but will describe all the known theories. Nine ten-dimensional theories based on the action (11.2.3) are known, though six have tachyons and so are consistent only in the same sense as the bosonic string. Of the three tachyon-free theories, two have spacetime supersymmetry and these are our main interest. In this section we construct the two supersymmetric theories and in the next the seven nonsupersymmetric theories.

In the IIA and IIB superstrings the GSO projection acted separately on the left- and right-moving sides. This will be also true in any supersymmetric heterotic theory. The world-sheet current associated with spacetime symmetry is $\mathscr{V}_\mathbf{s}$ as in eq. (10.4.25), with $\mathbf{s}$ in the **16**. In order for the corresponding charge to be well defined, the OPE of this current with any vertex operator must be single-valued. For the right-moving spinor part

of the vertex operator, the spin eigenvalue $\mathbf{s}'$ must then satisfy

$$\mathbf{s}\cdot\mathbf{s}' + \frac{l}{2} \in \mathbf{Z} \tag{11.2.7}$$

for all $\mathbf{s} \in \mathbf{16}$, where l is -1 in the NS sector and $-\frac{1}{2}$ in the R sector. Taking $\mathbf{s} = (\frac{1}{2},\frac{1}{2},\frac{1}{2},\frac{1}{2},\frac{1}{2})$, this condition is precisely the right-moving GSO projection

$$\exp(\pi i \tilde{F}) = 1 \,; \tag{11.2.8}$$

any other $\mathbf{s} \in \mathbf{16}$ gives the same condition.

Now let us try a GSO projection on the left-moving spinors also. That is, we take periodicities

$$\lambda^A(w + 2\pi) = \pm\lambda^A(w) \tag{11.2.9}$$

with the same sign on all 32 components, and impose

$$\exp(\pi i F) = 1 \tag{11.2.10}$$

for the left-moving fermion number. It is easily verified by means of bosonization that the OPE is local and closed, just as in the IIA and IIB strings. Combine the 32 real fermions into 16 complex fermions,

$$\lambda^{K\pm} = 2^{-1/2}(\lambda^{2K-1} \pm i\lambda^{2K}) \,, \quad K = 1,\ldots,16 \,. \tag{11.2.11}$$

These can then be bosonized in terms of 16 left-moving scalars $H^K(z)$. By analogy to the definition of F in the type II string define

$$F = \sum_{K=1}^{16} q_K \,, \tag{11.2.12}$$

where $\lambda^{K\pm}$ has $q_K = \pm 1$. Then F is additive so the OPE is closed, and the projection (11.2.10) guarantees that there are no branch cuts with the R sector vertex operators. Note that in the bosonized description we have 26 left-moving and 10 right-moving bosons, so the theory (11.2.3) really is a fusion (heterosis) of the bosonic and type II strings. We will emphasize the fermionic description in the present section, returning to the bosonic description later.

Modular invariance is straightforward. The partition function for the λ is

$$Z_{16}(\tau) = \frac{1}{2}\left[Z^0{}_0(\tau)^{16} + Z^0{}_1(\tau)^{16} + Z^1{}_0(\tau)^{16} + Z^1{}_1(\tau)^{16}\right] . \tag{11.2.13}$$

The modular transformations just permute the four terms, with no phase under $\tau \to -1/\tau$ and a phase of $\exp(2\pi i/3)$ under $\tau \to \tau + 1$. The latter cancels the opposite phase from the partition function $Z_\psi^+(\tau)^*$ of $\tilde{\psi}$. The form (11.2.13) parallels that of $Z_\psi^+(\tau)$ in the type II string but with all +

signs. This is necessary from several points of view. With 32 rather than 8 fermions, the signs in the modular transformations are raised to the fourth power and so the first three terms must enter with a common sign. As usual the $Z^1{}_1$ term transforms only into itself and its sign depends on the chirality in the R sector. Three other theories, defined by flipping the chirality in one or both R sectors, are physically equivalent. Also, the relative minus sign in the first and second terms of $Z_\psi^+(\tau)$ came from the F of the superconformal ghosts, which we do not have on the left-moving side of the heterotic string. The relative minus sign in the first and third terms came from spacetime statistics, but the λ are spacetime scalars and so are their R sector states. So modular invariance, conservation of F by the OPE, and spacetime spin-statistics are all consistent with the partition function (11.2.13).

We now find the lightest states. The right-moving side is the same as in the type II string, with no tachyon and $\mathbf{8}_v + \mathbf{8}$ at the massless level. On the left-moving side, the normal ordering constant in the left-moving transverse Hamiltonian $H^\perp = \alpha' m^2/4$ is

$$\text{NS:}\ -\frac{8}{24}-\frac{32}{48}=-1\ ,\quad \text{R:}\ -\frac{8}{24}+\frac{32}{24}=+1\ . \tag{11.2.14}$$

The left-moving NS ground state is therefore a tachyon. The first excited states

$$\lambda^A_{-1/2}|0\rangle_{\text{NS}} \tag{11.2.15}$$

have $H^\perp = -\frac{1}{2}$ but are removed by projection (11.2.10): the NS ground state now has $\exp(\pi i F) = +1$ because there is no contribution from ghosts. A state of $H^\perp = 0$ can be obtained in two ways:

$$\alpha^i_{-1}|0\rangle_{\text{NS}}\ ,\quad \lambda^A_{-1/2}\lambda^B_{-1/2}|0\rangle_{\text{NS}}\ . \tag{11.2.16}$$

The λ^A transform under an $SO(32)$ internal symmetry. Under the full symmetry $SO(8)$ spin $\times SO(32)$, the NS ground state is invariant, $(\mathbf{1},\mathbf{1})$. The second state in (11.2.16) is antisymmetric under $A \leftrightarrow B$, so the massless states (11.2.16) transform as $(\mathbf{8}_v,\mathbf{1}) + (\mathbf{1},[2])$. The antisymmetric tensor representation is the adjoint of $SO(32)$, with dimension $32 \times 31/2 = \mathbf{496}$.

Table 11.2 summarizes the tachyonic and massless states on each side. The left-movers are given with their $SO(8) \times SO(32)$ quantum numbers and the right-movers with their $SO(8)$ quantum numbers. Closed string states combine left- and right-moving states at the same mass. The left-moving side, like the bosonic string, has a would-be tachyon, but there is no right-mover to pair it with so the theory is tachyon-free. At the massless level, the product

$$(\mathbf{8}_v,\mathbf{1}) \times (\mathbf{8}_v+\mathbf{8}) = (\mathbf{1},\mathbf{1}) + (\mathbf{28},\mathbf{1}) + (\mathbf{35},\mathbf{1}) + (\mathbf{56},\mathbf{1}) + (\mathbf{8}',\mathbf{1}) \tag{11.2.17}$$

Table 11.2. *Low-lying heterotic string states.*

m^2	NS	R	$\tilde{\text{NS}}$	$\tilde{\text{R}}$
$-4/\alpha'$	$(\mathbf{1},\mathbf{1})$	-	-	-
0	$(\mathbf{8}_v,\mathbf{1}) + (\mathbf{1},\mathbf{496})$	-	$\mathbf{8}_v$	$\mathbf{8}$

is the type I supergravity multiplet. The product

$$(\mathbf{1},\mathbf{496}) \times (\mathbf{8}_v+\mathbf{8}) = (\mathbf{8}_v,\mathbf{496}) + (\mathbf{8},\mathbf{496}) \tag{11.2.18}$$

is an $N = 1$ gauge multiplet in the adjoint of $SO(32)$. The latter is therefore a gauge symmetry in spacetime.

This is precisely the same massless content as the type I open plus closed $SO(32)$ theory. However, these two theories have different massive spectra. In the open string, the gauge quantum numbers are carried by an $SO(32)$ vector at each endpoint, so even at the massive levels there will never be more than a rank 2 tensor representation of the gauge group. In the heterotic string, the gauge quantum numbers are carried by fields that propagate on the whole world sheet. At massive levels any number of these can be excited, allowing arbitrarily large representations of the gauge group. Remarkably, however, we will see in chapter 14 that the type I and heterotic $SO(32)$ theories are one and the same.

The second heterotic string theory is obtained by dividing the λ^A into two sets of 16 with independent boundary conditions,

$$\lambda^A(w+2\pi) = \begin{cases} \eta\lambda^A(w)\ , & A = 1,\dots,16\ , \\ \eta'\lambda^A(w)\ , & A = 17,\dots,32\ , \end{cases} \tag{11.2.19}$$

with η and η' each ± 1. Correspondingly, there are the operators

$$\exp(\pi i F_1)\ ,\quad \exp(\pi i F_1')\ , \tag{11.2.20}$$

which anticommute with λ^A for $A = 1,\dots,16$ and $A = 17,\dots,32$ respectively. Take separate GSO projections on the right-movers and the two sets of left-movers. That is, sum over the $2^3 = 8$ possible combinations of periodicities with the projections

$$\exp(\pi i F_1) = \exp(\pi i F_1') = \exp(\pi i \tilde{F}) = 1\ . \tag{11.2.21}$$

Again closure and locality of the OPE and modular invariance are easily verified. In particular the partition function

$$Z_8(\tau)^2 = \frac{1}{4}\left[Z^0{}_0(\tau)^8 + Z^0{}_1(\tau)^8 + Z^1{}_0(\tau)^8 + Z^1{}_1(\tau)^8\right]^2 \tag{11.2.22}$$

transforms in the same way as $Z_\psi^\pm$ and Z_{16}. It is important here that the fermions are in groups of 16, so that the minus signs from $Z_\psi^\pm$ (which was for eight fermions) are squared.

As before, the lightest states on the right-hand side are the massless $\mathbf{8}_v + \mathbf{8}$. On the left-hand side, the sector NS–NS′ again has a normal ordering constant of -1, so the ground state is tachyonic but finds no matching state on the right. The first excited states, at $m^2 = 0$, are

$$\alpha^i_{-1}|0\rangle_{\mathrm{NS,NS'}}\ ,$$
$$\lambda^A_{-1/2}\lambda^B_{-1/2}|0\rangle_{\mathrm{NS,NS'}}\ ,\quad 1 \le A,B \le 16 \text{ or } 17 \le A,B \le 32\ . \tag{11.2.23}$$

There is a difference here from the $SO(32)$ case: because there are separate GSO projections on each set of 16, A and B must come from the same set. Since the $SO(32)$ symmetry is partly broken by the boundary conditions, we classify states by the surviving $SO(16) \times SO(16)$. The states (11.2.23) include the antisymmetric tensor adjoint representation for each $SO(16)$, with dimension $16 \times 15/2 = \mathbf{120}$.

In the left-moving R–NS′ sector the normal ordering constant is

$$-\frac{8}{24} + \frac{16}{24} - \frac{16}{48} = 0\ , \tag{11.2.24}$$

so the ground states are massless. Making eight raising and eight lowering operators out of the 16 λ^A zero modes produces a 256-dimensional spinor representation of the first $SO(16)$. The GSO projection separates it into two irreducible representations, $\mathbf{128} + \mathbf{128}'$, the former being in the spectrum. The NS–R′ sector produces a $\mathbf{128}$ of the other $SO(16)$, and the R–R′ sector again has no massless states.

In all, the $SO(8)$ spin $\times\ SO(16) \times SO(16)$ content of the massless level of the left-hand side is

$$(\mathbf{8}_v, \mathbf{1}, \mathbf{1}) + (\mathbf{1}, \mathbf{120}, \mathbf{1}) + (\mathbf{1}, \mathbf{1}, \mathbf{120}) + (\mathbf{1}, \mathbf{128}, \mathbf{1}) + (\mathbf{1}, \mathbf{1}, \mathbf{128})\ . \tag{11.2.25}$$

Combining these with the right-moving $\mathbf{8}_v$ gives, for each $SO(16)$, massless vector bosons which transform as $\mathbf{120} + \mathbf{128}$. Consistency of the spacetime theory requires that massless vectors transform under the *adjoint* representation of the gauge group. There is indeed a group, the exceptional group E_8, that has an $SO(16)$ subgroup under which the E_8 adjoint $\mathbf{248}$ transforms as $\mathbf{120} + \mathbf{128}$. Evidently this second heterotic string theory has gauge group $E_8 \times E_8$. The world-sheet theory has a full $E_8 \times E_8$ symmetry, even though only an $SO(16) \times SO(16)$ symmetry is manifest in the present description. The additional currents are given by bosonization as

$$\exp\left[i\sum_{K=1}^{16} q_K H^K(z)\right]\ . \tag{11.2.26}$$

This is a spin field, just as in the fermion vertex operator (10.4.25). For the first E_8 the charges are

$$q_K = \begin{cases} \pm\frac{1}{2}\ , & K = 1,\dots 8 \\ 0\ , & K = 9,\dots 16 \end{cases}\ ,\qquad \sum_{K=1}^{16} q_K \in 2\mathbf{Z}\ , \tag{11.2.27}$$

and vice versa for the second. These are indeed $(1,0)$ operators. The massless spectrum is the $d = 10$, $N = 1$ supergravity multiplet plus an $E_8 \times E_8$ gauge multiplet. The $SO(8)$ spin $\times\ E_8 \times E_8$ quantum numbers of the massless fields are

$$(\mathbf{1},\mathbf{1},\mathbf{1}) + (\mathbf{28},\mathbf{1},\mathbf{1}) + (\mathbf{35},\mathbf{1},\mathbf{1}) + (\mathbf{56},\mathbf{1},\mathbf{1}) + (\mathbf{8}',\mathbf{1},\mathbf{1})$$
$$+ (\mathbf{8}_v,\mathbf{248},\mathbf{1}) + (\mathbf{8},\mathbf{248},\mathbf{1}) + (\mathbf{8}_v,\mathbf{1},\mathbf{248}) + (\mathbf{8},\mathbf{1},\mathbf{248})\ . \tag{11.2.28}$$

Consistency requires the fermions to be in groups of 16. We could make a modular-invariant theory using groups of eight, the left-moving partition function being $(Z^{\pm}_{\psi})^4$. However, we have seen that modular invariance requires minus signs in $Z^{\pm}_{\psi}$. These signs would give negative weight to left-moving R sector states and would correspond to the projection $\exp(\pi i F) = -1$ in the NS sector. The first is inconsistent with spin-statistics because these states are spacetime scalars, and the second is inconsistent with closure of the OPE thus making the interactions inconsistent. The $SO(32)$ and $E_8 \times E_8$ theories are the only supersymmetric heterotic strings in ten dimensions.

11.3 Other ten-dimensional heterotic strings

The other heterotic string theories can all be constructed from a single theory by the twisting construction introduced in section 8.5. The 'least twisted' theory, in the sense of having the smallest number of path integral sectors, corresponds to the diagonal modular invariant

$$\frac{1}{4}\Big[Z^0{}_0(\tau)^{16}Z^0{}_0(\tau)^{*4} - Z^0{}_1(\tau)^{16}Z^0{}_1(\tau)^{*4}$$
$$-Z^1{}_0(\tau)^{16}Z^1{}_0(\tau)^{*4} - Z^1{}_1(\tau)^{16}Z^1{}_1(\tau)^{*4}\Big]\ . \tag{11.3.1}$$

This invariant corresponds to taking all fermions, λ^A and $\tilde{\psi}^\mu$, to be simultaneously periodic or antiperiodic on each cycle of the torus. In terms of the spectrum the world-sheet fermions are either all R or all NS, with the diagonal projection

$$\exp[\pi i(F + \tilde{F})] = 1\ . \tag{11.3.2}$$

This theory is consistent except for a tachyon, the state

$$\lambda^A_{-1/2}|0\rangle_{\mathrm{NS,NS}}\ ,\quad m^2 = -\frac{2}{\alpha'}\ ,\quad \exp(\pi i F) = \exp(\pi i \tilde{F}) = -1\ , \tag{11.3.3}$$

which transforms as a vector under $SO(32)$. At the massless level are the states

$$\alpha^i_{-1}\tilde{\psi}^j_{-1/2}|0\rangle_{\mathrm{NS,NS}}\,, \quad \lambda^A_{-1/2}\lambda^B_{-1/2}\tilde{\psi}^j_{-1/2}|0\rangle_{\mathrm{NS,NS}}\,, \tag{11.3.4}$$

which are the graviton, dilaton, antisymmetric tensor and $SO(32)$ gauge bosons. There are fermions in the spectrum, but the lightest are at $m^2 = 4/\alpha'$.

Now let us twist by various symmetries. Consider first the $\mathbf{Z}_2$ generated by $\exp(\pi i\tilde{F})$. Combined with the diagonal projection (11.3.2) this gives the total projection

$$\frac{1+\exp[\pi i(F+\tilde{F})]}{2}\cdot\frac{1+\exp(\pi i\tilde{F})}{2} = \frac{1+\exp(\pi iF)}{2}\cdot\frac{1+\exp(\pi i\tilde{F})}{2}\,. \tag{11.3.5}$$

This is the same as the projections (11.2.8) plus (11.2.10) defining the supersymmetric $SO(32)$ heterotic string. Also, the spatial twist by $\exp(\pi i\tilde{F})$ adds in the sectors in which the λ^A and $\tilde{\psi}^\mu$ have opposite periodicities. The twisted theory is thus the $SO(32)$ heterotic string. Twisting by $\exp(\pi iF)$ has the same effect.

Now consider twisting the diagonal theory by $\exp(\pi iF_1)$, which flips the sign of the first 16 λ^A and which was used to construct the $E_8\times E_8$ heterotic string. The resulting theory is nonsupersymmetric — as in eq. (11.2.8), a theory will be supersymmetric if and only if the projections include $\exp(\pi i\tilde{F}) = 1$. It has gauge group $E_8 \times SO(16)$ and a tachyon in the $(\mathbf{1},\mathbf{16})$. We leave it to the reader to verify this. A further twist by $\exp(\pi i\tilde{F})$ produces the supersymmetric $E_8 \times E_8$ heterotic string.

One can carry this further by dividing the λ^A into groups of 8, 4, 2, and 1 as follows. Write the SO(32) index A in binary form, $A = 1+d_1d_2d_3d_4d_5$, where each of the digits d_i is zero or one. Define the operators $\exp(\pi iF_i)$ for $i = 1,\ldots,5$ to anticommute with those λ^A having $d_i = 0$ and commute with those having $d_i = 1$. There are essentially five possible twist groups, with 2, 4, 8, 16, or 32 elements, generated respectively by choosing one, two, three, four or five of the $\exp(\pi iF_i)$ and forming all products. The first of these produces the $E_8\times SO(16)$ theory just described; the further twists produce the gauge groups $SO(24)\times SO(8)$, $E_7\times E_7\times SO(4)$, $SU(16)\times SO(2)$, and E_8. None of these theories is supersymmetric, and all have tachyons. A further twist by $\exp(\pi i\tilde{F})$ produces a supersymmetric theory which in each case is either the $SO(32)$ theory or the $E_8 \times E_8$ theory. These gauge symmetries are less manifest in this construction, with more of the currents coming from R sectors.

Let us review the logic of the twisting construction. The vertex operator corresponding to a sector twisted by a group element h produces branch cuts in the fields, but the projection onto h-invariant states means that these branch cuts do not appear in the products of vertex operators.

Since h is a symmetry the projection is preserved by interactions. On the torus, the sum over spatial and timelike twists is modular-invariant, and this generalizes to any genus. However, we have learned in section 10.7 that naive modular invariance of the sum over path integral boundary conditions is not enough, because in general there are anomalous phases in the modular transformations. Only for a right–left symmetric path integral do the phases automatically cancel.

At one loop the anomalous phases appear only in the transformation $\tau \to \tau + 1$, where they amount to a failure of the level-matching condition $L_0 - \tilde{L}_0 \in \mathbf{Z}$. It is further a theorem that for an Abelian twist group (like the products of $\mathbf{Z}_2$s considered here), the one-loop amplitude and in fact all amplitudes are modular-invariant precisely if in every twisted sector, before imposing the projection, there is an infinite number of level-matched states. The projection can then be defined to satisfy level matching. In the heterotic string, taking a sector in which k of the λ^A satisfy R boundary conditions and $32-k$ satisfy NS boundary conditions, the zero-point energy is

$$-\frac{8}{24}+\frac{k}{24}-\frac{(32-k)}{48} = -1+\frac{k}{16}\,. \tag{11.3.6}$$

The oscillators raise the energy by a multiple of $\frac{1}{2}$, so the energies on the left-moving side are $\frac{1}{16}k \bmod \frac{1}{2}$. On the right-moving side we are still taking the fermions to have common boundary conditions for Lorentz invariance, so the energies are multiples of $\frac{1}{2}$. Thus the level-matching condition is satisfied precisely if k is a multiple of eight. Closure of the OPE and spacetime spin-statistics actually require k to be a multiple of 16, as we have seen. The twists $\exp(\pi iF_i)$ were defined so that any product of them anticommutes with exactly 16 of the λ^A, satisfying this condition.

When the level-matching condition is satisfied, there can in fact be more than one modular-invariant and consistent theory. Consider a twisted theory with partition function

$$Z = \frac{1}{\mathrm{order}(H)}\sum_{h_1,h_2\in H} Z_{h_1,h_2}\,, \tag{11.3.7}$$

where h_1 and h_2 are the spatial and timelike periodicities on the torus. Then the theory with partition function

$$Z = \frac{1}{\mathrm{order}(H)}\sum_{h_1,h_2\in H} \epsilon(h_1,h_2)Z_{h_1,h_2} \tag{11.3.8}$$

is also consistent (modular-invariant with closed and local OPE) provided

that the phases $\epsilon(h_1, h_2)$ satisfy

$$\epsilon(h_1, h_2) = \epsilon(h_2, h_1)^{-1} , \tag{11.3.9a}$$

$$\epsilon(h_1, h_2)\epsilon(h_1, h_3) = \epsilon(h_1, h_2 h_3) , \tag{11.3.9b}$$

$$\epsilon(h, h) = 1 . \tag{11.3.9c}$$

In terms of $\hat{h}_2$ defined in the original twisted theory, the new twisted theory is no longer projected onto H-invariant states, but onto states satisfying

$$\hat{h}_2 |\psi\rangle_{h_1} = \epsilon(h_1, h_2)^{-1} |\psi\rangle_{h_1} \tag{11.3.10}$$

in the sector twisted by h_1. In other words, states are now eigenvectors of $\hat{h}$, with a sector-dependent phase; equivalently we have made a sector-dependent redefinition

$$\hat{h} \to \epsilon(h_1, h)\hat{h} . \tag{11.3.11}$$

The phase factor $\epsilon(h_1, h_2)$ is known as *discrete torsion.*

There is one interesting possibility for discrete torsion in the theories above, in the group generated by $\exp(\pi i F_1)$ and $\exp(\pi i \tilde{F})$ that produces the $E_8 \times E_8$ string from the diagonal theory. For

$$(h_1, h_2) = \left(\exp[\pi i(k_1 F_1 + l_1 \tilde{F})], \exp[\pi i(k_2 F_1 + l_2 \tilde{F})]\right) \tag{11.3.12}$$

the phase

$$\epsilon(h_1, h_2) = (-1)^{k_1 l_2 + k_2 l_1} \tag{11.3.13}$$

satisfies the conditions (11.3.9). It modifies the projection from

$$\exp(\pi i F_1) = \exp(\pi i F_1') = \exp(\pi i \tilde{F}) = 1 , \tag{11.3.14}$$

which produces to the supersymmetric $E_8 \times E_8$ string, to

$$\exp[\pi i(F_1 + \alpha_1' + \tilde{\alpha})] = \exp[\pi i(F_1' + \alpha_1 + \tilde{\alpha})] = \exp[\pi i(\tilde{F} + \alpha_1 + \alpha_1')] = 1 . \tag{11.3.15}$$

The notation parallels that in eq. (10.7.11): under $w \to w + 2\pi$, the $\tilde{\psi}^\mu$, the first 16 λ^A, and the second 16 λ^A pick up phases $-\exp(-\pi i\tilde{\alpha})$, $-\exp(\pi i \alpha_1)$, and $-\exp(\pi i \alpha_1')$ respectively. The αs are conserved by the OPE so the projections are consistent. In other words, the spectrum consists of the sectors

$$\begin{aligned} &(\mathrm{NS}+, \mathrm{NS}+, \mathrm{NS}+) , \\ &(\mathrm{NS}-, \mathrm{NS}-, \mathrm{R}+) , \ (\mathrm{NS}-, \mathrm{R}+, \mathrm{NS}-) , \ (\mathrm{NS}+, \mathrm{R}-, \mathrm{R}-) , \\ &(\mathrm{R}+, \mathrm{NS}-, \mathrm{NS}-) , \ (\mathrm{R}-, \mathrm{R}-, \mathrm{NS}+) , \ (\mathrm{R}-, \mathrm{NS}+, \mathrm{R}-) , \\ &(\mathrm{R}+, \mathrm{R}+, \mathrm{R}+) \end{aligned}$$

where the notation refers respectively to the $\tilde{\psi}^\mu$, the first 16 λ^A, and the second 16 λ^A.

Gravitinos, in the sectors $(\mathrm{R}\pm, \mathrm{NS}+, \mathrm{NS}+)$, are absent from the spectrum. So also are tachyons, which are in the $(\mathrm{NS}-, \mathrm{NS}-, \mathrm{NS}+)$ and $(\mathrm{NS}-, \mathrm{NS}+, \mathrm{NS}-)$ sectors. The twists leave an $SO(16) \times SO(16)$ gauge symmetry. Classifying states by their $SO(8)$ spin $\times SO(16) \times SO(16)$ quantum numbers, one finds the massless spectrum

$$\begin{aligned} (\mathrm{NS}+, \mathrm{NS}+, \mathrm{NS}+) : \ & (\mathbf{1}, \mathbf{1}, \mathbf{1}) + (\mathbf{28}, \mathbf{1}, \mathbf{1}) + (\mathbf{35}, \mathbf{1}, \mathbf{1}) \\ & + (\mathbf{8}_v, \mathbf{120}, \mathbf{1}) + (\mathbf{8}_v, \mathbf{1}, \mathbf{120}) , \\ (\mathrm{R}+, \mathrm{NS}-, \mathrm{NS}-) : \ & (\mathbf{8}, \mathbf{16}, \mathbf{16}) , \\ (\mathrm{R}-, \mathrm{R}-, \mathrm{NS}+) : \ & (\mathbf{8}', \mathbf{128}', \mathbf{1}) , \\ (\mathrm{R}-, \mathrm{NS}+, \mathrm{R}-) : \ & (\mathbf{8}', \mathbf{1}, \mathbf{128}') . \end{aligned}$$

This shows that a tachyon-free theory without supersymmetry is possible.

11.4 A little Lie algebra

In the open string the gauge charges are carried by the Chan–Paton degrees of freedom at the endpoints. In the closed string the charges are carried by fields that move along the string. We saw this earlier for the Kaluza–Klein gauge symmetry and the enhanced gauge symmetries that appear when the bosonic string is compactified, and now we see it again in the heterotic string. In the following sections we will discuss these closed string gauge symmetries in a somewhat more systematic way, but first we need to introduce a few ideas from Lie algebra. Space forbids a complete treatment; we focus on some basic ideas and some specific results that will be needed later.

Basic definitions

A Lie algebra is a vector space with an antisymmetric product $[T, T']$. In terms of a basis T^a the product is

$$[T^a, T^b] = i f^{ab}{}_c T^c \tag{11.4.1}$$

with $f^{ab}{}_c$ the *structure constants.* The product is required to satisfy the Jacobi identity

$$[T^a, [T^b, T^c]] + [T^b, [T^c, T^a]] + [T^c, [T^a, T^b]] = 0 . \tag{11.4.2}$$

The associated Lie group is generated by the exponentials

$$\exp(i\theta_a T^a) , \tag{11.4.3}$$

with the θ_a real. For a compact group, the associated compact Lie algebra has a positive inner product

$$(T^a, T^b) = d^{ab} , \tag{11.4.4}$$

which is invariant,

$$([T,T'],T'')+(T',[T,T''])=0\,. \tag{11.4.5}$$

This invariance is equivalent to the statement that f^{abc} is completely antisymmetric, where d^{ab} is used to raise the index.

We are interested in simple Lie algebras, those having no nontrivial closed subalgebras. A general compact algebra is a sum of simple algebras and $U(1)$s. For a simple algebra the inner product is unique up to normalization, and there is a basis of generators in which it is simply δ^{ab}. For any representation r of the Lie algebra (any set of matrices $t^a_{r,ij}$ satisfying (11.4.1) with the given $f^{ab}{}_c$), the trace is invariant and so for a simple Lie algebra is proportional to d^{ab},

$$\mathrm{Tr}(t^a_r t^b_r)=T_r d^{ab} \tag{11.4.6}$$

from some constant T_r. Also, $t^a_r t^b_r d_{ab}$ commutes with all the t^c_r and so is proportional to the identity,

$$t^a_r t^b_r d_{ab}=Q_r \tag{11.4.7}$$

with Q_r the *Casimir invariant* of the representation r.

The classical Lie algebras are

- $SU(n)$: Traceless Hermitean $n\times n$ matrices. The corresponding group consists of unitary matrices with unit determinant.[1] This algebra is also denoted A_{n-1}.

- $SO(n)$: Antisymmetric Hermitean $n\times n$ matrices. The corresponding group $SO(n,\mathbf{R})$ consists of real orthogonal matrices with unit determinant. For $n=2k$ this algebra is also denoted D_k. For $n=2k+1$ it is denoted B_k.

- $Sp(k)$: Hermitean $2k\times 2k$ matrices with the additional condition

$$MTM^{-1}=-T^T\,. \tag{11.4.8}$$

  Here the superscript T denotes the transpose, and

$$M=i\begin{bmatrix}0 & I_k\\ -I_k & 0\end{bmatrix} \tag{11.4.9}$$

  with I_k the $k\times k$ identity matrix. The corresponding group consists of unitary matrices U with the additional property

$$MUM^{-1}=(U^T)^{-1}\,. \tag{11.4.10}$$

[1] To be precise the Lie algebra determines only the local structure of the group. Many groups, differing only by discrete identifications, will have a common Lie algebra.

Confusingly, the notation $Sp(2k)$ is also used for this group. It is also denoted C_k.

From each of the compact groups one obtains various noncompact groups by multiplying some generators by i. For example, the traceless *imaginary* matrices generate the group $SL(n,\mathbf{R})$ of real matrices of unit determinant. The group $SO(m,n,\mathbf{R})$ preserving a Lorentzian inner product is similarly obtained from $SO(m+n)$. Another noncompact group is generated by imaginary rather than Hermitean matrices satisfying the symplectic condition (11.4.8) and consists of real matrices satisfying (11.4.10). This noncompact group is also denoted $Sp(k)$ or $Sp(2k)$; occasionally $USp(2k)$ is used to distinguish the compact unitary case.

Such noncompact groups do not appear in Yang–Mills theory (the result would not be unitary) but they have other applications. Some of the $SL(n,\mathbf{R})$ and $SO(m,n,\mathbf{R})$ appear as low energy symmetries in compactified string theory, as discussed in section B.5 and chapter 14. The real form of $Sp(k)$ is an invariance of the Poisson bracket in classical mechanics.

Roots and weights

A useful description of any Lie algebra h begins with a maximal set of commuting generators H^i, $i=1,\ldots,\mathrm{rank}(g)$. The remaining generators E^α can be taken to have definite charge under the H^i,

$$[H^i,E^\alpha]=\alpha^i E^\alpha\,. \tag{11.4.11}$$

The rank(g)-dimensional vectors α^i are known as *roots*. It is a theorem that there is only one generator for a given root so the notation E^α is unambiguous. The Jacobi identity determines the rest of the algebra to be

$$[E^\alpha,E^\beta]=\begin{cases}\epsilon(\alpha,\beta)E^{\alpha+\beta} & \text{if } \alpha+\beta \text{ is a root}\,,\\ 2\alpha\cdot H/\alpha^2 & \text{if } \alpha+\beta=0\,,\\ 0 & \text{otherwise}\,.\end{cases} \tag{11.4.12}$$

The products $\alpha\cdot H$ and α^2 are defined by contraction with d_{ij}, the inverse of the inner product (11.4.4) restricted to the commuting subalgebra. Taking the trace with H^i, the second equation determines the normalization $(E^\alpha,E^{-\alpha})=2/\alpha^2$. The constants $\epsilon(\alpha,\beta)$ take only the values ± 1.

The matrices t^i_r that represent H^i can all be taken to be diagonal. Their simultaneous eigenvalues w^i, combined into vectors

$$(w^1,\ldots,w^{\mathrm{rank}(g)})\,, \tag{11.4.13}$$

are the weights, equal in number to the dimension of the representation. The roots are the same as the weights of the adjoint representation.

Examples:

- For $SO(2k) = D_k$, consider the k 2×2 blocks down the diagonal and let H^i be

$$\begin{bmatrix} 0 & i \\ -i & 0 \end{bmatrix} \tag{11.4.14}$$

in the ith block and zero elsewhere. This is a maximal commuting set. The $2k$-vectors $(1, \mp i, 0, \ldots, 0)$ have eigenvalues

$$(\pm 1, 0^{k-1}) \tag{11.4.15}$$

under the k H^i; these are weights of the vector representation. The other weights are the same with the ± 1 in any other position.

The adjoint representation is the antisymmetric tensor, which is contained in the product of two vector representations. The weights are additive so the roots are obtained by adding any distinct (because of the antisymmetry) pair of vector weights. This gives

$$(+1, +1, 0^{k-2}),\ (+1, -1, 0^{k-2}),\ (-1, -1, 0^{k-2}), \tag{11.4.16}$$

and all permutations of these. The k zero roots obtained by adding any weight and its negative are just the H^i.

The diagonal generators (11.4.14) are the same as are used in section B.1 to construct the spinor representations. In the spinor representation the weights w^i are given by all k-vectors with components $\pm\frac{1}{2}$, with the $\mathbf{2^{k-1}}$ having an even number of $-\frac{1}{2}$s and the $\mathbf{2^{k-1}}'$ an odd number.

- For $SO(2k+1) = B_k$, one can take the same set of diagonal generators with a final row and a final column of zeros. The weights in the vector representation are the same as above plus (0^k) from the added row. The additional generators have roots

$$(\pm 1, 0^{k-1}) \tag{11.4.17}$$

and all permutations.

- The adjoint of $Sp(k) = C_k$ is the *symmetric* tensor, so one can obtain the roots as for $SO(2k)$ except that the vector weights need not be distinct. The resulting roots are those of $SO(2k)$ together with

$$(\pm 2, 0^{k-1}) \tag{11.4.18}$$

and permutations. It is usually conventional to normalize the generators such that the longest root has length-squared two, so we must divide all the roots by $2^{1/2}$.

- For $SU(n) = A_{n-1}$ it is useful first to consider $U(n)$, even though this algebra is not simple. The n commuting generators H^i can be taken to have a 1 in the ii position and zeros elsewhere. The charged generator with a 1 in the ij position then has eigenvalue $+1$ under H^i and -1 under H^j: the roots are all permutations of

$$(+1, -1, 0^{n-2}). \tag{11.4.19}$$

Note that all roots lie in the hyperplane $\sum_i \alpha^i = 0$; this is because all eigenvalues of the $U(1)$ generator are zero. The roots of $SU(n)$ are just the roots of $U(n)$ regarded as lying in this hyperplane.

- We have stated that the E_8 generators decompose into the adjoint plus one spinor of $SO(16)$. The commuting generators of $SO(16)$ can also be taken as commuting generators of E_8, so the roots of E_8 are the roots of $SO(16)$ plus the weights of the spinor, namely all permutations of the roots (11.4.16) plus

$$(+\tfrac{1}{2}, +\tfrac{1}{2}, +\tfrac{1}{2}, +\tfrac{1}{2}, +\tfrac{1}{2}, +\tfrac{1}{2}, +\tfrac{1}{2}, +\tfrac{1}{2}) \tag{11.4.20}$$

and the roots obtained from this by an even number of sign flips. Equivalently this set is described by

$$\alpha^i \in \mathbf{Z} \text{ for all } i\ , \text{ or } \alpha^i \in \mathbf{Z} + \tfrac{1}{2} \text{ for all } i\ , \tag{11.4.21a}$$

and

$$\sum_{i=1}^{8} \alpha^i \in 2\mathbf{Z}\ , \qquad \sum_{i=1}^{8} (\alpha^i)^2 = 2\ . \tag{11.4.21b}$$

For A_n, D_k, and E_8 (and E_6 and E_7, which we have not yet described), all roots are of the same length. These are referred to as *simply-laced* algebras. For B_k and C_k (and F_4 and G_2) there are roots of two different lengths so one refers to long and short roots.

A quantity that will be useful later is the *dual Coxeter number* $h(g)$ of the Lie algebra g, defined by

$$-\sum_{c,d} f^{ac}{}_d f^{bd}{}_c = h(g)\psi^2 d^{ab}\ . \tag{11.4.22}$$

Here ψ is any long root. For reference, we give the values for all simple Lie algebras in table 11.3. The definition (11.4.22) makes $h(g)$ independent of the arbitrary normalization of the inner product d^{ab} because the inverse appears in $\psi^2 = \psi^i \psi^j d_{ij}$.

Useful facts for grand unification

The exceptional group E_8 is connected to the groups appearing in grand unification through a series of subgroups. This will play a role in the com-

Table 11.3. *Dimensions and Coxeter numbers for simple Lie algebras.*

	$SU(n)$	$SO(n)$, $n \geq 4$	$Sp(k)$	E_6	E_7	E_8	F_4	G_2
$\dim(g)$	n^2-1	$n(n-1)/2$	$2k^2+k$	78	133	248	52	14
$h(g)$	n	$n-2$	$k+1$	12	18	30	9	4

pactification of the heterotic string, and so we record without derivation the necessary results.

The first subgroup is

$$E_8 \to SU(3) \times E_6 \,. \tag{11.4.23}$$

We have not described E_6 explicitly, but the reader can reproduce this and the decomposition (11.4.24) from the known properties of spinor representations, as well as the further decomposition of the E_6 representations in table 11.4 (exercise 11.5). In simple compactifications of the $E_8 \times E_8$ string, the fermions of the Standard Model can all be thought of as arising from the **248**-dimensional adjoint representation of one of the E_8s. It is therefore interesting to trace the fate of this representation under the successive symmetry breakings. Under $E_8 \to SU(3) \times E_6$,

$$\mathbf{248} \to (\mathbf{8},\mathbf{1}) + (\mathbf{1},\mathbf{78}) + (\mathbf{3},\mathbf{27}) + (\bar{\mathbf{3}},\overline{\mathbf{27}}) \,. \tag{11.4.24}$$

That is, the adjoint of E_8 contains the adjoints of the subgroups, with half the remaining 162 generators transforming as a triplet of $SU(3)$ and a complex **27**-dimensional representation of E_6 and half as the conjugate of this. Further subgroups are shown in table 11.4. The first three subgroups correspond to successive breaking of E_6 down to the Standard Model group through smaller grand unified groups; the fourth is an alternate breaking pattern.

It is a familiar fact from grand unification that precisely one $SU(3) \times SU(2) \times U(1)$ generation of quarks and leptons is contained in the **10** plus $\bar{\mathbf{5}}$ of $SU(5)$. Tracing back further, we see that a generation fits into the single representation **16** of $SO(10)$, together with an additional state $\mathbf{1}_{-5}$. This extra state is neutral under $SU(5)$, and so under $SU(3) \times SU(2) \times U(1)$, and can be regarded as a right-handed neutrino. Going back to E_6, the **27** contains the 15 states of a single generation plus 12 additional states. Relative to $SU(5)$ unification, $SO(10)$ and E_6 are more unified in the sense that a generation is contained within a single representation, but less economical in that the representation contains additional unseen states as well. In fact, the latter may not be such a

Table 11.4. *Subgroups and representations of grand unified groups.*

E_6	$\to$	$SO(10) \times U(1)$
78	$\to$	$\mathbf{45}_0 + \mathbf{16}_{-3} + \overline{\mathbf{16}}_3 + \mathbf{1}_0$
27	$\to$	$\mathbf{1}_4 + \mathbf{10}_{-2} + \mathbf{16}_1$
$SO(10)$	$\to$	$SU(5) \times U(1)$
45	$\to$	$\mathbf{24}_0 + \mathbf{10}_4 + \overline{\mathbf{10}}_{-4} + \mathbf{1}_0$
16	$\to$	$\mathbf{10}_{-1} + \bar{\mathbf{5}}_3 + \mathbf{1}_{-5}$
10	$\to$	$\mathbf{5}_2 + \bar{\mathbf{5}}_{-2}$
$SU(5)$	$\to$	$SU(3) \times SU(2) \times U(1)$
10	$\to$	$(\mathbf{3},\mathbf{2})_1 + (\bar{\mathbf{3}},\mathbf{1})_{-4} + (\mathbf{1},\mathbf{1})_6$
$\bar{\mathbf{5}}$	$\to$	$(\bar{\mathbf{3}},\mathbf{1})_2 + (\mathbf{1},\mathbf{2})_{-3}$
E_6	$\to$	$SU(3) \times SU(3) \times SU(3)$
78	$\to$	$(\mathbf{8},\mathbf{1},\mathbf{1}) + (\mathbf{1},\mathbf{8},\mathbf{1}) + (\mathbf{1},\mathbf{1},\mathbf{8}) + (\mathbf{3},\mathbf{3},\mathbf{3}) + (\bar{\mathbf{3}},\bar{\mathbf{3}},\bar{\mathbf{3}})$
27	$\to$	$(\mathbf{3},\bar{\mathbf{3}},\mathbf{1}) + (\mathbf{1},\mathbf{3},\bar{\mathbf{3}}) + (\bar{\mathbf{3}},\mathbf{1},\mathbf{3})$

difficulty. To see why, consider the decomposition of the **27** of E_6 under $SU(3) \times SU(2) \times U(1) \subset SU(5) \subset SO(10) \subset E_6$:

$$\begin{aligned}\mathbf{27} \to\ & (\mathbf{3},\mathbf{2})_1 + (\bar{\mathbf{3}},\mathbf{1})_{-4} + (\mathbf{1},\mathbf{1})_6 + (\bar{\mathbf{3}},\mathbf{1})_2 + (\mathbf{1},\mathbf{2})_{-3} \\ & + [\mathbf{1}_0] \\ & + [(\bar{\mathbf{3}},\mathbf{1})_2 + (\mathbf{3},\mathbf{1})_{-2}] + [(\mathbf{1},\mathbf{2})_{-3} + (\mathbf{1},\mathbf{2})_3] + [\mathbf{1}_0] \,.\end{aligned} \tag{11.4.25}$$

The first line lists one generation, the second the extra state appearing in the **16** of $SO(10)$, and the third the additional states in the **27** of E_6. The subset within each pair of square brackets is a real representation of $SU(3) \times SU(2) \times U(1)$. The significance of this is that for a real representation r, the CPT conjugate also is in the representation r, and so the combined gauge plus $SO(2)$ helicity representation for the particles plus their antiparticles is $(r,+\frac{1}{2}) + (r,-\frac{1}{2})$. This is the same as for a *massive* spin-$\frac{1}{2}$ particle in representation r, so it is consistent with the gauge and spacetime symmetries for these particles to be massive. In the most general invariant action, all particles in [ ] brackets will have large (of order the unification scale) masses. It is notable that for any of the $\mathbf{10} + \bar{\mathbf{5}}$ of $SU(5)$, the **16** of $SO(10)$, or the **27** of E_6, the natural $SU(3) \times SU(2) \times U(1)$ spectrum is precisely a standard generation of quarks and leptons.

11.5 Current algebras

The gauge boson vertex operators in the heterotic string are of the form $j(z)\bar{\psi}^{\mu}(\bar{z})e^{ik\cdot X}$, where $j(z)$ is either a fermion bilinear $\lambda^A\lambda^B$ or a spin field (11.2.26). Similarly the gauge boson vertex operators for the toroidally compactified bosonic string were of the form $j(z)\bar{\partial}X^{\mu}(\bar{z})e^{ik\cdot X}$ with j being ∂X^m for the Kaluza–Klein gauge bosons or an exponential for the enhanced gauge symmetry (or the same with right and left reversed). All these currents are holomorphic $(1,0)$ operators. In this section we consider general properties of such currents.

Let us consider in a general CFT the set of $(1,0)$ currents $j^a(z)$. Conformal invariance requires their OPE to be of the form

$$j^a(z)j^b(0) \sim \frac{k^{ab}}{z^2} + \frac{ic^{ab}{}_c}{z}j^c(0) \tag{11.5.1}$$

with k^{ab} and $c^{ab}{}_c$ constants. Dimensionally, the z^{-2} term must be a c-number and the z^{-1} term must be proportional to a current. The Laurent coefficients

$$j^a(z) = \sum_{m=-\infty}^{\infty} \frac{j_m^a}{z^{m+1}} \tag{11.5.2}$$

thus satisfy a closed algebra

$$[j_m^a, j_n^b] = mk^{ab}\delta_{m,-n} + ic^{ab}{}_c j_{m+n}^c \,. \tag{11.5.3}$$

In particular,

$$[j_0^a, j_0^b] = ic^{ab}{}_c j_0^c \,. \tag{11.5.4}$$

That is, the $m = 0$ modes form a Lie algebra g, and

$$c^{ab}{}_c = f^{ab}{}_c \,. \tag{11.5.5}$$

We focus first on the case of simple g. The $j_1^a\, j_0^b\, j_{-1}^c$ Jacobi identity requires that

$$f^{bc}{}_d k^{ad} + f^{ba}{}_d k^{dc} = 0 \,. \tag{11.5.6}$$

This is the same relation as that defining the Lie algebra inner product d^{ab}, and since we are assuming g to be simple it must be that

$$k^{ab} = \hat{k}d^{ab} \tag{11.5.7}$$

for some constant $\hat{k}$. The algebra (11.5.3) is variously known as a *current algebra*, an *affine Lie algebra*, or an *(affine) Kac–Moody algebra*. The currents are $(1,0)$ tensors, so

$$[L_m, j_n^a] = nj_{m+n}^a \,. \tag{11.5.8}$$

Physically, the j_n^a generate position-dependent g-transformations. This is possible in quantum field theory because there is a local current. The *central extension* or *Schwinger term* $\hat{k}$ must always be positive in a unitary theory. To show this, note that

$$\hat{k}d^{aa} = \langle 1|\,[j_1^a, j_{-1}^a]\,|1\rangle = \|\, j_{-1}^a|1\rangle\|^2 \tag{11.5.9}$$

(no sum on a). For a compact Lie algebra d^{aa} is positive and so $\hat{k}$ must be nonnegative. It can vanish only if $j_{-1}^a|1\rangle = 0$, but the vertex operator for $j_{-1}^a|1\rangle$ is precisely the current j^a: any matrix element of j^a can be obtained by gluing $j_{-1}^a|1\rangle$ into the world-sheet. Thus $\hat{k} = 0$ only if the current vanishes identically.

The coefficient $\hat{k}$ is quantized. To show this, consider any root α. Defining

$$J^3 = \frac{\alpha\cdot H}{\alpha^2} \,, \quad J^{\pm} = E^{\pm\alpha} \,, \tag{11.5.10}$$

one finds from the general form (11.4.12) that these satisfy the $SU(2)$ algebra

$$[J^3, J^{\pm}] = J^{\pm} \,, \quad [J^+, J^-] = 2J^3 \,. \tag{11.5.11}$$

The reader can verify that the two sets

$$\frac{\alpha\cdot H_0}{\alpha^2} \,, \quad E_0^{\alpha} \,, \quad E_0^{-\alpha} \,, \tag{11.5.12a}$$

$$\frac{\alpha\cdot H_0 + \hat{k}}{\alpha^2} \,, \quad E_1^{\alpha} \,, \quad E_{-1}^{-\alpha} \tag{11.5.12b}$$

also satisfy the $SU(2)$ algebra. The first is just the usual center-of-mass Lie algebra, while the second is known as *pseudospin*. From familiar properties of $SU(2)$, $2J_3$ must be an integer, and so $2\hat{k}/\alpha^2$ must be an integer. This condition is most stringent if α is taken to be one of the long roots of the algebra (denoted ψ). The *level*

$$k = \frac{2\hat{k}}{\psi^2} \tag{11.5.13}$$

is then a nonnegative integer, and positive for a nontrivial current.

It is common to normalize the Lie algebra inner product to give the long roots length-squared two, so that $\hat{k} = k$ is the coefficient of the leading term in the OPE. We will usually do this in examples, as we have done in giving the roots of various Lie algebras in the previous section. Incidentally, it follows that with this normalization the generators (11.4.14) are normalized, so the $SO(n)$ inner product is half of the vector representation trace. Similarly the inner product for $SU(n)$ such that the long roots have length-squared two is equal to the trace in the fundamental representation. In general expressions we will keep the inner product arbitrary, inserting explicit factors of ψ^2 so that results are independent of the normalization.

We will, however, take henceforth a basis for the generators such that $d^{ab} = \delta^{ab}$.

The level represents the relative magnitude of the z^{-2} and z^{-1} terms in the OPE. For $U(1)$ the structure constant is zero and only the z^{-2} term appears. Hence there is no analog of the level. It is convenient to normalize all the $U(1)$ currents to

$$j^a(z)j^b(0) \sim \frac{\delta^{ab}}{z^2} . \tag{11.5.14}$$

From this OPE and holomorphicity it follows that each $U(1)$ current algebra is isomorphic to a free boson CFT,

$$j^a = i\partial H^a . \tag{11.5.15}$$

We will often use this equivalence.

The current algebra in the heterotic string consisted of n real fermions $\lambda^A(z)$. The currents

$$i\lambda^A\lambda^B \tag{11.5.16}$$

form an $SO(n)$ algebra. The maximal set of commuting currents is $i\lambda^{2K-1}\lambda^{2K}$ for $K = 1,\ldots,[n/2]$. These correspond to the generators (11.4.14), which are normalized such that roots (11.4.16) have length-squared two. The level is then the coefficient of the leading term in the OPE; this is $1/z^2$, so the level is $k = 1$. The case $n = 3$ is an exception: there are no long roots, only the short roots ± 1, so we must rescale the diagonal current to $2^{1/2}i\lambda^1\lambda^2$ and the level is $k = 2$.

For any real representation r of any Lie algebra, one can construct from $\dim(r)$ real fermions the currents

$$\frac{i}{2}\lambda^A\lambda^B t^a_{r,AB} . \tag{11.5.17}$$

These satisfy a current algebra with level $k = T_r/\psi^2$, with T_r defined in eq. (11.4.6). The case in the previous paragraph is the n-dimensional vector representation of $SO(n)$, for which $T_R = \psi^2$. As another example, nk fermions transforming as k copies of the vector representation give level k.

As a final example consider the $SU(2)$ symmetry at the self-dual point of toroidal compactification. The current is $\exp[2^{1/2}iH(z)]$. The current $i\partial H$ is then normalized so that the weight (from the OPE) is $2^{1/2}$, with length-squared two. The OPE of $i\partial H$ with itself starts as $1/z^2$, so the level is again $k = 1$.

In some cases one may have sectors in which some currents are not periodic, $j^a(w + 2\pi) = R^{ab}j^b(w)$, where R^{ab} is any automorphism of the algebra. In these, the modes of the currents are fractional and satisfy a *twisted* affine Lie algebra.

The Sugawara construction

In current algebras with conformal symmetry, there is a remarkable connection between the energy-momentum tensor and the currents, which leads to a great deal of interesting structure. Define the operator

$$:jj(z_1): = \lim_{z_2\to z_1}\left(j^a(z_1)j^a(z_2) - \frac{\hat{k}\dim(g)}{z_{12}^2}\right) , \tag{11.5.18}$$

with the sum on a implicit. We first wish to show that up to normalization the OPE of $:jj:$ with j^a is the same as that of T_B with j^a. This takes a bit of effort; the same calculation is organized in a different way in exercise 11.7.

The OPE of the product $:jj:$ is not the same as the product of the OPEs, because the two currents in $:jj:$ are closer to each other than they are to the third current; we must make a less direct argument using holomorphicity. Consider the following product:

$$\begin{aligned} j^a(z_1)j^a(z_2)j^c(z_3) = {} & \frac{\hat{k}}{z_{31}^2}j^c(z_2) + \frac{if^{cad}}{z_{31}}j^d(z_1)j^a(z_2) + \frac{\hat{k}}{z_{32}^2}j^c(z_1) \\ & + \frac{if^{cad}}{z_{32}}j^a(z_1)j^d(z_2) + \text{terms holomorphic in } z_3 . \end{aligned} \tag{11.5.19}$$

We have used the current–current OPE to determine the singularities as z_3 approaches z_1 or z_2, with a holomorphic remainder. In this relation take $z_2 \to z_1$ and make a Laurent expansion in z_{21}, being careful to expand both the operator products and the explicit z_2 dependence. Keep the term of order z_{21}^0 (there is some cancellation from the antisymmetry of f^{cad}) to obtain

$$\begin{aligned} :jj(z_1): j^c(z_3) &\sim \frac{2\hat{k}}{z_{13}^2}j^c(z_1) + \frac{f^{cad}f^{ead}}{z_{13}^2}j^e(z_1) \\ &= \frac{2\hat{k} + h(g)\psi^2}{z_{13}^2}j^c(z_1) \\ &= (k + h(g))\psi^2\left[\frac{1}{z_{13}^2}j^c(z_3) + \frac{1}{z_{13}}\partial j^c(z_3)\right] . \end{aligned} \tag{11.5.20}$$

Here $h(g)$ is again the dual Coxeter number. Define

$$T_B^s(z) = \frac{1}{(k + h(g))\psi^2} :jj(z): . \tag{11.5.21}$$

The OPE of T_B^s with the current is the same as that of the energy-momentum tensor $T_B(z)$,

$$T_B^s(z)j^c(0) \sim T_B(z)j^c(0) . \tag{11.5.22}$$

Now repeat the above with $j^c(z_3)$ replaced by $T_B^s(z_3)$,

$$j^a(z_1)j^a(z_2)T_B^s(z_3) = \frac{1}{z_{31}^2}j^a(z_1)j^a(z_2) + \frac{1}{z_{31}}\partial j^a(z_1)j^a(z_2) + \frac{1}{z_{32}^2}j^a(z_1)j^a(z_2) + \frac{1}{z_{32}}j^a(z_1)\partial j^a(z_2) + \text{terms holomorphic in } z_3 . \tag{11.5.23}$$

Again expand in z_{21} and keep the term of order z_{21}^0 to obtain

$$T_B^s(z_1)T_B^s(z_3) \sim \frac{c^{g,k}}{2z_{13}^4} + \frac{2}{z_{13}^2}T_B^s(z_3) + \frac{1}{z_{13}}\partial T_B^s(z_3) \tag{11.5.24}$$

with

$$c^{g,k} = \frac{k\dim(g)}{k+h(g)} . \tag{11.5.25}$$

This is of the standard form for an energy-momentum tensor, with central charge $c^{g,k}$. The Laurent coefficients

$$L_0^s = \frac{1}{(k+h(g))\psi^2}\left(j_0^a j_0^a + 2\sum_{n=1}^{\infty} j_{-n}^a j_n^a\right) , \tag{11.5.26a}$$

$$L_m^s = \frac{1}{(k+h(g))\psi^2}\sum_{n=-\infty}^{\infty} j_n^a j_{m-n}^a , \quad m \neq 0 , \tag{11.5.26b}$$

satisfy a Virasoro algebra with this central charge. The vanishing of the normal ordering constant in L_0^s can be deduced by noting that holomorphicity requires L_0^s and also j_n^a for $n \geq 0$ to annihilate the state $|1\rangle$.

We have used the jj OPE to determine the $:jj::jj:$ OPE. We could not do this directly, because the jj OPE is valid only for two operators close compared to all others, and in this case there are two additional currents in the vicinity. Naive application of the OPE would give the wrong normalization for T^s and $c^{g,k}$. The argument above uses the OPE only where it is valid, and then takes advantage of holomorphicity. The operator T_B^s constructed from the product of two currents is known as a *Sugawara energy-momentum tensor.*

Finding the Sugawara tensor for a $U(1)$ current algebra is easy. With the normalization (11.5.14) it is simply

$$T_B^s = \frac{1}{2} :jj: , \tag{11.5.27}$$

as one sees by writing the current in terms of a free boson, $j = i\partial H$.

The tensor T_B^s may or may not be equal to the total T_B of the CFT. Define

$$T_B' = T_B - T_B^s . \tag{11.5.28}$$

Since the T_B and T_B^s OPEs with j^a have the same singular terms, the product

$$T_B'(z_1)j^a(z_2) \sim 0 \tag{11.5.29}$$

is nonsingular. Since T_B^s itself is constructed from the currents, this implies $T_B^s T_B' \sim 0$. Then

$$T_B'(z)T_B'(0) = T_B(z)T_B(0) - T_B^s(z)T_B^s(0) - T_B'(z)T_B^s(0) - T_B^s(z)T_B'(0) \sim \frac{c'}{2z^4} + \frac{2}{z^2}T_B'(0) + \frac{1}{z}T_B'(0) , \tag{11.5.30}$$

the standard TT OPE with central charge

$$c' = c - c^{g,k} . \tag{11.5.31}$$

The internal theory thus separates into two decoupled CFTs. One has an energy-momentum tensor T_B^s constructed entirely from the current, and the other an energy-momentum tensor T_B' that commutes with the current. We will use the term *current algebra* to refer to the first factor alone, since the two CFTs are completely independent. For a unitary CFT c' must be nonnegative and so

$$c^{g,k} \leq c , \tag{11.5.32}$$

and T_B' is trivial precisely if

$$c^{g,k} = c , \tag{11.5.33}$$

in which case $T_B = T_B^s$.

We now consider examples. The dual Coxeter number can be written as a sum over the roots. For any simply-laced algebra, $h(g) + 1 = \dim(g)/\text{rank}(g)$, and so

$$c^{g,k} = \frac{k\dim(g)\,\text{rank}(g)}{\dim(g) + (k-1)\text{rank}(g)} . \tag{11.5.34}$$

For any simply-laced algebra at $k = 1$, the central charge is therefore

$$c^{g,1} = \text{rank}(g) . \tag{11.5.35}$$

For the $E_8 \times E_8$ and $SO(32)$ heterotic strings, this is the same as the central charge of the free fermion representation, and for the free boson representation of the next section: these are Sugawara theories. The operator $:jj:$ looks as though it should be quartic in the fermions, but by using the OPE and the antisymmetry of the fermions one finds that T_B^s reduces to the usual $-\frac{1}{2}\lambda^A\partial\lambda^A$.

Another example is $SU(2) = SO(3)$, for which

$$c^{g,k} = \frac{3k}{2+k} = 1, \frac{3}{2}, \frac{9}{5}, 2, \frac{15}{7}, \ldots \to 3 . \tag{11.5.36}$$

We have seen the first CFT in this series (the self-dual point of toroidal compactification) and the second (free fermions). Most levels do not have a free-field representation. For any current algebra the central charge lies in the range

$$\text{rank}(g) \leq c^{g,k} \leq \dim(g) \,. \tag{11.5.37}$$

The first equality holds only for a simply-laced algebra at level one, and the second only for an Abelian algebra or in the limit $k \to \infty$.

Primary fields

By acting repeatedly with the lowering operators j_n^a with $n > 0$, one reaches a *highest weight* or *primary* state of the current algebra, a state annihilated by all the j_n^a for $n > 0$. It is therefore also annihilated by the L_n^s for $n > 0$, eq. (11.5.26), so is a highest weight state of the Virasoro algebra. The center-of-mass generators j_0^a take primary states into primary states, so the latter form a representation of the algebra g,

$$j_0^a |r, i\rangle = |r, j\rangle t^a_{r,ji} \,, \tag{11.5.38}$$

with r (not summed) labeling the representation. It then follows that

$$\begin{aligned} L_0^s |r, i\rangle &= \frac{1}{(k + h(g))\psi^2} |r, k\rangle t^a_{r,kj} t^a_{r,ji} \\ &= \frac{Q_r}{(k + h(g))\psi^2} |r, i\rangle \,, \end{aligned} \tag{11.5.39}$$

with Q_r the Casimir (11.4.7). The weights of the primary fields are thus determined in terms of the algebra, level, and representation,

$$h_r = \frac{Q_r}{(k + h(g))\psi^2} = \frac{Q_r}{2\hat{k} + Q_g} \,, \tag{11.5.40}$$

where Q_g is the Casimir for the adjoint representation. For $SU(2)$ at level k, the weight of the spin-j primary is

$$h_j = \frac{j(j+1)}{k+2} \,. \tag{11.5.41}$$

It is also true that at any given level, only a finite number of representations are possible for the primary states. For any root α of g and any weight λ of r, the $SU(2)$ algebra (11.5.12b) implies that

$$\begin{aligned} \langle r, \lambda | \, [E_1^\alpha, E_{-1}^{-\alpha}] \, | r, \lambda \rangle &= 2 \langle r, \lambda | (\alpha \cdot H_0 + \hat{k}) | r, \lambda \rangle / \alpha^2 \\ &= 2(\alpha \cdot \lambda + \hat{k})/\alpha^2 \,. \end{aligned} \tag{11.5.42}$$

The left-hand side is $\|E_{-1}^{-\alpha} |r, \lambda\rangle\|^2 \geq 0$, and so $\hat{k} \geq -\alpha \cdot \lambda$. Combining this with the same for $-\alpha$ gives

$$\hat{k} \geq |\alpha \cdot \lambda| \tag{11.5.43}$$

for all weights λ of r. Taking α to be a long root ψ, the level must satisfy

$$k \geq \frac{2|\psi \cdot \lambda|}{\psi^2} = 2J^3 \,, \tag{11.5.44}$$

where J^3 refers to the $SU(2)$ algebra (11.5.12a) constructed from the charges j_0^a and the root ψ. For $g = SU(2)$ the statement is that the spin j of any primary state can be at most $\frac{1}{2}k$. For example at $k = 1$, only the representations **1** and **2** are possible. For $g = SU(3)$ at $k = 1$, only the **1**, **3**, and $\bar{\mathbf{3}}$ can appear. For $g = SU(n)$ at level k, only representations whose Young tableau has k or fewer columns can appear.

The expectation values of primary fields are completely determined by symmetry. We defer the details to chapter 15.

Finally, let us briefly discuss the gauge symmetries of the type I theory in this same abstract language. The matter part of the gauge boson vertex operator is

$$\dot{X}^\mu \lambda^a e^{ik \cdot X} \tag{11.5.45}$$

on the boundary, where the λ^a are weight 0 fields. In a unitary CFT such λ^a must be constant by the equations of motion. The OPE is then

$$\lambda^a(y_1)\lambda^b(y_2) = \left[\theta(y_1 - y_2) d^{ab}{}_c + \theta(y_2 - y_1) d^{ba}{}_c\right] \lambda^c(y_2) \,, \tag{11.5.46}$$

so the λ^a form a multiplicative algebra with structure constants $d^{ab}{}_c$. The antisymmetric part of $d^{ab}{}_c$ is the structure constant of the gauge Lie algebra. This is an abstract description of the Chan–Paton factor. The requirement that the λ^a algebra be associative has been shown to forbid the gauge group $E_8 \times E_8$.

11.6 The bosonic construction and toroidal compactification

We have seen in the construction of winding state vertex operators in section 8.2 that we may consider independent left- and right-moving scalars. Let us try to construct a heterotic string with 26 left-movers and 10 right-movers, which together with the ψ^μ give the correct central charge. The main issue is the spectrum of $k_{L,R}$; as in section 8.4 we use dimensionless momenta

$$l_{L,R} = (\alpha'/2)^{1/2} k_{L,R} \tag{11.6.1}$$

in much of the discussion. Recall that an ordinary noncompact dimension corresponds to a left- plus a right-mover with $l_L^\mu = l_R^\mu = l^\mu$ taking continuous values; let there be $d \leq 10$ noncompact dimensions. The remaining momenta,

$$(l_L^m, l_R^n) \,, \quad d \leq m \leq 25 \,, \; d \leq n \leq 9 \,, \tag{11.6.2}$$

take values in some lattice Γ. From the discussion of Narain compactification in section 8.4, we know that the requirements for a consistent CFT are locality of the OPE plus modular invariance. After taking the GSO projection on the right-movers, the conditions on Γ are precisely as in the bosonic case. Defining the product

$$l \circ l' = l_L \cdot l'_L - l_R \cdot l'_R \,, \tag{11.6.3}$$

the lattice must be an *even self-dual Lorentzian lattice of signature* $(26-d, 10-d)$,

$$l \circ l \in 2\mathbf{Z} \quad \text{for all } l \in \Gamma \,, \tag{11.6.4a}$$

$$\Gamma = \Gamma^* \,. \tag{11.6.4b}$$

As in the bosonic case, where the signature was $(26-d, 26-d)$, all such lattices have been classified. Consider first the maximum possible number of noncompact dimensions, $d = 10$. In this case, the $\circ$ product has only positive signs, so the l_L^m form an even self-dual *Euclidean* lattice of dimension 16. Even self-dual Euclidean lattices exist only when the dimension is a multiple of 8, and for dimension 16 there are exactly two such lattices, Γ_{16} and $\Gamma_8 \times \Gamma_8$. The lattice Γ_{16} is the set of all points of the form

$$(n_1, \ldots, n_{16}) \quad \text{or} \quad (n_1 + \tfrac{1}{2}, \ldots, n_{16} + \tfrac{1}{2}) \,, \tag{11.6.5a}$$

$$\sum_i n_i \in 2\mathbf{Z} \tag{11.6.5b}$$

for any integers n_i. The lattice Γ_8 is similarly defined to be all points

$$(n_1, \ldots, n_8) \quad \text{or} \quad (n_1 + \tfrac{1}{2}, \ldots, n_8 + \tfrac{1}{2}) \,, \tag{11.6.6a}$$

$$\sum_i n_i \in 2\mathbf{Z} \,. \tag{11.6.6b}$$

The left-moving zero-point energy is -1 as in the bosonic string, so the massless states would have left-moving vertex operators ∂X^μ, ∂X^m, or $e^{ik_L \cdot X(z)}$ with $l_L^2 = 2$. Tensored with the usual right-moving $\mathbf{8}_v + \mathbf{8}$, the first gives the usual graviton, dilaton, and antisymmetric tensor. The 16 ∂X^m currents form a maximal commuting set corresponding to the m-momenta. The momenta l_L^m are the charges under these and so are the roots of the gauge group. For Γ_{16}, the points of length-squared two are just the $SO(32)$ roots (11.4.16). For Γ_8 the points of length-squared two are the E_8 roots (11.4.21). Thus the two possible lattices give the same two gauge groups, $SO(32)$ and $E_8 \times E_8$, found earlier. The commuting currents have singularity $1/z^2$, so $k = 1$ again.

It is easy to see that the earlier fermionic construction and the present bosonic one are equivalent under bosonization. The integral points on the lattices (11.6.5) and (11.6.6) map to the NS sectors of the current algebra and the half-integral points to the R sectors. The constraint that the total

k_L^m be even is the GSO projection on the left-movers in each theory. We have seen in the previous section that the dynamics of a current algebra is completely determined by its symmetry, so we can give a representation-independent description of the left-movers as an $SO(32)$ or $E_8 \times E_8$ level one current algebra.[2]

Let us note some general results about Lie algebras and lattices. The set of all integer linear combinations of the roots of a Lie algebra g is known as the *root lattice* Γ_g of g. Now take any representation r and let λ be any weight of r. The set of points $\lambda + v$ for all $v \in \Gamma_g$ is denoted Γ_r. It can be shown by considering various $SU(2)$ subgroups that for a simply-laced Lie algebra with roots of length-squared two,

$$\Gamma_r \subset \Gamma_g^* \,. \tag{11.6.7}$$

The union of all Γ_r is the weight lattice Γ_w, and[3]

$$\Gamma_w = \Gamma_g^* \,. \tag{11.6.8}$$

For example, the weight lattice of $SO(2n)$ has four sublattices:

$$(0): \ 0 + \text{any root} \,; \tag{11.6.9a}$$

$$(v): \ (1,0,0,\ldots,0) + \text{any root} \,; \tag{11.6.9b}$$

$$(s): \ (\tfrac{1}{2},\tfrac{1}{2},\tfrac{1}{2},\ldots,\tfrac{1}{2}) + \text{any root} \,; \tag{11.6.9c}$$

$$(c): \ (-\tfrac{1}{2},\tfrac{1}{2},\tfrac{1}{2},\ldots,\tfrac{1}{2}) + \text{any root} \,. \tag{11.6.9d}$$

These are respectively the root lattice, the lattice containing the weights of the vector representation, and the lattices containing the weights of the two 2^{n-1}-dimensional spinor representations. The lattice Γ_8 is the root lattice of E_8 and is also the weight lattice because it is self-dual. The root lattice of $SO(32)$ gives the integer points in Γ_{16}. The full Γ_{16} is the root lattice plus one spinor lattice of $SO(32)$.

The level one current algebra for any simply-laced Lie algebra g can similarly be represented by rank(g) left-moving bosons, the momentum lattice being the root lattice of g with the roots scaled to length-squared two. The constants $\epsilon(\alpha,\beta)$ appearing in the Lie algebra (11.4.12) can then be determined from the vertex operator OPE; this is one situation where the explicit form of the cocycle is needed. A modular-invariant CFT can be obtained by taking also rank(g) right-moving bosons, with the momentum lattice being

$$\Gamma = \sum_r \Gamma_r \times \tilde{\Gamma}_r \,. \tag{11.6.10}$$

[2] To be precise it is still necessary to specify the spectrum, which amounts to specifying which primary fields appear. Modular invariance generally restricts the possibilities greatly.

[3] For the nonsimply-laced algebras $Sp(k)$ and $SO(2k+1)$, these same relations hold between the weight lattice of one and the dual of the root lattice of the *other*.

That is, the spectrum runs over all sublattices of the weight lattice, with the left- and right-moving momenta taking values in the same sublattice.

Toroidal compactification

In parallel to the bosonic case, all even self-dual lattices of signature $(26-d, 10-d)$ can be obtained from any single lattice by $O(26-d, 10-d, \mathbf{R})$ transformations. Again start with any given solution Γ_0; for example, this could be either of the ten-dimensional theories with all compact dimensions orthogonal and at the $SU(2) \times SU(2)$ radius. Then any lattice

$$\Gamma = \Lambda\Gamma_0 , \quad \Lambda \in O(26-d, 10-d, \mathbf{R}) \tag{11.6.11}$$

defines a consistent heterotic string theory. As in the bosonic case there is an equivalence

$$\Lambda_1\Lambda\Lambda_2\Gamma_0 \cong \Lambda\Gamma_0 , \tag{11.6.12}$$

where

$$\Lambda_1 \in O(26-d, \mathbf{R}) \times O(10-d, \mathbf{R}) , \quad \Lambda_2 \in O(26-d, 10-d, \mathbf{Z}) . \tag{11.6.13}$$

The moduli space is then

$$\frac{O(26-d, 10-d, \mathbf{R})}{O(26-d, \mathbf{R}) \times O(10-d, \mathbf{R}) \times O(26-d, 10-d, \mathbf{Z})} . \tag{11.6.14}$$

The discrete T-duality group $O(26-d, 10-d, \mathbf{Z})$ of invariances of Γ_0 is understood to act on the right.

Now consider the unbroken gauge symmetry. There are $26-d$ gauge bosons with vertex operators $\partial X^m \tilde{\psi}^\mu$ and $10-d$ with vertex operators $\partial X^\mu \tilde{\psi}^m$. These are the original 16 commuting symmetries of the ten-dimensional theory plus $10-d$ Kaluza–Klein gauge bosons and $10-d$ more from compactification of the antisymmetric tensor. In addition there are gauge bosons $e^{ik_L \cdot X_L} \tilde{\psi}^\mu$ for every point on the lattice Γ such that

$$l_L^2 = 2 , \quad l_R = 0 . \tag{11.6.15}$$

There are no gauge bosons from points with $l_R \neq 0$ because the mass of such a state will be at least $\frac{1}{2} l_R^2$. For generic boosts Λ, giving generic points in the moduli space, there are no points in Γ with $l_R = 0$ and so no additional gauge bosons; the gauge group is $U(1)^{36-2d}$. At special points the gauge symmetry is enhanced. Obviously one can get $SO(32) \times U(1)^{20-2d}$ or $E_8 \times E_8 \times U(1)^{20-2d}$ from compactifying the original ten-dimensional theory on a torus without Wilson lines, just as in field theory. However, as in the bosonic string, there are stringy enhanced gauge symmetries at special points in moduli space. For example, the lattice $\Gamma_{26-d,10-d}$, defined by analogy to the lattices Γ_8 and Γ_{16}, gives rise to $SO(52-2d) \times U(1)^{10-d}$. As in the bosonic case, the low energy physics near the point of enhanced

symmetry is the Higgs mechanism. All groups obtained in this way have rank $36-2d$. This is the maximum in perturbation theory, but we will see in chapter 19 that nonperturbative effects can lead to larger gauge symmetries.

The number of moduli, from the dimensions of the SO groups, is

$$\frac{1}{2}\left[(36-2d)(35-2d) - (26-d)(25-d) - (10-d)(9-d)\right] = (26-d)(10-d) . \tag{11.6.16}$$

As in the bosonic string these can be interpreted in terms of backgrounds for the fields of the ten-dimensional gauge theory. The compact components of the metric and antisymmetric tensor give a total of $(10-d)^2$ moduli just as before. In addition there can be Wilson lines, constant backgrounds for the gauge fields A_m. As discussed in chapter 8, due to the potential $\mathrm{Tr}([A_m, A_n]^2)$ the fields in different directions commute along flat directions and so can be chosen to lie in a $U(1)^{16}$ subgroup. Thus there are $16(10-d)$ parameters in A_m for $(26-d)(10-d)$ in all.

In chapter 8 we studied quantization with antisymmetric tensor and open string Wilson line backgrounds. Here we leave the details to the exercises and quote the result. If we compactify $x^m \cong x^m + 2\pi R$ with constant backgrounds G_{mn}, B_{mn}, and A_m^I, then canonical quantization gives

$$k_{Lm} = \frac{n_m}{R} + \frac{w^n R}{\alpha'}(G_{mn} + B_{mn}) - q^I A_m^I - \frac{w^n R}{2} A_n^I A_m^I , \tag{11.6.17a}$$

$$k_L^I = (q^I + w^m R A_m^I)(2/\alpha')^{1/2} , \tag{11.6.17b}$$

$$k_{Rm} = \frac{n_m}{R} + \frac{w^n R}{\alpha'}(-G_{mn} + B_{mn}) - q^I A_m^I - \frac{w^n R}{2} A_n^I A_m^I , \tag{11.6.17c}$$

where n_m and w^m are integers and q^I is on the Γ_{16} or $\Gamma_8 \times \Gamma_8$ lattice depending on which string has been compactified. The details are left to exercise 11.10. Let us note that with the gauge fields set to zero this reduces to the bosonic result (8.4.7). The terms in k_{Lm} and k_{Rm} that are linear in A^I come from the effect of the Wilson line on the periodicity, as in eq. (8.6.7). The term in k_L^I that is linear in A^I comes about as follows. For a string that winds around the compact dimension, the Wilson line implies that the current algebra fermions are no longer periodic. The corresponding vertex operator (10.3.25) shows that the momentum is shifted. Finally, the terms quadratic in A^I can be most easily checked by verifying that $\alpha' k \circ k/2$ is even.

To compare this spectrum with the Narain description one must go to coordinates in which $G_{m'n'} = \delta_{m'n'}$ so that $k_{m'} = e_{m'}{}^n k_n$, the tetrad being defined by $\delta_{p'q'} = e_{p'}{}^m e_{q'}{}^n G_{mn}$. The discrete T-duality group is generated by T-dualities on the separate axes, large spacetime coordinate transform-

ations, and quantized shifts of the antisymmetric tensor background and Wilson lines.

There is an interesting point here. Because the coset space (11.6.14) is the general solution to the consistency conditions, we must obtain this same set of theories whether we compactify the $SO(32)$ theory or the $E_8 \times E_8$ theory. From another point of view, note that the coset space is noncompact because of the Lorentzian signature — one can go to the limit of infinite Narain boost. Such a limit corresponds physically to taking one or more of the compact dimensions to infinite radius. Then one such limit gives the ten-dimensional $SO(32)$ theory, while another gives the ten-dimensional $E_8 \times E_8$ theory. Clearly one should think of all the different toroidally compactified heterotic strings as different states in a single theory. The two ten-dimensional theories are then distinct limits of this single theory.

Let us make the connection between these theories more explicit. Compactify the $SO(32)$ theory on a circle of radius R, with $G_{99} = 1$ and Wilson line

$$RA_9^I = \mathrm{diag}\left(\tfrac{1}{2}^8, 0^8\right) . \tag{11.6.18}$$

Adjoint states with one index from $1 \le A \le 16$ and one from $17 \le A \le 32$ are antiperiodic due to the Wilson line, so the gauge symmetry is reduced to $SO(16) \times SO(16)$. Now compactify the $E_8 \times E_8$ theory on a circle of radius R' with $G_{99} = 1$ and Wilson line

$$R'A_9^I = \mathrm{diag}\left(1, 0^7, 1, 0^7\right) . \tag{11.6.19}$$

The integer-charged states from the $SO(16)$ root lattice in each E_8 remain periodic while the half-integer charged states from the $SO(16)$ spinor lattices become antisymmetric. Again the gauge symmetry is $SO(16) \times SO(16)$. To see the relation between these two theories, focus on the states that are neutral under $SO(16) \times SO(16)$, those with $k_L^I = 0$. In both theories these are present only for $w = 2m$ even, because of the shift in k_L^I. The respective neutral spectra are

$$k_{L,R} = \frac{\tilde{n}}{R} \pm \frac{2mR}{\alpha'} , \quad k'_{L,R} = \frac{\tilde{n}'}{R'} \pm \frac{2m'R'}{\alpha'} , \tag{11.6.20}$$

with the subscript 9 suppressed. The primes denote the $E_8 \times E_8$ theory, and $\tilde{n} = n + 2m$, $\tilde{n}' = n' + 2m'$. We have used the explicit form of the Wilson line in each case, as well as the fact that $k_L^I = 0$. Under $(\tilde{n}, m) \leftrightarrow (m', \tilde{n}')$ and $(k_L, k_R) \leftrightarrow (k'_L, -k'_R)$, the spectra are identical if $RR' = \alpha'/2$. This symmetry extends to the full spectrum.

Finally let us ask how realistic a theory one obtains by compactification down to four dimensions. At generic points of moduli space the massless spectrum is given by dimensional reduction, simply classifying states by their four-dimensional symmetries. Analyzing the spectrum in terms of the four-dimensional $SO(2)$ helicity, the $SO(8)$ spins decompose as

$$\mathbf{8}_v \to +1,\ 0^6,\ -1 , \tag{11.6.21a}$$

$$\mathbf{8} \to +\tfrac{1}{2}^4,\ -\tfrac{1}{2}^4 , \tag{11.6.21b}$$

and so

$$\mathbf{8}_v \times \mathbf{8}_v \to +2,\ +1^{12},\ 0^{38},\ -1^{12},\ -2 , \tag{11.6.22a}$$

$$\mathbf{8} \times \mathbf{8}_v \to \tfrac{3}{2}^4,\ \tfrac{1}{2}^{28},\ -\tfrac{1}{2}^{28},\ -\tfrac{3}{2}^4 . \tag{11.6.22b}$$

From the supergravity multiplet there is a graviton, with helicities ± 2. There are four gravitinos, each with helicities $\pm\frac{3}{2}$. Toroidal compactification does not break any supersymmetry. Since in four dimensions the supercharge has four components, the 16 supersymmetries reduce to $d = 4$, $N = 4$ supersymmetry. The supergravity multiplet also includes 12 Kaluza–Klein and antisymmetric tensor gauge bosons, some fermions, and 36 moduli for the compactification. The final two spin-zero states are the dilaton and the axion. In four dimensions a two-tensor $B_{\mu\nu}$ is equivalent to a scalar (section B.4). This is the axion, whose physics we will discuss further in chapter 18.

In ten dimensions the only fields carrying gauge charge are the gauge field and gaugino. These reduce as discussed in section B.6 to an $N = 4$ vector multiplet — a gauge field, four Weyl spinors, and six scalars, all in the adjoint. For enhanced gauge symmetries, which are not present in ten dimensions, one still obtains the same $N = 4$ vector multiplet because of the supersymmetry. Compactification with $N = 4$ supersymmetry cannot give rise to the Standard Model because the fermions are necessarily in the adjoint of the gauge group. One gravitino is good, as we will explain in more detail in section 16.2, but four are too much of a good thing. We will see in chapter 16 that a fairly simple orbifold twist reduces the supersymmetry to $N = 1$ and gives a realistic spectrum.

Supersymmetry and BPS states

A little thought shows that the supersymmetry algebra of the toroidally compactified theory must be of the form[4]

$$\{Q_\alpha, Q_\beta^\dagger\} = 2P_\mu(\Gamma^\mu\Gamma^0)_{\alpha\beta} + 2P_{Rm}(\Gamma^m\Gamma^0)_{\alpha\beta} . \tag{11.6.23}$$

This differs from the simple dimensional reduction of the ten-dimensional algebra in that we have replaced P_m with P_{Rm}, the total right-moving

[4] For clarity a projection operator is omitted — all spinor indices in this equation must be in the **16**.

momentum k_{Rm} of all strings in a given state. These are equal only for a state of total winding number zero. To obtain the algebra (11.6.23) directly from a string calculation requires some additional machinery that we will not develop until the next chapter. However, it is clear that the algebra must take this form because the spacetime supersymmetry involves only the right-moving side of the heterotic string.

Let us look for Bogomolnyi–Prasad–Sommerfield (BPS) states, states that are annihilated by some of the Q_α. Take the expectation value of the algebra (11.6.23) in any state $|\psi\rangle$ of a single string of mass M in its rest frame. The left-hand side is a nonnegative matrix. The right-hand side is

$$2(M + k_{Rm}\Gamma^m\Gamma^0)_{\alpha\beta} \,. \tag{11.6.24}$$

The zero eigenvectors of this matrix are the supersymmetries that annihilate $|\psi\rangle$. Since $(k_{Rm}\Gamma^m\Gamma^0)^2 = k_R^2$, the eigenvalues of the matrix (11.6.24) are

$$2(M \pm |k_R|) \,, \tag{11.6.25}$$

with half having each sign. A BPS state therefore has $M^2 = k_R^2$. Recalling the heterotic string mass-shell conditions on the right-moving side,

$$M^2 = \begin{cases} k_R^2 + 4(\tilde{N} - \frac{1}{2})/\alpha' & (\text{NS}) \,, \\ k_R^2 + 4\tilde{N}/\alpha' & (\text{R}) \,, \end{cases} \tag{11.6.26}$$

the BPS states are those for which the right-movers are in an R ground state or in an NS state with one $\psi_{-1/2}$ excited. The latter are the lowest NS states to survive the GSO projection, so it makes sense to change terminology at this point and call them ground states as well. The BPS states are then precisely those states for which the right-moving side is in its $\mathbf{8}_v + \mathbf{8}$ ground state, but with arbitrarily large k_R. These can be paired with many possible states on the left-moving side. The left-moving mass-shell condition is

$$M^2 = k_L^2 + 4(N - 1)/\alpha' \tag{11.6.27}$$

or

$$N = 1 + \alpha'(k_R^2 - k_L^2)/4 = 1 - n_m w^m \,. \tag{11.6.28}$$

Any left-moving oscillator state is possible, as long as the compact momenta and winding satisfy the condition (11.6.28). For any given left-moving state, the 16 right-moving states $\mathbf{8}_v + \mathbf{8}$ form an ultrashort multiplet of the supersymmetry algebra, as compared to the 256 states in a normal massive multiplet.

It is interesting to look at the ten-dimensional origin of the modified

supersymmetry algebra (11.6.23). Rewrite the algebra as

$$\{Q_\alpha, Q_\beta^\dagger\} = 2P_M(\Gamma^M\Gamma^0)_{\alpha\beta} - 2\frac{\Delta X_m}{2\pi\alpha'}(\Gamma^m\Gamma^0)_{\alpha\beta} \,, \tag{11.6.29}$$

where ΔX^m is the total winding of the string. Consider the limit that the compactification radii become macroscopic, so that a winding string is macroscopic as well. The central charge term in the supersymmetry algebra must be proportional to a conserved charge, so we are looking for a charge proportional to the length ΔX of a string. Indeed, the string couples to the antisymmetric two-tensor field as

$$\frac{1}{2\pi\alpha'}\int_M B = \frac{1}{2}\int d^{10}x\, j^{MN}(x)B_{MN}(x) \tag{11.6.30a}$$

$$j^{MN}(x) = \frac{1}{2\pi\alpha'}\int_M d^2\sigma\,(\partial_1 X^M \partial_2 X^N - \partial_1 X^N \partial_2 X^M)\delta^{10}(x - X(\sigma)) \,. \tag{11.6.30b}$$

This is the natural generalization of the gauge coupling of a point particle, as discussed in section B.4. Integrating the current at fixed time gives the charge

$$Q^M = \int d^9x\, j^{M0} = \frac{1}{2\pi\alpha'}\int dX^M \,, \tag{11.6.31}$$

the integral running along the world-line of the string. The full supersymmetry algebra is then

$$\{Q_\alpha, Q_\beta^\dagger\} = 2(P_M - Q_M)(\Gamma^M\Gamma^0)_{\alpha\beta} \,. \tag{11.6.32}$$

In ten noncompact dimensions the charge (11.6.31) vanishes for any finite closed string but can be carried by an infinite string, for example an infinite straight string which would arise as the $R \to \infty$ limit of a winding string. It is often useful to contemplate such macroscopic strings, which of course have infinite total mass and charge but finite values per unit length. Under compactification the combination $P_m - Q_m$ is the right-moving gauge charge. The left-moving charges do not appear in the supersymmetry algebra.

It is natural to wonder whether the algebra (11.6.32) is now complete, and in fact it is not. Consider compactification to four dimensions at a generic point in the moduli space where the gauge symmetry is broken to $U(1)^{22}$. Grand unified theories in which the $U(1)$ of the Standard Model is embedded in a simple group always have magnetic monopoles arising from the quantization of topologically nontrivial classical solutions. String theory is not an ordinary grand unified theory but it also has magnetic monopoles. Compactification of the heterotic string leads to three kinds of gauge symmetry: the ten-dimensional symmetries, the Kaluza–Klein symmetries, and the antisymmetric tensor symmetries. For each there is

a corresponding monopole solution: the 't Hooft–Polyakov monopole, the Kaluza–Klein monopole, and the H-monopole respectively. Of course since the various charges are interchanged by the $O(22,6,\mathbf{Z})$ T-duality, the monopoles must be as well. Monopole charges appear in the supersymmetry algebra; in the present case it is again the right-moving charges that appear. In the low energy supergravity theory there is a symmetry that interchanges the electric and magnetic charges, so they must appear in the supersymmetry algebra in a symmetric way. We will discuss similar central charge terms extensively in chapters 13 and 14.

Exercises

11.1 Show that the operators (10.7.21) with appropriate normalization generate the full $N = 2$ superconformal algebra (11.1.4).

11.2 Show that if a $(\frac{3}{2},0)$ constraint j_F is not tensor, then $L_1 \cdot j_F$ is a nonvanishing $(\frac{1}{2},0)$ constraint, and a linear combination of $L_{-1} \cdot L_1 \cdot j_F$ and j_F is a tensor $(\frac{3}{2},0)$ constraint.

11.3 Show that if we take the GSO projection on the λ^4 in groups of eight, modular invariance is inconsistent with spacetime spin-statistics. Show that the OPE does not close.

11.4 (a) Find the massless and tachyonic states in the theory obtained by twisting the diagonal theory on the group generated by $\exp(\pi i F_1)$.
(b) Do the same for the group generated by $\exp(\pi i F_1)$ and $\exp(\pi i F_2)$.

11.5 (a) The decompositions of the spinor representation under $SO(16) \to SO(6) \times SO(10)$ and under $SO(6) \to SU(3) \times U(1)$ are obtained in section B.1. Use this to show that the adjoint of E_8 decomposes into $SU(3)$ representations with the degeneracies (11.4.24). The 78 generators neutral under $SU(3)$ must form a closed algebra: this is E_6.
(b) Use the same decompositions to show that the E_6 representations decompose as shown in table 11.4 under $E_6 \to SO(10) \times U(1)$.
(c) In a similar way obtain the decompositions shown in table 11.4 for $SO(10) \to SU(5) \times U(1)$.

11.6 Repeat parts (a) and (b) of the previous exercise for $SO(16) \to SO(4) \times SO(12)$ and $SO(4) \to SU(2) \times U(1)$ to obtain the analogous properties of E_7.

11.7 Show that $:jj(0):= j^a_{-1} \cdot j^a_{-1} \cdot 1$. Act with the Laurent expansion (11.5.2) for $j^c(z)$ and verify the OPE (11.5.20) in the Sugawara construction. Similarly verify the OPE (11.5.24).

11.8 For the free-fermion currents (11.5.16) for $SO(n)$, verify that the Sugawara construction gives the usual bilinear T_B.

11.9 Show that the lattice

$$\Gamma = \sum_r \Gamma_r \times \tilde{\Gamma}_r$$

is even and self-dual, where Γ_r is a weight sublattice of $SO(22)$, $\tilde{\Gamma}_r$ the same weight sublattice for $SO(6)$, and the sum runs over the four sublattices of $SO(2n)$. Show that this gives a four-dimensional compactification of the heterotic string with $SO(44) \times U(1)^6$ gauge symmetry.

11.10 (a) Verify the spectrum (11.6.17) for one compact dimension with a Wilson line background only.
(b) For the full spectrum (11.6.17), verify that $\alpha' k \circ k/2$ is even for any state and that $\alpha' k \circ k'/2$ is integral for any pair of states. The $\circ$ product is

$$k \circ k = k_L^I k_L^{\prime I} + G^{mn}(k_{Lm}k'_{Ln} - k_{Rm}k'_{Rn}) .$$

(c) (Optional) Verify the full result (11.6.17) by canonical quantization. Recall that the antisymmetric tensor background has already been treated in chapter 8. Reference: Narain, Sarmadi, & Witten (1987).

11.11 In the $E_8 \times E_8$ string, the currents $i\partial H^I$ plus the vertex operators for the points of length two form a set of (1,0) currents satisfying the $E_8 \times E_8$ algebra. From the $1/z$ term in the OPE, find the commutation relations of E_8. Be sure to include the cocycle in the vertex operator.

11.12 Find the Hagedorn temperatures of the type I, II, and heterotic string theories. Use the result (7.2.30) for the asymptotics of the partition function to express the Hagedorn temperature in general form.

12
Superstring interactions

In this chapter we will examine superstring interactions from two complementary points of view. First we study the interactions of the massless degrees of freedom, which are highly constrained by supersymmetry. The first section discusses the tree-level interactions, while the second discusses an important one-loop effect: the anomalies in local spacetime symmetries. We then develop superstring perturbation theory. We introduce superfields and super-Riemann surfaces to give superconformal symmetry a geometric interpretation, and calculate a variety of tree-level and one-loop amplitudes.

12.1 Low energy supergravity

The ten-dimensional supersymmetric string theories all have 32 or 16 supersymmetry generators. This high degree of supersymmetry completely determines the low energy action.

Type IIA superstring

We begin by discussing the field theory that has the largest possible spacetime supersymmetry and Poincaré invariance, namely eleven-dimensional supergravity. As explained in the appendix, the upper limit on the dimension arises because nontrivial consistent field theories cannot have massless particles with spins greater than two.

This theory would seem to have no direct connection to superstring theory, which requires ten dimensions. Our immediate interest in it is that, as discussed in section B.5, its supersymmetry algebra is the same as that of the IIA theory. The action of the latter can therefore be obtained by dimensional reduction, toroidal compactification keeping only fields that are independent of the compact directions. For now this is just a trick to

take advantage of the high degree of supersymmetry, but in chapter 14 we will see that there is much more going on.

The eleven-dimensional supergravity theory has two bosonic fields, the metric G_{MN} and a 3-form potential $A_{MNP} \equiv A_3$ with field strength F_4. Higher-dimensional supergravities contain many different p-form fields; to distinguish these from one another we will denote the rank by an italicized subscript, as opposed to numerical tensor indices which are written in roman font. In terms of the $SO(9)$ spin of massless states, the metric gives a traceless symmetric tensor with 44 states, and the 3-form gives a rank 3 antisymmetric tensor with 84 states. The total number of bosonic states is then 128, equal to the dimension of the $SO(9)$ spinor.

The bosonic part of the action is given by

$$2\kappa_{11}^2 S_{11} = \int d^{11}x(-G)^{1/2}\left(R - \frac{1}{2}|F_4|^2\right) - \frac{1}{6}\int A_3 \wedge F_4 \wedge F_4 \ . \quad (12.1.1)$$

The form action, written out fully, is proportional to

$$\int d^dx(-G)^{1/2}|F_p|^2 = \int d^dx \frac{(-G)^{1/2}}{p!} G^{M_1N_1}\dots G^{M_pN_p}F_{M_1\dots M_p}F_{N_1\dots N_p} \ . \quad (12.1.2)$$

The $p!$ cancels the sum over permutations of the indices, so that each independent component appears with coefficient 1. Forms are written as tensors with lower indices in order that their gauge transformations do not involve the metric.

We will take such results from the literature without derivation. Our interest is only in certain general features of the various actions, and we will not write out the full fermionic terms or supersymmetry transformations. For the supergravities arising from string theories, one can verify the action by comparison with the low energy limits of string amplitudes; a few such calculations are given later in the chapter and in the exercises. Also, many important features, such as the coupling of the dilaton, will be understood from general reasoning.

Now dimensionally reduce as in section 8.1. The general metric that is invariant under translations in the 10-direction is

$$\begin{aligned} ds^2 &= G^{11}_{MN}(x^\mu)dx^M dx^N \\ &= G^{10}_{\mu\nu}(x^\mu)dx^\mu dx^\nu + \exp(2\sigma(x^\mu))[dx^{10} + A_\nu(x^\mu)dx^\nu]^2 \ . \end{aligned} \quad (12.1.3)$$

Here M, N run from 0 to 10 and μ, ν from 0 to 9. We have added a superscript 11 to the metric appearing in the earlier supergravity action and introduced a new ten-dimensional metric $G^{10}_{\mu\nu} \neq G^{11}_{\mu\nu}$. The ten-dimensional metric will appear henceforth, so the superscript 10 will be omitted.

The eleven-dimensional metric (12.1.3) reduces to a ten-dimensional metric, a gauge field A_1, and a scalar σ. The potential A_3 reduces to two

potentials A_3 and A_2, the latter coming from components where one index is along the compact 10-direction. The three terms (12.1.1) become

$$S_1 = \frac{1}{2\kappa_{10}^2}\int d^{10}x\,(-G)^{1/2}\left(e^{\sigma}R - \frac{1}{2}e^{3\sigma}|F_2|^2\right), \tag{12.1.4a}$$

$$S_2 = -\frac{1}{4\kappa_{10}^2}\int d^{10}x\,(-G)^{1/2}\left(e^{-\sigma}|F_3|^2 + e^{\sigma}|\tilde{F}_4|^2\right), \tag{12.1.4b}$$

$$S_3 = -\frac{1}{4\kappa_{10}^2}\int A_2\wedge F_4\wedge F_4 = -\frac{1}{4\kappa_{10}^2}\int A_3\wedge F_3\wedge F_4\ . \tag{12.1.4c}$$

We have compactified the theory on a circle of coordinate radius $2\pi R$ and defined $\kappa_{10}^2 = \kappa_{11}^2/2\pi R$. To keep the normalization of the form actions uniform we have rescaled $A_{2,3} \to (2\pi R)^{-1/2}A_{2,3}$. The normalization of the kinetic terms is canonical for $2\kappa_{10}^2 = 1$.

In the action (12.1.4) we have defined

$$\tilde{F}_4 = dA_3 - A_1\wedge F_3\ , \tag{12.1.5}$$

the second term arising from the components $G^{\mu 10}$ in the 4-form action (12.1.2). We will use $F_{p+1} = dA_p$ to denote the simple exterior derivative of a potential, while field strengths with added terms are distinguished by a tilde as in eq. (12.1.5). Note that the action contains several terms where p-form potentials appear, rather than their exterior derivatives, but which are still gauge invariant. These are known as *Chern–Simons* terms, and we see that they are of two types. One involves the wedge product of one potential with any number of field strengths, and it is gauge invariant as a consequence of the Bianchi identities for the field strengths. The other appears in the kinetic term for the modified field strength (12.1.5). The second term in $\tilde{F}_4$ has a gauge variation

$$-d\lambda_0\wedge F_3 = -d(\lambda_0\wedge F_3). \tag{12.1.6}$$

It is canceled by a transformation

$$\delta' A_3 = \lambda_0\wedge F_3\ , \tag{12.1.7}$$

which is in addition to the usual $\delta A_3 = d\lambda_2$. In the present case, the Kaluza–Klein gauge transformation λ_0 originates from reparameterization of x^{10}, and the transformation (12.1.7) is simply part of the eleven-dimensional tensor transformation. Since the combination $\tilde{F}_4$ is invariant under both χ_0 and χ_2 transformations we should regard it as the physical field strength, but with a nonstandard Bianchi identity

$$d\tilde{F}_4 = -F_2\wedge F_3\ . \tag{12.1.8}$$

Poincaré duality of the form theory, developed in section B.4 for forms without Chern–Simons terms, interchanges these two kinds of Chern–Simons term.

The fields of the reduced theory are the same as the bosonic fields of the IIA string, as they must be. In particular the scalar σ must be the dilaton Φ, up to some field redefinition. The terms in the action have a variety of σ-dependences. Recall that the string coupling constant is determined by the value of the dilaton. As discussed in section 3.7, this means that after appropriate field redefinitions the tree-level spacetime action is multiplied by an overall factor $e^{-2\Phi}$, and otherwise depends on Φ only through its derivatives. 'Appropriate redefinitions' means that the fields are the same as those appearing in the string world-sheet sigma model action.

Since we have arrived at the action (12.1.4) without reference to string theory, we have no idea as yet how these fields are related to those in the world-sheet action. We will proceed by guesswork, and then explain the result in world-sheet terms. First redefine

$$G_{\mu\nu} = e^{-\sigma}G_{\mu\nu}(\text{new}), \quad \sigma = \frac{2\Phi}{3}\ . \tag{12.1.9}$$

The original metric will no longer appear, so to avoid cluttering the equations we do not put a prime on the new metric. Then

$$S_{\text{IIA}} = S_{\text{NS}} + S_{\text{R}} + S_{\text{CS}}\ , \tag{12.1.10a}$$

$$S_{\text{NS}} = \frac{1}{2\kappa_{10}^2}\int d^{10}x\,(-G)^{1/2}e^{-2\Phi}\left(R + 4\partial_\mu\Phi\partial^\mu\Phi - \frac{1}{2}|H_3|^2\right), \tag{12.1.10b}$$

$$S_{\text{R}} = -\frac{1}{4\kappa_{10}^2}\int d^{10}x\,(-G)^{1/2}\left(|F_2|^2 + |\tilde{F}_4|^2\right), \tag{12.1.10c}$$

$$S_{\text{CS}} = -\frac{1}{4\kappa_{10}^2}\int B_2\wedge F_4\wedge F_4\ . \tag{12.1.10d}$$

Note that $R \to e^{\sigma}R + \ldots$, that $(-G)^{1/2} \to e^{-5\sigma}(-G)^{1/2}$, and that the form action (12.1.2) scales as $e^{(p-5)\sigma}$.

We have regrouped terms according to whether the fields are in the NS–NS or R–R sector of the string theory; the Chern–Simons action contains both. It will be useful to distinguish R–R from NS–NS forms, so for the R–R fields we henceforth use C_p and F_{p+1} for the potential and field strength, and for the NS–NS fields B_2 and H_3. Also, we will use A_1 and F_2 for the open string and heterotic gauge fields, and B_2 and H_3 for the heterotic antisymmetric tensor.

The NS action now involves the dilaton in standard form. Eq. (12.1.9) is the unique redefinition that does this. The R action does not have the expected factor of $e^{-\Phi}$, but can be brought to this form by the further redefinition

$$C_1 = e^{-\Phi}C_1'\ , \tag{12.1.11}$$

after which

$$\int d^{10}x(-G)^{1/2}|F_2|^2 = \int d^{10}x(-G)^{1/2}e^{-2\Phi}|F_2'|^2 , \quad (12.1.12a)$$

$$F_2' \equiv dC_1' - d\Phi \wedge C_1' , \quad (12.1.12b)$$

and similarly for F_3 and C_4. The action (12.1.12) makes explicit the dilaton dependence of the loop expansion, but at the cost of complicating the Bianchi identity and gauge transformation,

$$dF_2' = d\Phi \wedge F_2' , \quad \delta C_1' = d\chi_0' - \chi_0' d\Phi . \quad (12.1.13)$$

For this reason the form (12.1.10) is usually used. For example, in a time-dependent dilaton field, it is the charge to which the unprimed fields couple that will be conserved.

Let us now make contact with string theory and see why the background R–R fields appearing in the world-sheet action have the more complicated properties (12.1.13). We work at the linearized level, in terms of the vertex operators

$$\mathscr{V}_\alpha \tilde{\mathscr{V}}_\beta (C\Gamma^{\mu_1\dots\mu_p})_{\alpha\beta} e_{\mu_1\dots\mu_p}(X) . \quad (12.1.14)$$

Here $\mathscr{V}_\alpha$ is the R ground state vertex operator (10.4.25) and $\Gamma^{\mu_1\dots\mu_p} = \Gamma^{[\mu_1}\dots\Gamma^{\mu_p]}$. The nontrivial physical state conditions are from $G_0 \sim p_\mu \psi_0^\mu$ and $\tilde{G}_0 \sim p_\mu \tilde{\psi}_0^\mu$, and amount to two Dirac equations, one acting on the left spinor index and one on the right:

$$\Gamma^\nu \Gamma^{\mu_1\dots\mu_p} \partial_\nu e_{\mu_1\dots\mu_p}(X) = \Gamma^{\mu_1\dots\mu_p}\Gamma^\nu \partial_\nu e_{\mu_1\dots\mu_p}(X) = 0 . \quad (12.1.15)$$

By antisymmetrizing all $p+1$ gamma matrices and keeping anticommutators one obtains

$$\Gamma^\nu \Gamma^{\mu_1\dots\mu_p} = \Gamma^{\nu\mu_1\dots\mu_p} + p\eta^{\nu[\mu_1}\Gamma^{\mu_2\dots\mu_p]} , \quad (12.1.16a)$$

$$\Gamma^{\mu_1\dots\mu_p}\Gamma^\nu = (-1)^p\Gamma^{\nu\mu_1\dots\mu_p} + (-1)^{p+1}p\eta^{\nu[\mu_1}\Gamma^{\mu_2\dots\mu_p]} . \quad (12.1.16b)$$

The Dirac equations (12.1.15) are then equivalent to

$$de_p = d{*}e_p = 0 . \quad (12.1.17)$$

These are first order equations, unlike the second order equations encountered previously for bosonic fields. In fact, they have the same form as the field equation and Bianchi identity for a p-form field strength. Thus we identify the function $e_{\mu_1\dots\mu_p}(X)$ appearing in the vertex operator as the R–R *field strength* rather than potential. To confirm this, observe that in the IIA theory the spinors in the R–R vertex operator (12.1.14) have opposite chirality and so their product in table 10.1 contains forms of even rank, the same as the IIA R–R field strengths.

This has one consequence that will be important later on. Amplitudes for R–R forms will always contain a power of the momentum and so

vanish at zero momentum. The zero-momentum coupling of a gauge field measures the charge, so this means that *strings are neutral under all R–R gauge fields.*

The derivation of the field equations (12.1.17) was for a flat background. Now let us consider the effect of a dilaton gradient. It is convenient that the linear dilaton background gives rise to the free CFT (10.1.22),

$$T_F = i(2/\alpha')^{1/2}\psi^\mu \partial X_\mu - 2i(\alpha'/2)^{1/2}\Phi_{,\mu}\partial\psi^\mu , \quad (12.1.18a)$$

$$G_0 \sim (\alpha'/2)^{1/2}\psi_0^\mu(p_\mu + i\Phi_{,\mu}) , \quad (12.1.18b)$$

and similarly for $\tilde{T}_F$ and $\tilde{G}_0$. The field equations are modified to

$$(d - d\Phi\wedge)e_p = (d - d\Phi\wedge) * e_p = 0 . \quad (12.1.19)$$

Thus the Bianchi identity and field equation for the string background fields are modified in the fashion deduced from the action. There is no such modification for the NS–NS tensor. It couples to the world-sheet through its potential,

$$\frac{1}{2\pi\alpha'}\int_M B_2 . \quad (12.1.20)$$

This is invariant under $\delta B_2 = d\chi_1$ independent of the dilaton, and so $H_3 = dB_2$ is invariant and $dH_3 = 0$.

Massive IIA supergravity

There is a generalization of the IIA supergravity theory which has no simple connection with eleven-dimensional supergravity but which plays a role in string theory. The IIA theory has a 2-form and a 4-form field strength, and by Poincaré duality a 6-form and an 8-form as well,

$$\tilde{F}_6 = *\tilde{F}_4 , \quad \tilde{F}_8 = *F_2 ; \quad (12.1.21)$$

again, a tilde denotes a field strength with a nonstandard Bianchi identity. The pattern suggests we also consider a 10-form $F_{10} = dC_9$. The free field equation would be

$$d{*}F_{10} = 0 , \quad (12.1.22)$$

and since $*F_{10}$ is a scalar this means that

$$F_{10} = \text{constant} . \quad (12.1.23)$$

Thus there are no propagating degrees of freedom. Nevertheless, such a field would have a physical effect, since it would carry energy density. This is closely analogous to an electric field F_2 in two space-time dimensions, where there are no propagating photons but there is an energy density and a linear potential that confines charges.

Such a field can indeed be included in IIA supergravity. The action is

$$S'_{\rm IIA} = \tilde{S}_{\rm IIA} - \frac{1}{4\kappa_{10}^2}\int d^{10}x(-G)^{1/2}M^2 + \frac{1}{2\kappa_{10}^2}\int MF_{10} \ . \qquad (12.1.24)$$

Here $\tilde{S}_{\rm IIA}$ is the earlier IIA action (12.1.10) with the substitutions

$$F_2 \to F_2 + MB_2 \ , \quad F_4 \to F_4 + \frac{1}{2}MB_2 \wedge B_2 \ , \quad \tilde{F}_4 \to \tilde{F}_4 + \frac{1}{2}MB_2 \wedge B_2 \ . \qquad (12.1.25)$$

The scalar M is an auxiliary field, meaning that it appears in the action without derivatives (and in this case only quadratically). Thus it can be integrated out, at the cost of introducing a rather nonlinear dependence on B_2.

We will see in the next chapter that this massive supergravity does arise in the IIA string. To put the 9-form potential in perspective, observe that the maximum-rank potential that gives rise to a propagating field in ten dimensions is an 8-form, whose 9-form field strength is dual to a 1-form. The latter is just the gradient of the R–R scalar field C_0. A 10-form potential also fits in ten dimensions but does not give rise to propagating states. We saw in section 10.8 that this does exist in the type I string, so we should not be surprised that the 9-form will appear in string theory as well.

Type IIB superstring

For low energy IIB supergravity there is a problem due to the self-dual field strength $H_5 = *H_5$. As discussed in section B.4 there is no covariant action for such a field, but the following comes close:

$$S_{\rm IIB} = S_{\rm NS} + S_{\rm R} + S_{\rm CS} \ , \qquad (12.1.26a)$$

$$S_{\rm NS} = \frac{1}{2\kappa_{10}^2}\int d^{10}x(-G)^{1/2}e^{-2\Phi}\left(R + 4\partial_\mu\Phi\partial^\mu\Phi - \frac{1}{2}|H_3|^2\right) , \qquad (12.1.26b)$$

$$S_{\rm R} = -\frac{1}{4\kappa_{10}^2}\int d^{10}x(-G)^{1/2}\left(|F_1|^2 + |\tilde{F}_3|^2 + \frac{1}{2}|\tilde{F}_5|^2\right) , \qquad (12.1.26c)$$

$$S_{\rm CS} = -\frac{1}{4\kappa_{10}^2}\int C_4 \wedge H_3 \wedge F_3 \ , \qquad (12.1.26d)$$

where

$$\tilde{F}_3 = F_3 - C_0 \wedge H_3 \ , \qquad (12.1.27a)$$

$$\tilde{F}_5 = F_5 - \frac{1}{2}C_2 \wedge H_3 + \frac{1}{2}B_2 \wedge F_3 \ . \qquad (12.1.27b)$$

The NS–NS action is the same as in IIA supergravity, while the R–R and Chern–Simons actions are closely parallel in form. The equation of motion and Bianchi identity for $\tilde{H}_5$ are

$$d*\tilde{F}_5 = d\tilde{F}_5 = H_3 \wedge F_3 \ . \qquad (12.1.28)$$

Recall that the spectrum of the IIB string includes the degrees of freedom of a self-dual 5-form field strength. The field equations from the action (12.1.26) are consistent with

$$*\tilde{F}_5 = \tilde{F}_5 \qquad (12.1.29)$$

but they do not imply it. This must be imposed as an added constraint on the *solutions;* it cannot be imposed on the action or else the wrong equations of motion result.

This formulation is satisfactory for a classical treatment but it is not simple to impose the constraint in the quantum theory. This will not be important for our purposes, and we leave further discussion to the references. Our main interest in this action is a certain $SL(2,\mathbf{R})$ symmetry. Let

$$G_{{\rm E}\mu\nu} = e^{-\Phi/2}G_{\mu\nu} \ , \quad \tau = C_0 + ie^{-\Phi} \ , \qquad (12.1.30a)$$

$$\mathcal{M}_{ij} = \frac{1}{{\rm Im}\,\tau}\begin{bmatrix} |\tau|^2 & -{\rm Re}\,\tau \\ -{\rm Re}\,\tau & 1 \end{bmatrix} , \quad F_3^i = \begin{bmatrix} H_3 \\ F_3 \end{bmatrix} . \qquad (12.1.30b)$$

Then

$$S_{\rm IIB} = \frac{1}{2\kappa_{10}^2}\int d^{10}x(-G_{\rm E})^{1/2}\left(R_{\rm E} - \frac{\partial_\mu\bar{\tau}\partial^\mu\tau}{2({\rm Im}\,\tau)^2} - \frac{1}{2}|F_1|^2 - \frac{\mathcal{M}_{ij}}{2}F_3^i\cdot F_3^j - \frac{1}{2}|\tilde{F}_5|^2\right) - \frac{\epsilon_{ij}}{4\kappa_{10}^2}\int C_4 \wedge F_3^i \wedge F_3^j \ , \qquad (12.1.31)$$

the Einstein metric (12.1.30a) being used everywhere. This is invariant under the following $SL(2,\mathbf{R})$ symmetry:

$$\tau' = \frac{a\tau + b}{c\tau + d} \ , \qquad (12.1.32a)$$

$$F_3'^i = \Lambda^i_j F_3^j \ , \quad \Lambda^i_j = \begin{bmatrix} d & c \\ b & a \end{bmatrix} , \qquad (12.1.32b)$$

$$\tilde{F}'_5 = \tilde{F}_5 \ , \quad G'_{{\rm E}\mu\nu} = G_{{\rm E}\mu\nu} \ , \qquad (12.1.32c)$$

with a, b, c, and d real numbers such that $ad - bc = 1$. The $SL(2,\mathbf{R})$ invariance of the τ kinetic term is familiar, and that of the F_3 kinetic term follows from

$$\mathcal{M}' = (\Lambda^{-1})^T\mathcal{M}\Lambda^{-1} \ . \qquad (12.1.33)$$

This $SL(2,\mathbf{R})$ invariance is as claimed in the second line of table B.3. Any given value τ is invariant under an $SO(2,\mathbf{R})$ subgroup so the moduli space is the coset $SL(2,\mathbf{R})/SO(2,\mathbf{R})$. If we now compactify on tori, the moduli

and other fields fall into multiplets of the larger symmetries indicated in the table and the low energy action has the larger symmetry.

Observe that this $SL(2,\mathbf{R})$ mixes the two 2-form potentials. We know that the NS–NS form couples to the string and the R–R form does not. The $SL(2,\mathbf{R})$ might thus seem to be an accidental symmetry of the low energy theory, not relevant to the full string theory. Indeed, this was assumed for some time, but now we know better. As we will explain in chapter 14, the discrete subgroup $SL(2,\mathbf{Z})$ is an exact symmetry.

Type I superstring

To obtain the type I supergravity action requires three steps: set to zero the IIB fields C_0, B_2, and C_4 that are removed by the Ω projection; add the gauge fields, with appropriate dilaton dependence for an open string field; and, modify the F_3 field strength. This gives

$$S_{\rm I} = S_{\rm c} + S_{\rm o} \,, \tag{12.1.34a}$$

$$S_{\rm c} = \frac{1}{2\kappa_{10}^2}\int d^{10}x\,(-G)^{1/2}e^{-2\Phi}\left(R + 4\partial_\mu\Phi\partial^\mu\Phi - \frac{1}{2}|\tilde{F}_3|^2\right) , \tag{12.1.34b}$$

$$S_{\rm o} = -\frac{1}{2g_{10}^2}\int e^{-\Phi}{\rm Tr}_{\rm v}(|F_2|^2) \,. \tag{12.1.34c}$$

The open string $SO(32)$ potential and field strength are written as matrix-valued forms A_1 and F_2, which are in the vector representation as indicated by the subscript on the trace. Here

$$\tilde{F}_3 = dC_2 - \frac{\kappa_{10}^2}{g_{10}^2}\omega_3 \,, \tag{12.1.35}$$

and ω_3 is the *Chern–Simons 3-form*

$$\omega_3 = {\rm Tr}_{\rm v}\left(A_1 \wedge dA_1 + \frac{2}{3}A_1 \wedge A_1 \wedge A_1\right) . \tag{12.1.36}$$

Again the modification of the field strength implies a modification of the gauge transformation. Under an ordinary gauge transformation $\delta A_1 = d\lambda - i[A_1,\lambda]$, the Chern-Simons form transforms as

$$\delta\omega_3 = d{\rm Tr}_{\rm v}(\lambda F_2). \tag{12.1.37}$$

Thus it must be that

$$\delta C_2 = \frac{\kappa_{10}^2}{g_{10}^2}{\rm Tr}_{\rm v}(\lambda F_2) \,. \tag{12.1.38}$$

The 2-form transformation $\delta C_2 = d\chi_1$ is unaffected.

The action appears to contain two parameters, κ_{10} with units of L^4 and g_{10} with units of L^3. We can think of κ_{10} as setting the scale because it is dimensionful, but there is one dimensionless combination $\kappa_{10}g_{10}^{-4/3}$. However, under an additive shift $\Phi \to \Phi + \zeta$, the couplings change $\kappa_{10} \to e^{\zeta}\kappa_{10}$ and $g \to e^{\zeta/2}g$ and so this ratio can be set to any value by a change of the background. Thus the low energy theory reflects the familiar string property that the coupling is not a fixed parameter but depends on the dilaton. The form of the action (12.1.34) is fixed by supersymmetry, but when we consider this as the low energy limit of string theory there is a relation between the closed string coupling κ_{10}, the open string coupling g_{10}, and the type I α'. We will derive this in the next chapter, from a D-brane calculation, as we did for the corresponding relation in the bosonic string.

Heterotic strings

The heterotic strings have the same supersymmetry as the type I string and so we expect the same action. However, in the absence of open strings or R–R fields the dilaton dependence should be $e^{-2\Phi}$ throughout:

$$S_{\rm het} = \frac{1}{2\kappa_{10}^2}\int d^{10}x\,(-G)^{1/2}e^{-2\Phi}\left[R + 4\partial_\mu\Phi\partial^\mu\Phi - \frac{1}{2}|\tilde{H}_3|^2 + \frac{\kappa_{10}^2}{g_{10}^2}{\rm Tr}_{\rm v}(|F_2^2|)\right] . \tag{12.1.39}$$

Here

$$\tilde{H}_3 = dB_2 - \frac{\kappa_{10}^2}{g_{10}^2}\omega_3 \,, \quad \delta B_2 = \frac{\kappa_{10}^2}{g_{10}^2}{\rm Tr}_{\rm v}(\lambda F_2) \tag{12.1.40}$$

are the same as in the type I string, with the form renamed to reflect the fact that it is from the NS sector.

Because of the high degree of supersymmetry, the type I and heterotic actions can differ only by a field redefinition. Indeed the reader can check that with the type I and heterotic fields related by

$$G_{{\rm I}\mu\nu} = e^{-\Phi_{\rm h}}G_{{\rm h}\mu\nu} \,, \quad \Phi_{\rm I} = -\Phi_{\rm h} \,, \tag{12.1.41a}$$

$$\tilde{F}_{{\rm I}3} = \tilde{H}_{{\rm h}3} \,, \quad A_{{\rm I}I} = A_{{\rm h}I} \,, \tag{12.1.41b}$$

the action (12.1.34) becomes the action (12.1.39). For the heterotic string, the relation among κ_{10}, g_{10}, and α' will be obtained later in this chapter, by two different methods; it is, of course, different from the relation in the type I theory.

For $E_8 \times E_8$ there is no vector representation, but it is convenient to use a normalization that is uniform with $SO(32)$. In place of ${\rm Tr}_{\rm v}(t^at^b)$ in the action use $\frac{1}{30}{\rm Tr}_{\rm a}(t^at^b)$. This has the property that for fields in any $SO(16)\times SO(16)$ subgroup it reduces to ${\rm Tr}_{\rm v}(t^at^b)$.

12.2 Anomalies

It is an important phenomenon that some classical symmetries are anomalous, meaning that they are not preserved by quantization. We encountered this for the Weyl anomaly in chapter 3. We also saw there that if the left- and right-moving central charges were not equal there was an anomaly in two-dimensional coordinate invariance.

In general, anomalies in local symmetries make a theory inconsistent, as unphysical degrees of freedom no longer decouple. Anomalies in global symmetries are not harmful but imply that the symmetry is no longer exact. Both kinds of anomalies play a role in the Standard Model. Potential local anomalies in gauge and coordinate invariance cancel among the quarks and leptons of each generation. Anomalies in global chiral symmetries of the strong interaction are important in accounting for the π^0 decay rate and the η' mass.

In this section we consider potential anomalies in the spacetime gauge and coordinate invariances in the various string theories. If the theories we have constructed are consistent these anomalies must be absent, and in fact they are. Although this can be understood in purely string theoretic terms it can also be understood from analysis of the low energy field theory, and it is useful to take both points of view.

We can analyze anomalies from the purely field theoretic point of view because of the odd property that they are both short-distance and long-distance effects. They are short-distance in the sense that they arise because the measure cannot be defined — the theory cannot be regulated — in an invariant way. They are long-distance in the sense that this impossibility follows entirely from the nature of the massless spectrum.

Let us illustrate this with another two-dimensional example, which is also of interest in its own right. Suppose we have left- and right-moving current algebras with the same algebra g, with the coefficients of the Schwinger terms being $\hat{k}_{L,R}\delta^{ab}$. Couple a gauge field to the current,

$$S_{\rm int} = \int d^2z\,(j_z^a A_{\bar z}^a + j_{\bar z}^a A_z^a)\,. \tag{12.2.1}$$

The OPE determines the jj expectation value, so to second order the path integral is

$$Z[A] = \frac{1}{2}\int d^2z_1\,d^2z_2\left[\frac{\hat{k}_L}{z_{12}^2}A_{\bar z}^a(z_1,\bar z_1)A_{\bar z}^a(z_2,\bar z_2) + \frac{\hat{k}_R}{\bar z_{12}^2}A_z^a(z_1,\bar z_1)A_z^a(z_2,\bar z_2)\right]. \tag{12.2.2}$$

Now make a gauge transformation, which to leading order is $\delta A_I^a = d\chi^a$.

Integrate by parts and use $\partial_z(1/\bar z^2) = -2\pi\partial_{\bar z}\delta^2(z,\bar z)$ to o

$$\delta Z[A] = -\pi\int d^2z\,\lambda^a(z,\bar z)\left[\hat{k}_L\partial_z A_{\bar z}^a(z,\bar z) + \hat{k}_R\partial_{\bar z}A_z^a(z,\right.$$

Now, consider the case that $\hat{k}_L = \hat{k}_R = \hat{k}$, where

$$\delta Z[A] = \pi\hat{k}\,\delta\int d^2z\,A_z^a(z,\bar z)A_{\bar z}^a(z,\bar z)\,.$$

Then

$$\begin{aligned} Z'[A] &= Z[A] - \pi\hat{k}\,\delta\int d^2z\,A_z^a(z,\bar z)A_{\bar z}^a(z,\bar z) \\ &= \frac{\hat{k}}{2}\int d^2z_1\,d^2z_2\,\ln|z_{12}^2|\,F_{z\bar z}^a(z_1,\bar z_1)F_{z\bar z}^a(z \end{aligned}$$

is gauge-invariant.

Let us run through the logic here. The path integral (1 but its gauge variation is local. The latter is necessarily variation can be thought of as arising from the regula evaluate the path integral by brute force. Although the it is not in general the variation of a local operator. is the case for $\hat{k}_L = \hat{k}_R$ here, one can subtract that loc the action to restore gauge invariance. In fact, with a regulator the needed local term will be produced by automatically. The OPE is unambiguous only for nonze the OPE calculation above doesn't determine the local t know which regulator we choose to use.

The final form (12.2.5) is written in terms of the field Abelian theory the full path integral is just the exponen non-Abelian theory the higher order terms are more cor condition $\hat{k}_L = \hat{k}_R$ for the symmetry to be preserved is s sufficient.

The two-dimensional gravitational anomaly was sim from the z^{-4} term in the TT OPE. Also, if there is a z^{-3} then there is a *mixed anomaly:* the current has an anor to the curvature and the coordinate invariance an anor to the field strength.

Note that these anomalies are all odd under parity, b to $\hat{k}_L - \hat{k}_R$ or $c_L - c_R$. Parity-symmetric theories can be d using a Pauli–Villars regulator. Also, the anomalies are add additional massive degrees of freedom. This follows f decoupling argument. Massive degrees of freedom gi to $Z[A]$ which at asymptotically long distance looks momentum). Any gauge variation of this can therefore variation of a local operator, and removed by a cour

reason the anomalies in superstring theory are determined by the massless spectrum, independent of the stringy details at short distance.

A single fermion of charge q coupled to a $U(1)$ gauge field contributes q^2 to the jj OPE. The anomaly cancellation conditions for free fermions coupled to such a field are

$$\text{gauge anomaly:} \quad \sum_L q^2 - \sum_R q^2 = 0 \,, \tag{12.2.6a}$$

$$\text{gravitational anomaly:} \quad \sum_L 1 - \sum_R 1 = 0 \,, \tag{12.2.6b}$$

$$\text{mixed anomaly:} \quad \sum_L q - \sum_R q = 0 \,. \tag{12.2.6c}$$

In four dimensions things are slightly different. For dimensional reasons the dangerous amplitudes have three currents and the anomaly is quadratic in the field strengths and curvatures. The antiparticle of a left-handed fermion of charge q is a right-handed fermion of charge $-q$, so the two terms in the anomaly are automatically equal for odd powers of q and opposite for even powers (including the purely gravitational anomaly), leaving the conditions:

$$\text{gauge anomaly:} \quad \sum_L q^3 = 0 \,, \tag{12.2.7a}$$

$$\text{mixed anomaly:} \quad \sum_L q = 0 \,. \tag{12.2.7b}$$

If there is more than one gauge group the necessary and sufficient condition for anomaly cancellation is that the above hold for every linear combination of generators.

The IIA theory is parity-symmetric and so automatically anomaly-free, while the others have potential anomalies. In ten dimensions the anomaly involves amplitudes with six currents (the hexagon graph) and is of fifth order in the field strengths and curvatures. The calculation has been done in detail in the literature; we will not repeat it here but just quote the result. First we must establish some notation. For the gravitational field, it is convenient to work in the tangent space (tetrad) formalism. In this formalism there are two local symmetries, coordinate invariance and local Lorentz transformations

$$e_\mu{}^p(x)' = e_\mu{}^q(x)\Theta_q{}^p(x) \,. \tag{12.2.8}$$

Both are necessary for the decoupling of unphysical degrees of freedom, and in fact when there is a coordinate anomaly one can by adding counterterms to the action convert it to a Lorentz anomaly, which closely resembles a gauge anomaly. The Riemann tensor can be written $R_{\mu\nu}{}^p{}_q$, with mixed spacetime and tangent space indices, and in this way be regarded as a 2-form R_2 which is a $d \times d$ tangent space matrix. Similarly

$e_\mu{}^q$ is written as a one-form which is a column vector in tangent space, and the field strength is written as a matrix 2-form $F_2 = F_2^a t_r^a$; here r is the representation carried by the matter

The anomaly can be written in compact form in terms of an *anomaly polynomial*, a formal $(d+2)$-form $\hat{I}_{d+2}(R_2, F_2)$. This has the property that it is the exterior derivative of a $(d+1)$-form, whose variation is the exterior derivative of a d-form:

$$\hat{I}_{d+2} = d\hat{I}_{d+1} \,, \quad \delta\hat{I}_{d+1} = d\hat{I}_d \,. \tag{12.2.9}$$

The anomalous variation of the path integral is then

$$\delta \ln Z = \frac{-i}{(2\pi)^5} \int \hat{I}_d(F_2, R_2) \,. \tag{12.2.10}$$

The anomaly cancellation condition is that the total anomaly polynomial vanish.

In the ten-dimensional supergravity theories there are three kinds of chiral field: the spinors **8** and **8′**, the gravitinos **56** and **56′**, and the field strengths $[5]_+$ and $[5]_-$ of the IIB theory. Parity interchanges the two fields in each pair so these make opposite contributions to the anomaly. The anomaly polynomials have been calculated. For the Majorana–Weyl **8**,

$$\begin{aligned}\hat{I}_{\mathbf{8}}(F_2, R_2) = &-\frac{\mathrm{Tr}(F_2^6)}{1440} \\ &+ \frac{\mathrm{Tr}(F_2^4)\mathrm{tr}(R_2^2)}{2304} - \frac{\mathrm{Tr}(F_2^2)\mathrm{tr}(R_2^4)}{23040} - \frac{\mathrm{Tr}(F_2^2)[\mathrm{tr}(R_2^2)]^2}{18432} \\ &+ \frac{n\,\mathrm{tr}(R_2^6)}{725760} - \frac{n\,\mathrm{tr}(R_2^4)\mathrm{tr}(R_2^2)}{552960} + \frac{n\,[\mathrm{tr}(R_2^2)]^3}{1327104} \,.\end{aligned} \tag{12.2.11}$$

For the Majorana–Weyl **56**,

$$\hat{I}_{\mathbf{56}}(F_2, R_2) = -495\frac{\mathrm{tr}(R_2^6)}{725760} - 225\frac{\mathrm{tr}(R_2^4)\mathrm{tr}(R_2^2)}{552960} - 63\frac{[\mathrm{tr}(R_2^2)]^3}{1327104} \,. \tag{12.2.12}$$

For the self-dual tensor,

$$\hat{I}_{\mathrm{SD}}(R_2) = 992\frac{\mathrm{tr}(R_2^6)}{725760} + 448\frac{\mathrm{tr}(R_2^4)\mathrm{tr}(R_2^2)}{552960} + 128\frac{[\mathrm{tr}(R_2^2)]^3}{1327104} \,. \tag{12.2.13}$$

The 'tr' denotes the trace on the tangent space indices p, q. In this section we will write products and powers of forms without the $\wedge$, to keep expressions compact. The 'Tr' denotes the trace of the field strength in the representation carried by the fermion. In particular, $n = \mathrm{Tr}(1)$ is the dimension of the representation. If the representation r is reducible, $r = r_1 + r_2 + \ldots$, the corresponding traces add: $\mathrm{Tr}_r = \mathrm{Tr}_{r_1} + \mathrm{Tr}_{r_2} + \ldots$.

Now let us consider the anomalies in the various chiral string theories.

Type IIB anomalies

In type IIB supergravity there are two **8**′s, two **56**s, and one $[5]_+$, giving the total anomaly polynomial

$$\hat{I}_{\mathrm{IIB}}(R_2) = -2\hat{I}_{\mathbf{8}}(R_2) + 2\hat{I}_{\mathbf{56}}(R_2) + \hat{I}_{\mathrm{SD}}(R_2) = 0\,. \tag{12.2.14}$$

There are no gauge fields so only the three purely gravitational terms enter, and the coefficients of these conspire to produce zero total anomaly. From the point of view of the low energy theory, this is somewhat miraculous. In fact, it seems accidental that there are any consistent chiral theories at all. There are three anomaly terms that must vanish and three free parameters — the net number of **8** minus **8**′, of **56** minus **56**′, and of $[5]_+$ minus $[5]_-$. Barring a numerical coincidence the only solution would be that all these differences vanish, a nonchiral theory. One can view string theory as explaining this numerical coincidence: the conditions for the internal consistency of string theory are reasonably straightforward, and having satisfied them, the low energy theory must be nonanomalous.

The existence of consistent chiral theories is a beautiful example of the consistency of string theory, and is also of some practical importance. The fermion content of the Standard Model is chiral — the weak interactions violate parity. This chiral property seems to be an important clue, and it has been a difficulty for many previous unifying ideas. Of course, in string theory we are still talking about the ten-dimensional spectrum, but we will see in later chapters that there is some connection between chirality in higher dimensions and in four.

Type I and heterotic anomalies

The type I and heterotic strings have the same low energy limits so we can discuss their anomalies together. There is an immediate problem. The only charged chiral field is the **8**′, so there is apparently no possibility of cancellation of gauge and mixed anomalies. This is a paradox because we have claimed that these string theories were constructed to satisfy all the conditions for unitarity. Our arguments were perhaps heuristic in places, but it is not so hard to carry out an explicit string calculation at one loop and verify the decoupling of null states. This contradiction led Green and Schwarz to a careful study of the structure of the string amplitude, and they found a previously unknown, and canceling, contribution to the anomaly.

The assertion that the anomaly cannot be canceled by local counterterms takes into account only terms constructed from the gauge field and metric. Consider, however, the Chern–Simons interaction

$$S' = \int B_2 \mathrm{Tr}(F_2^4) \tag{12.2.15}$$

(in any representation r, for now). This is invariant under gauge transformations of the vector potential because it is constructed from the field strength, and under the 2-form transformation $\delta B_2 = d\chi_1$ using integration by parts and the Bianchi identity for the field strength. However, we have seen that in the $N = 1$ supergravity theory the 2-form has a nontrivial gauge transformation $\delta B_2 \propto \mathrm{Tr}(\lambda F_2)$, eq. (12.1.40). Then

$$\delta S' \propto \int \mathrm{Tr}(\lambda F_2)\mathrm{Tr}(F_2^4)\,. \tag{12.2.16}$$

This is of the form (12.2.9) with

$$\hat{I}_d \propto \mathrm{Tr}(\lambda F_2)\mathrm{Tr}(F_2^4)\,, \quad \hat{I}_{d+1} \propto \mathrm{Tr}(AF_2)\mathrm{Tr}(F_2^4)\,, \tag{12.2.17a}$$

$$\hat{I}_{d+2} \propto \mathrm{Tr}(F_2^2)\mathrm{Tr}(F_2^4)\,. \tag{12.2.17b}$$

Thus it can cancel an anomaly of this form. Similarly, the variation of

$$S'' = \int B_2 [\mathrm{Tr}(F_2^2)]^2 \tag{12.2.18}$$

can cancel the anomaly polynomial $[\mathrm{Tr}(F_2^2)]^3$.

The pure gauge anomaly polynomial has a different group-theoretic form $\mathrm{Tr_a}(F_2^6)$, now in the adjoint representation because the charged fields are gauginos. However, for certain algebras there are relations between the different invariants. For $SO(n)$, it is convenient to convert all traces into the vector representation. The fermions of the supergravity theory are always in the adjoint; in terms of the vector traces these are

$$\mathrm{Tr_a}(t^2) = (n-2)\mathrm{Tr_v}(t^2)\,, \tag{12.2.19a}$$

$$\mathrm{Tr_a}(t^4) = (n-8)\mathrm{Tr_v}(t^4) + 3\,\mathrm{Tr_v}(t^2)\mathrm{Tr_v}(t^2)\,, \tag{12.2.19b}$$

$$\mathrm{Tr_a}(t^6) = (n-32)\mathrm{Tr_v}(t^6) + 15\,\mathrm{Tr_v}(t^2)\mathrm{Tr_v}(t^4)\,. \tag{12.2.19c}$$

Here t is any linear combination of generators, but this implies the same relations for symmetrized products of different generators. Symmetrized products appear when the anomaly polynomial is expanded in sums over generators, because the 2-forms F_2^a and F_2^b commute.

The last of these identities implies that precisely for $SO(32)$ the gauge anomaly $\mathrm{Tr_a}(F_2^6)$ is equal to a product of lower traces and so can be canceled by the variations of S' and S''. This is the *Green–Schwarz mechanism*. This is of course the same SO group that arises in the type I and heterotic strings, and not surprisingly the necessary interactions occur in these theories with the correct coefficients.

Also for the group E_8, the sixth order trace can be reduced to lower order traces,

$$\mathrm{Tr}_a(t^4) = \frac{1}{100}[\mathrm{Tr}_a(t^2)]^2\,, \quad \mathrm{Tr}_a(t^6) = \frac{1}{7200}[\mathrm{Tr}_a(t^2)]^3\,. \tag{12.2.20}$$

Using the relation $\mathrm{Tr}_a(t^m) = \mathrm{Tr}_{a1}(t^m) + \mathrm{Tr}_{a2}(t^m)$, it follows that the sixth power trace can be reduced for $E_8 \times E_8$ as well (with only one factor of E_8 the *gravitational* anomaly does not cancel, as we will see).

Now let us consider the full anomaly, including mixed anomalies. Generalizing S' and S'' to

$$\int B_2 X_8(F_2, R_2)\,, \tag{12.2.21}$$

makes it possible to cancel an anomaly of the form $\mathrm{Tr}(F_2^2)X_8(F_2,R_2)$ for arbitrary 8-form $X_8(F_2,R_2)$. In addition, the B_2 field strength includes also a gravitational Chern–Simons term:

$$\tilde{H}_3 = dB_2 - c\omega_{3Y} - c'\omega_{3L} \tag{12.2.22}$$

with c and c' constants. Here $\omega_{3Y} = A_I dA_I + \frac{2}{3}A_I^3$ is the gauge Chern–Simons term as before and

$$\omega_{3L} = \omega_I d\omega_I + \frac{2}{3}\omega_I^3 \tag{12.2.23}$$

is the Lorentz Chern–Simons term, with $\omega_I \equiv \omega_\mu{}^p{}_q dx^\mu$ the spin connection. This has the property

$$\delta\omega_{3L} = d\mathrm{tr}(\Theta R_2)\,. \tag{12.2.24}$$

The combined Lorentz and Yang–Mills transformation law must then be

$$\delta A_I = d\lambda\,, \tag{12.2.25a}$$

$$\delta\omega_I = d\Theta\,, \tag{12.2.25b}$$

$$\delta B_2 = c\mathrm{Tr}(\lambda F_2) + c'\mathrm{tr}(\Theta R_2)\,. \tag{12.2.25c}$$

Again, we only indicate the leading, Abelian, terms. With this transformation the interaction (12.2.21) cancels an anomaly of the form

$$[c\mathrm{Tr}(F_2^2) + c'\mathrm{Tr}(R_2^2)]X_8(F_2,R_2)\,. \tag{12.2.26}$$

The gravitational Chern–Simons term was not included in the earlier low energy effective action because it is a higher derivative effect. The spin connection ω_I is proportional to the derivative of the tetrad, so the gravitational term in the field strength (12.2.22) contains three derivatives where the other terms contain one. However, its contribution is important in discussing the anomaly.

The chiral fields of $N = 1$ supergravity with gauge group g are the gravitino **56**, a neutral fermion **8′**, and an **8** gaugino in the adjoint representation, for total anomaly

$$\begin{aligned}\hat{I}_\mathrm{I} &= \hat{I}_{\mathbf{56}}(R_2) - \hat{I}_{\mathbf{8}}(R_2) + \hat{I}_{\mathbf{8}}(F_2,R_2)\\ &= \frac{1}{1440}\left\{-\mathrm{Tr}_a(F_2^6) + \frac{1}{48}\mathrm{Tr}_a(F_2^2)\mathrm{Tr}_a(F_2^4) - \frac{[\mathrm{Tr}_a(F_2^2)]^3}{14400}\right\}\\ &\qquad + (n-496)\frac{\mathrm{tr}(R_2^6)}{725760} - \frac{Y_4X_8}{384}\,.\end{aligned} \tag{12.2.27}$$

Here

$$Y_4 = \mathrm{tr}(R_2^2) - \mathrm{Tr}_a(F_2^2)\,, \tag{12.2.28a}$$

$$X_8 = \mathrm{tr}(R_2^4) + \frac{[\mathrm{tr}(R_2^2)]^2}{4} - \frac{\mathrm{Tr}_a(F_2^2)\mathrm{tr}(R_2^2)}{30} + \frac{\mathrm{Tr}_a(F_2^4)}{3} - \frac{[\mathrm{Tr}_a(F_2^2)]^2}{900}\,. \tag{12.2.28b}$$

The anomaly has been organized into a sum of three terms. The third is of the factorized form that can be canceled by the Green–Schwarz mechanism but the first two cannot, and so for the theory to be anomaly-free the combination of traces on the first line must vanish for the adjoint representation, and the total number of gauge generators must be 496. For the groups $SO(32)$ and $E_8 \times E_8$, both properties hold.[1] The net anomaly is then

$$\frac{Y_4X_8}{384}\,. \tag{12.2.29}$$

Of the various additional heterotic string theories constructed in the previous chapter, all but the diagonal theory are chiral, and in all cases the anomalies factorize.

In six-dimensional compactifications, some of which will be discussed in chapter 19, there can be multiple tensors. The Green–Schwarz mechanism can then cancel a sum of products Y_4X_4. Also, the same mechanism generalizes to forms of other rank; for example, a scalar in place of B_2 can cancel an anomaly Y_2X_d. For $d = 4$ this will arise in section 18.7.

Relation to string theory

From the low energy point of view, the cancellation of the anomaly involves several numerical accidents: the identity for the gauge traces, the correct number of generators, the factorized form (12.2.27). Again, these are explained by the existence of consistent string theories. In constructing new string theories, it is in principle not necessary to check the low

[1] They also hold for $E_8 \times U(1)^{248}$ and $U(1)^{496}$, but no corresponding string theories are known.

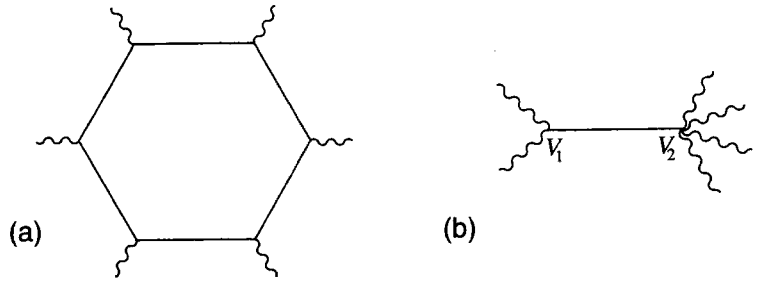


Fig. 12.1. Graphs contributing to the anomalies. One of the six external lines is a current and the others are gauge or gravitational fields: (a) hexagon graph; (b) canceling graph from exchange of $B_{\mu\nu}$ field.

energy anomaly, since this is guaranteed to vanish if the string consistency requirements have been satisfied. In practice, it is very useful as a check on the calculations and as a check that no subtle inconsistency has been overlooked.

In terms of Feynman graphs, the unphysical gauge and gravitational polarizations decouple by a cancellation between the two graphs of figure 12.1. The loop is the usual anomaly graph. The vertices of the tree graph come respectively from the H_3 kinetic term and the interaction (12.2.21). It is curious that a tree graph can cancel a loop, and it is interesting to look more closely at the coupling constant dependence. As discussed below eq. (12.1.11), in order to do the loop counting we need to write the R–R field as $C_2 = e^{-\Phi}C_2'$. Both vertices in figure 12.1(a) are then proportional to $e^{-\Phi}$ and so are 'half-loop' effects; they come from the disk amplitude. In the heterotic string no rescaling is needed. The vertex V_1 is proportional to $e^{-2\Phi}$ and so is a tree-level effect, while the vertex V_2 does not depend on the dilaton and so is actually a one-loop effect.

In each string theory, the hexagon loop and the tree graph arise from the same topology but different limits of moduli space. In the type I theory, the topology is the cylinder. The loop graph is from the short-cylinder limit and the tree graph from the long-cylinder limit. In the heterotic theory, the topology is the torus. The hexagon graph is from the limit $\tau_2 \to \infty$, while the tree graph is from the limit where two vertex operators approach one another.

In the heterotic string, the gauge group was determined by the requirement of modular invariance. In the type I string it was determined by cancellation of tadpole divergences. The relation with the field theory anomaly is as follows. One can prove the decoupling of null states formally in either field theory or string theory; the issue is whether terms from the UV limit invalidate the formal argument. In string theory these are the usual surface terms on moduli space. In the heterotic string the effective UV cutoff comes from the restriction of the integration to the fundamental region of moduli space. Surface terms from the boundary of the fundamental region cancel *if* the theory is modular-invariant. In the type I string the integration is not cut off but the 'UV' limit is reinterpreted as the IR limit of a closed string exchange, and the anomaly then vanishes if this converges.

12.3 Superspace and superfields

To formulate superstring perturbation theory it is useful to give superconformal symmetry a more geometric interpretation. To do this we need a *supermanifold*, a world-sheet with one ordinary complex coordinate z and one anticommuting complex coordinate θ, with

$$\theta^2 = \bar{\theta}^2 = \{\theta, \bar{\theta}\} = 0 \,. \tag{12.3.1}$$

What do we mean by anticommuting coordinates? Because of the anticommuting property, the Taylor series for any function of θ and $\bar{\theta}$ terminates. We can then think of any function on a supermanifold as the collection of ordinary functions appearing in the Taylor expansion. However, just as the operation '$\int d\theta$' has so many of the properties of ordinary integration that it is useful to call it integration, θ behaves so much like a coordinate that it is useful to think of a manifold with both ordinary and anticommuting coordinates.

We can think about ordinary conformal transformations as follows. Under a general change of world-sheet coordinates $z'(z, \bar{z})$ the derivative transforms as

$$\partial_z = \frac{\partial z'}{\partial z}\partial_{z'} + \frac{\partial \bar{z}'}{\partial z}\partial_{\bar{z}'} \,. \tag{12.3.2}$$

The conformal transformations are precisely those that take ∂_z into a multiple of itself.

Define the *superderivatives*,

$$D_\theta = \partial_\theta + \theta\partial_z \,, \quad D_{\bar{\theta}} = \partial_{\bar{\theta}} + \bar{\theta}\partial_{\bar{z}} \,, \tag{12.3.3}$$

which have the properties

$$D_\theta^2 = \partial_z \,, \quad D_{\bar{\theta}}^2 = \partial_{\bar{z}} \,, \quad \{D_\theta, D_{\bar{\theta}}\} = 0 \,. \tag{12.3.4}$$

A superconformal transformation $z'(z, \theta)$, $\theta'(z, \theta)$ is one that takes D_θ into a multiple of itself. From

$$D_\theta = D_\theta\theta'\partial_{\theta'} + D_\theta z'\partial_{z'} + D_\theta\bar{\theta}'\partial_{\bar{\theta}'} + D_\theta\bar{z}'\partial_{\bar{z}'} \,, \tag{12.3.5}$$

it follo

and so

Using

and th
genera
$f(z)$ ar

Infinite

with ϵ
confor
A $t\epsilon$

where
a conf
$\delta\theta = \epsilon$

$\delta\phi(z$

where
concer

Then

In ter

G_-

G_-

Either
transf
Ψ a

For the OPE, invariance under translations and rigid supersymmetry transformations implies that it is a function only of $z_1 - z_2 - \theta_1\theta_2$ and $\theta_1 - \theta_2$, and their conjugates. In this case,

$$X^\mu(z_1,\bar z_1)X^\nu(z_2,\bar z_2) \sim -\eta^{\mu\nu}\ln|z_1 - z_2 - \theta_1\theta_2|^2 , \tag{12.3.22}$$

as one can verify by expanding both sides in the anticommuting variables.

The superconformal ghost action is constructed from $(\lambda - \frac{1}{2}, 0)$ and $(1-\lambda, 0)$ tensor superfields B and C,

$$S_{BC} = \frac{1}{2\pi}\int d^2z\, d^2\theta\, BD_{\bar\theta}C . \tag{12.3.23}$$

The equation of motion is

$$D_{\bar\theta}B = D_{\bar\theta}C = 0 . \tag{12.3.24}$$

Acting on this equation with $D_{\bar\theta}$ gives $\partial_{\bar z}B = \partial_{\bar z}C = 0$, and so also $\partial_{\bar\theta}B = \partial_{\bar\theta}C = 0$. The equation of motion thus implies

$$B(z) = \beta(z) + \theta b(z) , \quad C(z) = c(z) + \theta\gamma(z) . \tag{12.3.25}$$

This is the same as the theory (10.1.17). The OPE is

$$B(z_1)C(z_2) \sim \frac{\theta_1 - \theta_2}{z_1 - z_2 - \theta_1\theta_2} = \frac{\theta_1 - \theta_2}{z_1 - z_2} . \tag{12.3.26}$$

The superfield form makes it easy to write down the nonlinear sigma model action

$$\begin{aligned} S &= \frac{1}{4\pi}\int d^2z\, d^2\theta\, [G_{\mu\nu}(X) + B_{\mu\nu}(X)]D_{\bar\theta}X^\nu D_\theta X^\mu \\ &= \frac{1}{4\pi}\int d^2z \Big\{ [G_{\mu\nu}(X) + B_{\mu\nu}(X)]\partial_z X^\mu \partial_{\bar z} X^\nu \\ &\qquad + G_{\mu\nu}(X)(\psi^\mu \mathscr{D}_{\bar z}\psi^\nu + \tilde\psi^\mu \mathscr{D}_z\tilde\psi^\nu) + \tfrac{1}{2}R_{\mu\nu\rho\sigma}(X)\psi^\mu\psi^\nu\tilde\psi^\rho\tilde\psi^\sigma \Big\} , \end{aligned} \tag{12.3.27}$$

after eliminating the auxiliary field. The Christoffel connection and antisymmetric tensor field strength combine in the covariant derivative,

$$\mathscr{D}_{\bar z}\psi^\nu = \partial_{\bar z}\psi^\nu + \Big[\Gamma^\nu{}_{\rho\sigma}(X) + \tfrac{1}{2}H^\nu{}_{\rho\sigma}(X)\Big]\partial_{\bar z}X^\rho\psi^\sigma , \tag{12.3.28a}$$

$$\mathscr{D}_z\tilde\psi^\nu = \partial_z\tilde\psi^\nu + \Big[\Gamma^\nu{}_{\rho\sigma}(X) - \tfrac{1}{2}H^\nu{}_{\rho\sigma}(X)\Big]\partial_z X^\rho\tilde\psi^\sigma . \tag{12.3.28b}$$

This describes a general NS–NS background in either type II string theory. R–R backgrounds are hard to describe in this framework because the superconformal transformations have branch cuts at the operators. The dilaton does not appear in the flat world-sheet action but does appear in the superconformal generators. The reader should beware of a common convention in the literature, $B_{\mu\nu}^{\text{here}} = 2B_{\mu\nu}^{\text{there}}$.

All the above applies to the heterotic string, using only $\bar\theta$ and not θ. One needs the superfields

$$X^\mu = X^\mu + i\bar\theta\tilde\psi^\mu , \tag{12.3.29a}$$

$$\lambda^A = \lambda^A + \bar\theta G^A . \tag{12.3.29b}$$

The field G^A is auxiliary. The nonlinear sigma model is

$$\begin{aligned} S &= \frac{1}{4\pi}\int d^2z\, d\bar\theta \Big\{ [G_{\mu\nu}(X) + B_{\mu\nu}(X)]\partial_z X^\mu D_{\bar\theta}X^\nu - \lambda^A \mathscr{D}_{\bar\theta}\lambda^A \Big\} \\ &= \frac{1}{4\pi}\int d^2z \Big\{ [G_{\mu\nu}(X) + B_{\mu\nu}(X)]\partial_z X^\mu \partial_{\bar z} X^\nu + G_{\mu\nu}(X)\tilde\psi^\mu \mathscr{D}_z\tilde\psi^\nu \\ &\qquad + \lambda^A \mathscr{D}_{\bar z}\lambda^A + \tfrac{1}{2}F^{AB}_{\rho\sigma}(X)\lambda^A\lambda^B\tilde\psi^\rho\tilde\psi^\sigma \Big\} , \end{aligned} \tag{12.3.30}$$

where $\mathscr{D}_z\tilde\psi^\nu$ is as above and

$$\mathscr{D}_{\bar\theta}\lambda^A = D_{\bar\theta}\lambda^A - iA^{AB}_\mu(X)\partial_{\bar\theta}X^\mu\lambda^B , \tag{12.3.31a}$$

$$\mathscr{D}_{\bar z}\lambda^A = \partial_{\bar z}\lambda^A - iA^{AB}_\mu(X)\partial_{\bar z}X^\mu\lambda^B . \tag{12.3.31b}$$

It is worth noting that the modified gauge transformation of the 2-form potential, which played an important role in the cancellation of spacetime anomalies, has a simple origin in terms of a world-sheet anomaly. A spacetime gauge transformation

$$\delta A^{AB}_\mu = D_\mu\chi^{AB} , \quad \delta\lambda^A = i\chi^{AB}\lambda^B \tag{12.3.32}$$

leaves the classical action invariant. However, this acts only on left-moving world-sheet fermions and so has an anomaly in the world-sheet path integral. We can use the result (12.2.3) with $\hat k_L = 1$, $\hat k_R = 0$, and

$$A^{AB}_{\bar z}(z,\bar z) = \frac{1}{2\pi}A^{AB}_\mu(X)\partial_{\bar z}X^\mu , \tag{12.3.33}$$

the factor of 2π correcting for the nonstandard normalization of the Noether current in CFT. Then after the addition of a counterterm,

$$\delta Z[A] = \frac{1}{8\pi}\int d^2z\, \mathrm{Tr}_{\mathrm{v}}[\chi(X)F_{\mu\nu}(X)]\partial_z X^\mu\partial_{\bar z}X^\nu . \tag{12.3.34}$$

This is precisely canceled if we also change the background,

$$\delta B_{\mu\nu} = \frac{1}{2}\mathrm{Tr}_{\mathrm{v}}(\chi F_{\mu\nu}) . \tag{12.3.35}$$

Comparing to the supergravity result (12.1.40) gives

$$\frac{\kappa_{10}^2}{g_{10}^2} = \frac{1}{2} \to \frac{\alpha'}{4} . \tag{12.3.36}$$

Noting that the left-hand side has units of L^2, we have restored α' by introducing one factor of $\alpha'/2$. This is the correct result for the relation

between gravitational and gauge couplings in the heterotic string. For future reference, let us note that if we study a vacuum with a nonzero dilaton, the physical couplings differ from the parameters in the action by an additional e^{Φ}, so that also

$$\frac{\kappa^2}{g_{\rm YM}^2} \equiv \frac{e^{2\Phi}\kappa_{10}^2}{e^{2\Phi}g_{10}^2} = \frac{\alpha'}{4} . \tag{12.3.37}$$

(We will discuss slightly differing conventions for the gauge coupling in chapter 18.)

Vertex operators

Recall that the bosonic string vertex operators came in two forms. The state–operator mapping gave them as $\check{c}c$ times a $(1,1)$ matter tensor. In the gauge-fixed Polyakov path integral this was the appropriate form for a vertex operator whose coordinate had been fixed. For an integrated vertex operator the $\check{c}c$ was omitted, replaced by a d^2z. The vertex operators of the superstring have a similar variety of forms, or *pictures*. We will derive this idea here by analogy to the bosonic string, and explain it in a more geometric way in section 12.5.

The state–operator mapping in chapter 10 gave the NS–NS vertex operators as

$$\delta(\gamma)\delta(\tilde{\gamma}) = e^{-\phi-\tilde{\phi}} \tag{12.3.38}$$

times a $(\frac{1}{2},\frac{1}{2})$ superconformal tensor. These are the analog of the fixed bosonic vertex operators. We have seen that the superconformal tensors are the lowest components of superfields, which do indeed correspond to the value of the superfield when θ and $\bar{\theta}$ are fixed at 0. Calling this tensor $\mathcal{O}$, eq. (12.3.15) gives the vertex operator integrated over θ and $\bar{\theta}$ as

$$\mathscr{V}^{0,0} = G_{-1/2}\tilde{G}_{-1/2}\cdot\mathcal{O} . \tag{12.3.39}$$

This operator appears without the $\delta(\gamma)\delta(\tilde{\gamma})$. The nonlinear sigma model action has just this form, the $d^2\theta$ integral of a $(\frac{1}{2},\frac{1}{2})$ superfield. It is conventional to label vertex operators by their ϕ and $\tilde{\phi}$ charges as here, so that an operator of charges $(q,\tilde{q})$ is said to be in the $(q,\tilde{q})$ *picture*. The θ-integrated operator (12.3.39) is in the (0,0) picture and the fixed operators

$$\mathscr{V}^{-1,-1} = e^{-\phi-\tilde{\phi}}\mathcal{O} \tag{12.3.40}$$

are in the $(-1,-1)$ picture. Of course, all of this extends to the open and heterotic cases with only one copy of the superconformal algebra, so we would have there the -1 and 0 pictures.

Let us consider as an example the massless states

$$\psi^{\mu}_{-1/2}\tilde{\psi}^{\nu}_{-1/2}|0;k\rangle_{\rm NS} , \tag{12.3.41}$$

with vertex operators

$$\mathscr{V}^{-1,-1} = g_c e^{-\phi-\tilde{\phi}}\psi^{\mu}\tilde{\psi}^{\nu}e^{ik\cdot X} . \tag{12.3.42}$$

The bosonic coordinates can be integrated or fixed independently of the fermionic ones, so for convenience we treated them as integrated. From

$$\begin{aligned} &G_{-1/2}\tilde{G}_{-1/2}\psi^{\mu}_{-1/2}\tilde{\psi}^{\nu}_{-1/2}|0;k\rangle_{\rm NS} \\ &= -(\alpha^{\mu}_{-1} + \alpha_0\cdot\psi_{-1/2}\psi^{\mu}_{-1/2})(\tilde{\alpha}^{\nu}_{-1} + \tilde{\alpha}_0\cdot\tilde{\psi}_{-1/2}\tilde{\psi}^{\nu}_{-1/2})|0;k\rangle_{\rm NS} , \end{aligned} \tag{12.3.43}$$

we obtain the integrated vertex operators

$$\mathscr{V}^{0,0} = -\frac{2g_c}{\alpha'}(i\partial_z X^{\mu} + \tfrac{1}{2}\alpha' k\cdot\psi\,\psi^{\mu})(i\partial_{\bar{z}}X^{\nu} + \tfrac{1}{2}\alpha' k\cdot\tilde{\psi}\,\tilde{\psi}^{\nu})e^{ik\cdot X} , \tag{12.3.44}$$

with α' again restored. Note the resemblance to the massless bosonic vertex operators, with additional fermionic terms. These additional terms correspond to the connection and curvature pieces in the nonlinear sigma models. For massless open string vectors,

$$\mathscr{V}^{-1} = g_o t^a\psi^{\mu}e^{ik\cdot X} , \tag{12.3.45a}$$

$$\mathscr{V}^{0} = g_o(2\alpha')^{-1/2}t^a(i\dot{X}^{\mu} + 2\alpha' k\cdot\psi\,\psi^{\mu})e^{ik\cdot X} , \tag{12.3.45b}$$

where t^a is the Chan–Paton factor. For heterotic string vectors,

$$\mathscr{V}^{-1} = g_c\hat{k}^{-1/2}j^a\tilde{\psi}^{\mu}e^{ik\cdot X} , \tag{12.3.46a}$$

$$\mathscr{V}^{0} = g_c(2/\alpha')^{1/2}\hat{k}^{-1/2}j^a(i\bar{\partial}X^{\mu} + \tfrac{1}{2}\alpha' k\cdot\tilde{\psi}\,\tilde{\psi}^{\mu})e^{ik\cdot X} . \tag{12.3.46b}$$

For convenient reference, we give the relations between the vertex operator normalizations and the various couplings in the low energy actions of section 12.1:

$$\text{type I:}\quad g_o = g_{\rm YM}(2\alpha')^{1/2} ; \quad g_{\rm YM} \equiv g_{10}e^{\Phi/2} , \tag{12.3.47a}$$

$$\text{heterotic:}\quad g_c = \frac{\kappa}{2\pi} = \frac{\alpha'^{1/2}g_{\rm YM}}{4\pi} ; \quad \kappa \equiv \kappa_{10}e^{\Phi} , \quad g_{\rm YM} \equiv g_{10}e^{\Phi} , \tag{12.3.47b}$$

$$\text{type I/II:}\quad g_c = \frac{\kappa}{2\pi} ; \quad \kappa \equiv \kappa_{10}e^{\Phi} . \tag{12.3.47c}$$

These can be obtained by comparing the calculations of the next section with the field theory amplitudes. Note that the amplitudes depend on the

background value of the dilaton in combination with the parameters κ_{10} and g_{10} from the action.

12.4 Tree-level amplitudes

It is now straightforward to guess the form of the tree-level amplitudes. In the next section we will justify this from a more geometric point of view.

We want the expectation value on the sphere or disk of the product of vertex operators with an appropriate number of bosonic and fermionic coordinates fixed. In the bosonic string it was necessary to fix three vertex operators on the sphere because of the existence of three c and three $\tilde{c}$ zero modes. There are two γ and two $\tilde{\gamma}$ zero modes on the sphere, namely $1, z$ and $1, \bar{z}$: these are holomorphic at infinity for a weight $-\frac{1}{2}$ field. We need this many factors of $\delta(\gamma)$ and $\delta(\tilde{\gamma})$, else the zero-mode integrals diverge. Thus we should fix the θ, $\bar{\theta}$ coordinates of two vertex operators. Similarly on the disk, we must fix the θ coordinates of two open string vertex operators.

We can also see this in the bosonized form. The anomaly in the ϕ current requires a total ϕ charge of -2 and a total $\tilde{\phi}$ charge of -2. Thus we need two vertex operators in the $(-1,-1)$ picture and the rest in the $(0,0)$ picture. For open strings on the disk (or heterotic strings on the sphere) we need two in the -1 picture and the rest in the 0 picture.

The R sector vertex operators have ϕ charge $-\frac{1}{2}$ from the ghost ground state (10.4.24). This is midway between the fixed and integrated pictures and does not have such a simple interpretation. Nevertheless, conservation of ϕ charge tells us that the sum of the ϕ charges must be -2. Thus for an amplitude with two fermions and any number of bosons we can use the pictures $-\frac{1}{2}$ for the fermions, -1 for one boson, and 0 for the rest. For four fermions and any number of bosons we can use the pictures $-\frac{1}{2}$ for the fermions and 0 for all the bosons. This is enough for all the cases we will treat in this section. To go to six or more fermions we clearly need to understand things better, as we will do in the next section.

Three-point amplitudes

Type I disk amplitudes: According to the discussion above, the type I three-boson amplitude is

$$\frac{1}{\alpha' g_o^2}\left\langle c\mathscr{V}_1^{-1}(x_1)c\mathscr{V}_2^{-1}(x_2)c\mathscr{V}_3^{0}(x_3)\right\rangle + (\mathscr{V}_1 \leftrightarrow \mathscr{V}_2)\,, \tag{12.4.1}$$

where we take $x_1 > x_2 > x_3$. The relevant expectation values for massless

amplitudes are

$$\langle c(x_1)c(x_2)c(x_3)\rangle = x_{12}x_{13}x_{23}\,, \tag{12.4.2a}$$

$$\left\langle e^{-\phi}(x_1)e^{-\phi}(x_2)\right\rangle = x_{12}^{-1}\,, \tag{12.4.2b}$$

$$\langle \psi^\mu(x_1)\psi^\nu(x_2)\rangle = \eta^{\mu\nu}x_{12}^{-1}\,, \tag{12.4.2c}$$

in the bc, $\beta\gamma$, and ψ CFTs, and

$$\left\langle \psi^\mu e^{ik_1\cdot X}(x_1)\,\psi^\nu e^{ik_2\cdot X}(x_2)\,(i\dot{X}^\rho + 2\alpha' k_3\cdot\psi\,\psi^\rho)e^{ik_3\cdot X}(x_3)\right\rangle$$
$$= 2i\alpha'(2\pi)^{10}\delta^{10}(\textstyle\sum_i k_i)\left(-\frac{\eta^{\mu\nu}k_1^\rho}{x_{12}x_{13}} - \frac{\eta^{\mu\nu}k_2^\rho}{x_{12}x_{23}} + \frac{\eta^{\mu\rho}k_3^\nu - \eta^{\nu\rho}k_3^\mu}{x_{13}x_{23}}\right) \tag{12.4.3}$$

in the combined $X\psi$ CFT. We have given the expectation value within each CFT a simple normalization and included an overall normalization factor $1/g_o^2\alpha'$, equal to the one (6.4.14) found in the bosonic theory. One can verify this normalization by a unitarity calculation as in the bosonic string, with the convention (12.4.1) that we sum separately over the reversed-cyclic orientation (which is always equal in this unoriented theory). That is, an n-particle amplitude is a sum of $(n-1)!$ orderings which are equal in pairs.

Combining these, using momentum conservation and transversality, and including the factor $g_o^3(2\alpha')^{-1/2}$ from the vertex operators, we obtain the type I three-gauge-boson amplitude

$$ig_{\rm YM}(2\pi)^{10}\delta^{10}(\textstyle\sum_i k_i)\,e_{1\mu}e_{2\nu}e_{3\rho}V^{\mu\nu\rho}\,{\rm Tr}_{\rm v}([t^{a_1},t^{a_2}]t^{a_3})\,, \tag{12.4.4}$$

where

$$V^{\mu\nu\rho} = \eta^{\mu\nu}k_{12}^\rho + \eta^{\nu\rho}k_{23}^\mu + \eta^{\rho\mu}k_{31}^\nu\,, \tag{12.4.5}$$

and $k_{ij} = k_i - k_j$. This is the ordinary Yang–Mills amplitude, with $g_{\rm YM}$ related to g_o as in eq. (12.3.47a) so as to agree with the definition in the low energy action. Unlike the bosonic open string amplitude (6.5.15) there is no k^3 term and so no F^3 term in the low energy effective action. Indeed, it is known that such a term is not allowed by the $d = 10$, $N = 1$ supersymmetry.

Now consider amplitudes with two fermions and a boson. The CFT amplitudes are

$$\left\langle e^{-\phi/2}(x_1)e^{-\phi/2}(x_2)e^{-\phi}(x_3)\right\rangle = x_{12}^{-1/4}x_{13}^{-1/2}x_{23}^{-1/2}\,, \tag{12.4.6a}$$

$$\langle \Theta_\alpha(x_1)\Theta_\beta(x_2)\rangle = x_{12}^{-5/4}C_{\alpha\beta}\,, \tag{12.4.6b}$$

$$\langle \Theta_\alpha(x_1)\Theta_\beta(x_2)\psi^\mu(x_3)\rangle = 2^{-1/2}(C\Gamma^\mu)_{\alpha\beta}\,x_{12}^{-3/4}x_{13}^{-1/2}x_{23}^{-1/2}\,. \tag{12.4.6c}$$

The ghost amplitude is a free-field calculation, and in principle the matter part can be done in this way as well using bosonization. However,

bosonization requires grouping the fermions in pairs and so spoils manifest Lorentz invariance. For explicit calculations it is often easier to use Lorentz and conformal invariance. The two-point amplitude (12.4.6b) is determined up to normalization by these symmetries. Note that $C_{\alpha\beta}$ is the charge conjugation matrix (section B.1), and that only for spinors of opposite chirality is it nonzero: in ten dimensions the product of like-chirality spinors does not include an invariant. The three-point amplitude (12.4.6c) is then deduced by using the OPE

$$\psi^\mu(x)\Theta_\alpha(0) = (2x)^{-1/2}\Theta_\beta(0)\Gamma^\mu_{\beta\alpha} + O(x^{1/2}) \tag{12.4.7}$$

to determine the x_3 dependence. This amplitude is nonvanishing only for spinors of like chirality. The gaugino-gaugino-gauge-boson amplitude, with respective polarizations $u_{1,2}$ and e^μ, is then[2]

$$ig_{\rm YM}(2\pi)^{10}\delta^{10}(\textstyle\sum_i k_i)\, e_\mu\, \bar u_1\Gamma^\mu u_2\, {\rm Tr}_{\rm v}([t^{a_1}, t^{a_2}]t^{a_3}) \,. \tag{12.4.8}$$

We have used $u_1^T C\Gamma^\mu u_2 = \bar u_1\Gamma^\mu u_2$, from the Majorana condition.[3]

Heterotic sphere amplitudes: The closed string three-point amplitudes are the products of open-string amplitudes. For the heterotic string we need the expectation values of two and three currents. The OPE gives

$$\left\langle j^a(z_1)j^b(z_2)\right\rangle = \frac{\hat k\delta^{ab}}{z_{12}^2} \tag{12.4.9a}$$

$$\left\langle j^a(z_1)j^b(z_2)j^c(z_3)\right\rangle = \frac{i\hat k f^{abc}}{z_{12}z_{13}z_{23}} \,, \tag{12.4.9b}$$

where the expectation value without insertions is normalized to unity. Each vertex operator thus needs a factor of $\hat k^{-1/2}$ to normalize the two-point function (as discussed in section 9.1). For the ten-dimensional heterotic string $k = 1$. In order to make contact with the discussion in the rest of this chapter, we will use the trace in the vector representation as the inner product, and then it follows from the discussion below eq. (11.5.13) that $\psi^2 = 1$ and $\hat k = \frac{1}{2}$. Including these factors, the normalization of the current algebra three-point function is

$$i\hat k^{-1/2} f^{abc} = 2^{1/2}{\rm Tr}_{\rm v}([t^a, t^b], t^c) \,. \tag{12.4.10}$$

The result can also be obtained from the free-fermion form $j^a = 2^{-1/2} i t^a_{AB}\lambda^A\lambda^B$, or from the free-boson form. Another necessary expectation value is

$$\left\langle \prod_{i=1}^3 ie_i\cdot\partial X e^{ik_i\cdot X}(z_i,\bar z_i)\right\rangle = \frac{\alpha'^2 e_{1\mu}e_{2\nu}e_{3\rho}T^{\mu\nu\rho}}{8iz_{12}z_{13}z_{23}} \,, \tag{12.4.11}$$

where

$$T^{\mu\nu\rho} = k_{23}^\mu\eta^{\nu\rho} + k_{31}^\nu\eta^{\rho\mu} + k_{12}^\rho\eta^{\mu\nu} + \frac{\alpha'}{8}k_{23}^\mu k_{31}^\nu k_{12}^\rho \,. \tag{12.4.12}$$

This is the same as for the bosonic string, section 6.6, where we have used the mass-shell condition $k_i^2 = 0$ and transversality $e_i\cdot k_i = 0$.

Now we can write all the massless three-point amplitudes. Including an overall factor $8\pi/\alpha' g_{\rm c}^2$ which is the same as in the bosonic string, the heterotic string three-gauge-boson amplitude is

$$4\pi g_{\rm c}\alpha'^{-1/2}(2\pi)^{10}\delta^{10}(\textstyle\sum_i k_i)e_{1\mu}e_{2\nu}e_{3\rho}V^{\mu\nu\rho}{\rm Tr}_{\rm v}([t^a,t^b],t^c) \,. \tag{12.4.13}$$

Up to the definition of the coupling this is the same as the open string amplitude (12.4.4). In particular there is no k^3 correction, again consistent with supersymmetry. Note that the vector part of this amplitude comes from the right-moving supersymmetric side. The heterotic amplitude for three massless neutral bosons (graviton, dilaton, or antisymmetric tensor) is

$$\pi i g_{\rm c}(2\pi)^{10}\delta^{10}(\textstyle\sum_i k_i)e_{1\mu\sigma}e_{2\nu\omega}e_{3\rho\lambda}T^{\mu\nu\rho}V^{\sigma\omega\lambda} \,. \tag{12.4.14}$$

One can relate the coupling $g_{\rm c}$ to the constants appearing in the heterotic string low energy action as in eq. (12.3.47b). In particular, the relation between $g_{\rm YM}$ and κ is in agreement with the anomaly result (12.3.37). The heterotic amplitude for two gauge bosons and one neutral boson is

$$\pi i g_{\rm c}(2\pi)^{10}\delta^{10}(\textstyle\sum_i k_i)\; e_{1\mu\nu}e_{2\rho}e_{3\sigma}k_{23}^\nu V^{\mu\rho\sigma}\delta^{ab} \,. \tag{12.4.15}$$

The antisymmetric part contains a Chern–Simons interaction, with $\omega_{3\,Y}$.

Type I/II sphere amplitudes: In any type I or II theory, the amplitude for three massless NS–NS bosons on the sphere is

$$\pi i g_{\rm c}(2\pi)^{10}\delta^{10}(\textstyle\sum_i k_i)e_{1\mu\sigma}e_{2\nu\omega}e_{3\rho\lambda}V^{\mu\nu\rho}V^{\sigma\omega\lambda} \,. \tag{12.4.16}$$

The normalization factor $8\pi/g_{\rm c}^2\alpha'$ and the relation $\kappa = 2\pi g_{\rm c}$ are the same as in other closed string theories.

The tensor structure is simpler than in the corresponding heterotic amplitude (12.4.14), with terms only of order k^2. The bosonic side of the heterotic string makes a more complicated contribution and the amplitude has terms of order k^2 and k^4. An R^2 correction to the action would give a three-point amplitude of order k^4, and an R^3 correction would give an amplitude of order k^6. Here 'R' is shorthand for the whole Riemann tensor, not just the Ricci scalar. The type I/II amplitude (12.4.16) implies

[2] In order that the gauge couplings of the gauge boson and gaugino agree — an indirect application of unitarity — we have normalized the fermion vertex operator as $g_{\rm o}\alpha'^{1/4}e^{-\phi/2}\Theta_\alpha e^{ik\cdot X}$.

[3] We are using standard field theory conventions, but to compare with much of the string literature one needs $g_{\rm YM}^{\rm here} = \frac{1}{2}g^{\rm there}$ and $u_i^{\rm here} = 2^{1/2}u_i^{\rm there}$.

no R^2 or R^3 corrections. In the heterotic string there is a correction of order R^2 but none of order R^3. The absence of R^2 and R^3 corrections in the type II theories is a consequence of the greater supersymmetry (32 generators rather than 16).

By taking two polarizations symmetric and one antisymmetric, there is in the heterotic string an order k^4 interaction of two gravitons and an antisymmetric tensor. An effective interaction built out of field strengths and curvatures would have *five* derivatives. The interaction we have found must therefore be the gravitational Chern–Simons interaction $H_3 \wedge *\omega_{3L}$, which figured in the heterotic anomaly cancellation. No such term was expected in the type II theories and none has appeared. We do need such a term in the type I theory, which has the same massless spectrum as the heterotic string and so needs the same Green–Schwarz cancellation. However, as explained at the end of section 12.2, in the type I theory this will come from the disk rather than the sphere. We can also understand this from the field redefinition (12.1.41). An R^2 interaction which is a tree-level heterotic effect maps

$$(-G_{\rm h})^{1/2} e^{-2\Phi_{\rm h}} R_{\rm h}^2 \to (-G_{\rm I})^{1/2} e^{-\Phi_{\rm I}} R_{\rm I}^2 \,, \tag{12.4.17}$$

which is the correct dilaton dependence for a disk or projective plane amplitude.

The various other three-point amplitudes are left as exercises.

Four-point amplitudes

All the four-point amplitudes of massless fields have been calculated. Many of the calculations are a bit tedious, though for supersymmetric strings the results tend to simplify. We will do a few simple calculations and quote some characteristic results, leaving the rest to the references.

Let us begin with the type I four-gaugino amplitude, each vertex operator being $g_{\rm o}(\alpha')^{1/4} t^a \mathscr{V}_\alpha e^{ik\cdot X} u_\alpha$. We need the expectation value of four $\mathscr{V}$s (of the same chirality). The OPE

$$\mathscr{V}_\alpha(z)\mathscr{V}_\beta(0) \sim \frac{(C\Gamma^\mu)_{\alpha\beta}}{2^{1/2} z} e^{-\phi} \psi_\mu \,, \tag{12.4.18}$$

follows from the three-point function (12.4.6c). Then

$$\begin{aligned} &\langle \mathscr{V}_\alpha(z_1)\mathscr{V}_\beta(z_2)\mathscr{V}_\gamma(z_3)\mathscr{V}_\gamma(z_4)\rangle \\ &= \frac{(C\Gamma^\mu)_{\alpha\beta}(C\Gamma_\mu)_{\gamma\delta}}{2z_{12}z_{23}z_{24}z_{34}} + \frac{(C\Gamma^\mu)_{\alpha\gamma}(C\Gamma_\mu)_{\delta\beta}}{2z_{13}z_{34}z_{32}z_{42}} + \frac{(C\Gamma^\mu)_{\alpha\delta}(C\Gamma_\mu)_{\beta\gamma}}{2z_{14}z_{42}z_{43}z_{23}} \,, \end{aligned} \tag{12.4.19}$$

from consideration of the singularities in z_1. An additional holomorphic term is forbidden because the expectation value (12.4.19) must fall as z_1^{-2} at infinity. Cancellation of the z_1^{-1} term further requires that

$$\Gamma^\mu_{\alpha\beta}\Gamma_{\mu\gamma\delta} + \Gamma^\mu_{\alpha\gamma}\Gamma_{\mu\delta\beta} + \Gamma^\mu_{\alpha\delta}\Gamma_{\mu\beta\gamma} = 0 \,. \tag{12.4.20}$$

This is indeed an identity, and plays an important role in ten-dimensional spacetime supersymmetry.

It is then straightforward to evaluate the rest of the amplitude. For the cyclic ordering 1234, let the vertex operators lie on the real axis and fix $x_1 = 0$, $x_3 = 1$, $x_4 \to \infty$ as usual to obtain

$$\begin{aligned} &\frac{i}{2} g_{\rm o}^2 (2\pi)^{10} \delta^{10}(\textstyle\sum_i k_i) {\rm Tr}_{\rm v}(t^{a_1}t^{a_4}t^{a_2}t^{a_3}) \int_0^1 dx_2\, x^{-\alpha' s - 1}(1-x)^{-\alpha' u - 1} \\ &\qquad \times (\bar u_1 \Gamma^\mu u_2\, \bar u_3 \Gamma^\mu u_4 + x\, \bar u_1 \Gamma^\mu u_3\, \bar u_2 \Gamma^\mu u_4) \,. \end{aligned} \tag{12.4.21}$$

Evaluating the integral and summing over cyclic orderings gives the final result

$$\begin{aligned} &-16 i g_{\rm YM}^2 \alpha'^2 (2\pi)^{10} \delta^{10}(\textstyle\sum_i k_i) K(u_1,u_2,u_3,u_4) \\ &\quad \times \left[{\rm Tr}_{\rm v}(t^{a_1}t^{a_2}t^{a_3}t^{a_4}) \frac{\Gamma(-\alpha' s)\Gamma(-\alpha' u)}{\Gamma(1-\alpha' s - \alpha' u)} + 2 \text{ permutations} \right] . \end{aligned} \tag{12.4.22}$$

The kinematic factor

$$K(u_1,u_2,u_3,u_4) = \frac{1}{8}(u\, \bar u_1 \Gamma^\mu u_2\, \bar u_3 \Gamma_\mu u_4 - s\, \bar u_1 \Gamma^\mu u_4\, \bar u_3 \Gamma_\mu u_2) \tag{12.4.23}$$

is fully antisymmetric in the spinors. We recall the definitions

$$s = -(k_1+k_2)^2 \,, \quad t = -(k_1+k_3)^2 \,, \quad u = -(k_1+k_4)^2 \,. \tag{12.4.24}$$

Replacing some of the gauginos with gauge bosons leads to the same form (12.4.22), with only the factor K altered. For four gauge bosons,

$$\begin{aligned} K(e_1,e_2,e_3,e_4) &= \frac{1}{8}\left(4 M^1_{\mu\nu} M^2_{\nu\sigma} M^3_{\sigma\rho} M^4_{\rho\mu} - M^1_{\mu\nu} M^2_{\nu\mu} M^3_{\sigma\rho} M^4_{\rho\sigma}\right) \\ &\qquad + 2 \text{ permutations} \\ &\equiv t^{\mu\nu\sigma\rho\alpha\beta\gamma\delta} k_{1\mu} e_{1\nu} k_{2\sigma} e_{2\rho} k_{3\alpha} e_{3\beta} k_{4\gamma} e_{4\delta} \,, \end{aligned} \tag{12.4.25}$$

where $M^i_{\mu\nu} = k_{i\mu} e_{i\nu} - e_{i\mu} k_{i\nu}$. The permutations replace the cyclic order 1234 with 1342 and 1423. The tensor t is antisymmetric within each $\mu_i \nu_i$ pair and symmetric under the interchange of two pairs, $\mu_i \nu_i$ with $\mu_j \nu_j$. This determines it to be a sum of the indicated two tensor structures. The result can also be written out

$$\begin{aligned} K(e_1,e_2,e_3,e_4) &= -\frac{1}{4}\left(s\, e_1\cdot e_4\, e_2\cdot e_3 + 2 \text{ permutations}\right) \\ &\quad + \frac{1}{2}\left(s\, e_1\cdot k_4\, e_3\cdot k_2\, e_2\cdot e_4 + 11 \text{ permutations}\right) . \end{aligned} \tag{12.4.26}$$

Each sum runs over all inequivalent terms obtained by permuting the four external lines.

It is interesting to consider the low energy limit of the bosonic amplitude. The expansion of the $\Gamma\Gamma/\Gamma$ factor begins

$$\frac{1}{\alpha'^2 su} - \frac{\pi^2}{6} + O(\alpha') , \tag{12.4.27}$$

the $O(\alpha'^{-1})$ term vanishing. We have used $\Gamma''(1) - \Gamma'(1)^2 = \zeta(2) = \pi^2/6$, where the zeta function is defined below. The leading term represents the Yang–Mills interaction in the low energy theory. Combined with the kinematic factor K it gives a sum of single poles, corresponding to exchange of massless gauge bosons, as well as the local quartic gauge interaction. The $O(\alpha'^0)$ terms correspond to a higher-derivative low energy interaction. To convert the scattering amplitude to a Lagrangian density replace $k_{[\mu}e_{\nu]} \cong -iF_{\mu\nu}/2g_{\rm YM}$ (so that the kinetic term has canonical normalization $\frac{1}{2}k^2 e^\mu e_\mu = \frac{1}{2}k^2$) and include a factor of $1/4!$ for the identical fields to obtain

$$\frac{\pi^2\alpha'^2}{2\times 4!\, g_{\rm YM}^2} t^{\mu\nu\sigma\rho\alpha\beta\gamma\delta} {\rm Tr}_{\rm v}(F_{\mu\nu}F_{\sigma\rho}F_{\alpha\beta}F_{\gamma\delta}) . \tag{12.4.28}$$

The net $g_{\rm YM}^{-2}$ is as expected for a tree-level string effect. The additional factor of α'^2 reflects the fact that this is a string correction to the low energy effective action, suppressed by the fourth power of the string length. The absence of an F^3 term is in agreement with the three-point amplitude.

The relation (6.6.23) between open and closed string tree amplitudes continues to hold in the superstring,

$$A_{\rm c}(s,t,u,\alpha',g_{\rm c}) = -\frac{\pi i g_{\rm c}^2\alpha'}{g_{\rm o}^4} A_{\rm o}(s,t,\tfrac{1}{4}\alpha',g_{\rm o}) A_{\rm o}(t,u,\tfrac{1}{4}\alpha',g_{\rm o})^* \sin\frac{\pi\alpha' t}{4} , \tag{12.4.29}$$

where the open string amplitudes represent just one of the six cyclic orderings, and the factors $(2\pi)^{10}\delta^{10}(\sum_i k_i)$ are omitted in $A_{\rm c,o}$. The type II amplitude with four massless NS–NS bosons is then

$$-\frac{i\kappa^2\alpha'^3}{4} \frac{\Gamma(-\frac{1}{4}\alpha' s)\Gamma(-\frac{1}{4}\alpha' t)\Gamma(-\frac{1}{4}\alpha' u)}{\Gamma(1+\frac{1}{4}\alpha' s)\Gamma(1+\frac{1}{4}\alpha' t)\Gamma(1+\frac{1}{4}\alpha' u)} K_{\rm c}(e_1,e_2,e_3,e_4) . \tag{12.4.30}$$

Here,

$$K_{\rm c}(e_1,e_2,e_3,e_4) = t^{\mu_1\nu_1\ldots\mu_4\nu_4} t^{\rho_1\sigma_1\ldots\rho_4\sigma_4} \prod_{j=1}^{4} e_{j\mu_j\rho_j} k_{j\nu_j} k_{j\sigma_j} . \tag{12.4.31}$$

The expansion of the ratio of gamma functions is

$$-\frac{64}{\alpha'^3 stu} - 2\zeta(3) + O(\alpha') \tag{12.4.32}$$

where the zeta function is

$$\zeta(k) = \sum_{m=1}^{\infty} \frac{1}{m^k} . \tag{12.4.33}$$

The first term is the low energy gravitational interaction; note that it is proportional to κ^2 with no α' dependence. From the normalization of the gravitational kinetic term, $e_{\mu\rho}k_\nu k_\sigma$ contracted into t becomes $R_{\mu\nu\sigma\rho}/4\kappa$; including a symmetry factor $1/4!$, the second term corresponds to an interaction

$$\frac{\zeta(3)\alpha'^3}{2^9\times 4!\,\kappa^2} t^{\mu_1\nu_1\ldots\mu_4\nu_4} t^{\rho_1\sigma_1\ldots\rho_4\sigma_4} R_{\mu_1\nu_1\rho_1\sigma_1} R_{\mu_2\nu_2\rho_2\sigma_2} R_{\mu_3\nu_3\rho_3\sigma_3} R_{\mu_4\nu_4\rho_4\sigma_4} . \tag{12.4.34}$$

This interaction, which is often identified by its distinctive coefficient $\zeta(3)$, has several interesting consequences; we will mention one in section 19.6. The absence of R^2 and R^3 corrections is again as expected from the three-point amplitude. For the heterotic string, the smaller supersymmetry allows more corrections.

We close with a few brief remarks about the heterotic amplitude with four gauginos or gauge bosons. The current algebra part of the amplitude is

$$\hat{k}^{-2}\langle j^{a_1}(z_1)j^{a_2}(z_2)j^{a_3}(z_3)j^{a_4}(z_4)\rangle = \frac{\delta^{a_1a_2}\delta^{a_3a_4}}{z_{12}^2 z_{34}^2} - \frac{f^{a_1a_2b}f^{ba_3a_4}}{\hat{k}z_{12}z_{23}z_{24}z_{34}} + (2\leftrightarrow 3) + (2\leftrightarrow 4) . \tag{12.4.35}$$

This is obtained by using the OPE to find the singularities in z_1. An additional holomorphic term is forbidden by the behavior at infinity. In fact, the $(1,0)$ current must fall off as z_1^{-2}, and the three terms of order z_1^{-1} do sum to zero by the Jacobi identity. Let us note further that $\delta^{a_1a_2} = {\rm Tr}_{\rm v}(t^{a_1}t^{a_2})$ and that

$$\begin{aligned} -\hat{k}^{-1} f^{a_1a_2b} f^{ba_3a_4} &= 2{\rm Tr}_{\rm v}([t^{a_1},t^{a_2}]t^b){\rm Tr}_{\rm v}(t^b[t^{a_3},t^{a_4}]) \\ &= 2{\rm Tr}_{\rm v}([t^{a_1},t^{a_2}][t^{a_3},t^{a_4}]) , \end{aligned} \tag{12.4.36}$$

where the last equality holds for $SO(32)$ (or for states in an $SO(16)\times SO(16)$ subgroup of $E_8\times E_8$) by completeness.

The remaining pieces of the amplitudes were obtained above, so it is straightforward to carry the calculation through. The amplitudes have the same factorized form (12.4.22) as in the type I theory, but with a more complicated group theory factor. In particular, the terms with two traces include effects from the exchange of massless supergravity states, which are of higher order in the type I theory.

All other three- and four-point massless amplitudes can be found in the references. We should mention that all of these were obtained first in the light-cone gauge, before the development of covariant methods. In fact,

while we have emphasized the covariant approach, for actual calculation the two methods are roughly comparable. The advantage of covariance is offset by the complication of the ghosts, and the realization of spacetime supersymmetry is more complicated.

12.5 General amplitudes

Pictures

Amplitudes should not depend on which vertex operators have their θ coordinates fixed. We demonstrate this in two different formalisms. The first, operator, method is particularly common in the older literature. The second leans more heavily on the BRST symmetry.

Let the two θ-fixed vertex operators also be $z,\bar{z}$-fixed, and use an $SL(2,\mathbf{C})$ transformation to bring them to 0 and ∞. In operator form, the amplitude becomes

$$\int d^2z_4 \dots d^2z_n \langle\!\langle \mathscr{V}_1^{-1} | \mathrm{T}[\mathscr{V}_3^0 \mathscr{V}_4^0 \dots \mathscr{V}_n^0] | \mathscr{V}_2^{-1} \rangle_{\text{matter}} . \tag{12.5.1}$$

We are working in the old covariant formalism, where the ghosts appear in a definite way. They then contribute only an overall factor to the amplitude, so we need only consider the matter part, as indicated. Then

$$|\mathscr{V}_2^{-1}\rangle = 2L_0^{\mathrm{m}} |\mathscr{V}_2^{-1}\rangle = \{G_{1/2}^{\mathrm{m}}, G_{-1/2}^{\mathrm{m}}\} |\mathscr{V}_2^{-1}\rangle = G_{1/2}^{\mathrm{m}} G_{-1/2}^{\mathrm{m}} |\mathscr{V}_2^{-1}\rangle , \tag{12.5.2}$$

using the physical state conditions. The $G_{-1/2}^{\mathrm{m}}$ converts $|\mathscr{V}_2^{-1}\rangle$ into $|\mathscr{V}_2^0\rangle$. The $G_{1/2}^{\mathrm{m}}$ can be moved to the left, the commutators making no contribution because of the superconformal invariance of the vertex operators, where it converts $\langle\!\langle \mathscr{V}_1^{-1}|$ to $\langle\!\langle \mathscr{V}_1^0|$. The final form

$$\int d^2z_4 \dots d^2z_n \langle\!\langle \mathscr{V}_1^0 | \mathrm{T}[\mathscr{V}_3^0 \mathscr{V}_4^0 \dots \mathscr{V}_n^0] | \mathscr{V}_2^0 \rangle_{\text{matter}} \tag{12.5.3}$$

has all matter vertex operators in the 0 picture.

The BRST argument starts by considering the *picture-changing operator (PCO)*

$$X(z) \equiv Q_{\mathrm{B}} \cdot \xi(z) = T_F(z)\delta(\beta(z)) - \partial b(z)\delta'(\beta(z)) , \tag{12.5.4}$$

where ξ is from bosonization of the superconformal ghosts. The calculation of $Q_{\mathrm{B}} \cdot \xi$ can be done in two ways. The first is to bosonize the BRST operator, expressing it in terms of ϕ, ξ, and η, calculate the OPE, and convert back. We will use a less direct but more instructive method. First, we claim that

$$\delta(\beta) \cong e^{\phi} . \tag{12.5.5}$$

The logic is exactly the same as that of $\delta(\gamma) \cong e^{-\phi}$. Now, it is generally true that

$$\gamma(z) f(\beta(0), \gamma(0)) \sim \frac{1}{z} \partial_\beta f(\beta(0), \gamma(0)) , \tag{12.5.6}$$

from all ways of contracting γ with a β in f. Now, we claim that the step function bosonizes as

$$\theta(\beta) \cong \xi . \tag{12.5.7}$$

Taking the OPE with $\gamma = e^{\phi}\eta$, this is consistent with the previous two equations, and this determines the left-hand side up to a function of γ alone; this function must vanish because both sides have a nonsingular product with $\beta = e^{-\phi}\partial\xi$. The explicit form (10.5.21) of the BRST current then gives

$$j_{\mathrm{B}}(z)\theta(\beta(0)) \sim -\frac{1}{z^2} b(z)\delta'(\beta(0)) + \frac{1}{z} T_F(z)\delta(\beta(0)) . \tag{12.5.8}$$

The two terms come from two or one $\gamma\beta$ contractions respectively. Integrating the current on a contour around the origin gives the result (12.5.4).

To understand the role of the PCO we need to examine an unusual feature of the $\beta\gamma$ bosonization. The (0,0) ξ field has one zero mode on the sphere, while the (1,0) η field has none. One factor of ξ is then needed to give a nonvanishing path integral. However, the only ghost factors in the vertex operators are $e^{-\phi}$ and $e^{-\phi/2}$. The correct rule is that the $\beta\gamma$ path integral is equal to the $\phi\eta\xi$ path integral with the various operators bosonized *and* with one additional $\xi(z)$ in the path integral. The position of the ξ insertion is irrelevant because the expectation value is proportional to the zero mode, which is constant. We can simply normalize

$$\langle \xi(z) \rangle = 1 . \tag{12.5.9}$$

To verify the decoupling of a null state we need to pull the BRST contour off the sphere. The ξ insertion would seem to be an obstruction, because the contour integral of the BRST charge around ξ is nonzero: it is just the definition (12.5.4) of the PCO. However, when the ξ insertion is replaced by X in this way, the path integral vanishes because of the ξ zero mode, and so there is no problem.

Now consider the path integral with one PCO and with the ξ insertion, as well as additional BRST-invariant operators. Then

$$X(z_1)\xi(z_2) = Q_{\mathrm{B}} \cdot \xi(z_1)\xi(z_2) = \xi(z_1) Q_{\mathrm{B}} \cdot \xi(z_2) = \xi(z_1) X(z_2) . \tag{12.5.10}$$

In the middle step we have pulled the BRST contour from $\xi(z_1)$ to $\xi(z_2)$ as in figure 12.2. There are two signs, from changing the order of Q_{B} and $\xi(z_1)$, and from changing the direction of the contour. Although $X(z)$ is formally null, its expectation value does not vanish because of the contour

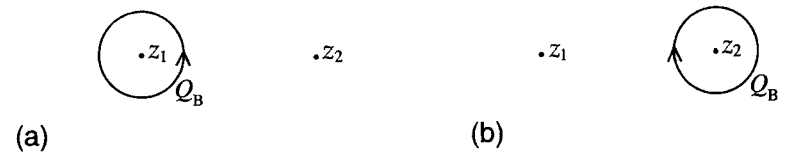


Fig. 12.2. Moving the PCO. The contour around z_1 in (a) is pulled around the sphere until it becomes a contour around z_2 in (b).

integral of Q_B around the ξ insertion. Unlike the same contribution in the previous paragraph, this does not vanish because the $\xi(z_1)$ remains to saturate the zero-mode integral. We already know that the path integral is independent of the position of the ξ insertion, so eq. (12.5.10) shows that it is also independent of the position of the PCO.

Consider now

$$\lim_{z\to 0} X(z)\mathscr{V}^{-1}(0)\,, \tag{12.5.11}$$

where for convenience we concentrate on the holomorphic side. The -1 picture vertex operator is $e^{-\phi}\mathcal{O}$ with $\mathcal{O}$ a matter superconformal primary. Consider now the term in $X(z)$ that involves the matter fields,

$$e^{\phi}T_F^{\rm m}(z)e^{-\phi}\mathcal{O}(0) = zT_F^{\rm m}(z)\mathcal{O}(0) + O(z^2)\,. \tag{12.5.12}$$

The $z\to 0$ limit picks out the coefficient of the z^{-1} in the matter OPE, which is precisely $G_{-1/2}\cdot\mathcal{O} = \mathscr{V}^0$, the 0 picture vertex operator. The purely ghost terms in X vanish as $z\to 0$, so that

$$\lim_{z\to 0} X(z)\mathscr{V}^{-1}(0) = \mathscr{V}^0(0)\,. \tag{12.5.13}$$

In the bosonic n-point amplitude with two -1 picture operators and $(n-2)$ 0 picture operators, we can pull a PCO out of each of the latter to be left with $(n-2)$ PCOs and n vertex operators, all of which are in the -1 picture. This is the 'natural' picture, the one given by the state–operator mapping. This also shows how to define a general tree-level amplitude, with n_B bosons and n_F (which must be even) fermions. Put all the bosons in the natural -1 picture, all the fermions in the natural $-\frac{1}{2}$ picture, and include $(n_B + \frac{1}{2}n_F - 2)$ PCOs. By taking some of the PCOs coincident with vertex operators, possibly more than one PCO at the same vertex operator, one obtains a representation with the vertex operators in higher pictures.

Finally, let us tie up a loose end. The operator product (12.4.18) is just the product of two spacetime supersymmetry currents, $\mathscr{V}_\alpha \equiv j_\alpha$. By the Ward identity and the supersymmetry algebra, we would expect the z^{-1} term to be the translation current. Instead it is $e^{-\phi}\psi^\mu$. However, this is the zero-momentum vector vertex operator in the -1 picture; if we move a PCO to the operator we get the 0 picture ∂X^μ which is indeed the translation current. So the algebra is correct. The $(1,0)$ operator $e^{-\phi}\psi^\mu$ is the translation current; which picture it appears in has no effect on the physics.

Super-Riemann surfaces

The preceding discussion suggests a natural generalization to all orders of perturbation theory. That is, string amplitudes are given by an integral over moduli space and the ghost plus matter path integral with the following insertions: the appropriate vertex operator for each incoming or outgoing string in the natural -1 or $-\frac{1}{2}$ picture, the b-ghosts for the measure on moduli space as in the bosonic string, plus the appropriate number of PCOs to give a sensible path integral. At genus g, the Riemann–Roch theorem gives the number of beta zero modes minus the number of gamma zero modes as $2g-2$. Equivalently, the total ϕ charge of the insertions must be $2g-2$. To obtain this, the total number of PCOs must be

$$n_X = 2g - 2 + n_B + \frac{n_F}{2}\,, \tag{12.5.14}$$

at arbitrary points; this is for the open string or one side of the closed string. The same formal arguments as in the case of the bosonic string show that this defines a consistent unitary theory. In particular, the PCOs are BRST-invariant and do not affect the decoupling of null states.

This prescription is sufficient for all the calculations we will carry out. However, in the remainder of this section we will develop superstring perturbation theory from a more general and geometric point of view. One reason for this is that the picture-changing prescription is rather *ad hoc* and it would be satisfying to see it derived in some way. Another is that this prescription actually has a subtle ambiguity at higher genus, which is best resolved from the more geometric point of view.

The needed idea is *supermoduli space*, the space of *super-Riemann surfaces (SRSs)*. These are defined by analogy to Riemann surfaces. Cover the surface with overlapping coordinate patches. The mth has coordinates z_m, θ_m. Patches are glued together with superconformal transformations. That is, if patches m and n overlap, identify points such that

$$z_m = f_{mn}(z_n) + \theta_n g_{mn}(z_n) h_{mn}(z_n)\,, \tag{12.5.15a}$$

$$\theta_m = g_{mn}(z_n) + \theta_n h_{mn}(z_n)\,, \tag{12.5.15b}$$

$$h_{mn}^2(z_n) = \partial_z f_{mn}(z_n) + g_{mn}(z_n)\partial_z g_{mn}(z_n)\,. \tag{12.5.15c}$$

The holomorphic functions f_{mn} and the anticommuting holomorphic functions g_{mn} define the SRS. Two SRSs are equivalent if there is a one-to-one mapping between them such that the respective coordinates are related by a superconformal transformation. Tensor fields are defined by analogy to tensors on an ordinary manifold, as functions in each patch with appropriate transformations between patches. Supermoduli space is the set of equivalence classes of super-Riemann surfaces. The coordinates on supermoduli space are the bosonic (even) moduli t_j and the anticommuting (odd) moduli v_a. The Riemann–Roch theorem gives the number of odd moduli minus the number of globally defined odd superconformal transformations as $2g-2$.

Again one can define all of this by Taylor expanding all functions in the anticommuting variables θ and v_a. The term in f_{mn} of order v_a^0 defines an ordinary (not super-) Riemann surface, and everything is expressed in terms of functions on this surface with the component form of the superconformal transformation between patches. Incidentally, z and $\bar{z}$ are no longer formally conjugates of one another on a SRS, particularly in the heterotic string where $\bar{z}$ transforms as the conjugate of eq. (12.5.15) while z transforms as on a 'bosonic' Riemann surface. However, if one defines everything by the Taylor expansion then z and $\bar{z}$ are again conjugates on the resulting ordinary Riemann surface.

For any SRS, setting the v_a to zero makes the anticommuting g_{mn} vanish and leaves

$$z_m = f_{mn}(z_n)\,, \tag{12.5.16a}$$

$$\theta_m = \theta_n h_{mn}(z_n)\,, \quad h_{mn}^2(z_n) = \partial_z f_{mn}(z_n)\,. \tag{12.5.16b}$$

The transformation of z defines a Riemann surface, but that of θ requires the additional choice of which square root to take in each h_{mn}. This choice is known as a *spin structure;* it is the same data one would need to put a spin-$\frac{1}{2}$ field on the surface. The signs are not all independent. If three patches overlap then the transition functions must satisfy the *cocycle condition*

$$h_{mn}h_{np}h_{pm} = 1\,. \tag{12.5.17}$$

Also, a coordinate change $\theta_p \to -\theta_p$ in the patch p_0 changes the signs of all the h_{pn}. The net result is that there is one meaningful sign for each nontrivial closed path on the surface, $2g$ for a genus g surface. These define 2^{2g} different spin structures, topologically distinct ways to put a spinor field on the surface.

Any sphere is equivalent to the one with two patches (z,θ), (u,ϕ) and transition functions

$$u = 1/z\,, \quad \phi = i\theta/z\,. \tag{12.5.18}$$

Clearly there are no spin structures. The index theorem implies superconformal Killing transformations. One can look for infinitesimal transformations as in the bosonic case, with the result that δf must be at most quadratic in z and δg linear in z. The general finite transformation is then of the superconformal form with

$$f(z) = \frac{\alpha z+\beta}{\gamma z+\delta}\,, \quad g(z) = \epsilon_1 + \epsilon_2 z \tag{12.5.19}$$

with $\alpha\delta-\beta\gamma=1$. In particular there are two odd transformations, ϵ_1 and ϵ_2, consistent with the Riemann–Roch theorem. These can be used to fix the odd coordinates of two NS vertex operators to zero.

A torus can be described as the (z,θ) plane modded by a group of rigid superconformal transformations,

$$(z,\theta) \cong (z+2\pi,\eta_1\theta) \cong (z+2\pi\tau,\eta_2\theta)\,. \tag{12.5.20}$$

The η_1 and η_2 are each ±1, defining the four spin structures. When θ changes sign around a loop, the bosonic and fermionic components of any superfield will have opposite periodicities, and in particular T_F will be antiperiodic. We thus denote the spin structures (P,P), (P,A), (A,P), and (A,A), giving the $z \to z+2\pi$ periodicity first. The periodicities on the right-moving side have the same form, with $\bar{\tau}$ the conjugate of τ but with independent $\tilde{\eta}_1$ and $\tilde{\eta}_2$.

On a torus the only holomorphic functions are the constants, so β and γ zero modes are possible only in the (P,P) case, in which case there is one of each. There is then an odd supermodulus v, giving rise to the more general periodicity[4]

$$(z,\theta) \cong (z+2\pi,\theta) \cong (z+2\pi\tau+\theta v,\theta+v)\,. \tag{12.5.21}$$

There is also the *superconformal Killing vector (SCKV)*

$$(z,\theta) \to (z+\theta\epsilon,\theta+\epsilon)\,. \tag{12.5.22}$$

The number of odd moduli minus the number of SCKVs is zero in all sectors, being $1-1$ for the (P,P) spin structure and $0-0$ for the others. The modular group and the fundamental region for τ are the same as in the bosonic string.

Returning to a general SRS, if the positions of n vertex operators are singled out then there is an additional nontrivial closed curve circling each, giving 2^{2g+n} spin structures altogether. The additional spin structures come from the choice of R or NS boundary conditions of the external

[4] We could introduce a second odd parameter into the $z+2\pi$ periodicity, but one of the two parameters can be removed by a linear redefinition of (z,θ). Also, it might appear that a similar generalization is possible in the antiperiodic case, but a coordinate redefinition returns the periodicity to the form (12.5.20).

strings. To describe the supermoduli space of SRSs with n_B NS vertex operators and n_F R vertex operators, it is useful to extend the approach used in section 5.4. First in the bosonic case, consider a specific patching together of a Riemann surface, with n marked points. We will define another Riemann surface as equivalent to this one if there is a one-to-one holomorphic mapping between the two which leaves the coordinates of the points invariant. That is, $f(z) - z$ must vanish linearly at the vertex operators. For simplicity we take each operator to be at $z = 0$ in its own tiny patch. Since we are modding by a smaller group, with two real conditions for each vertex operator, we obtain a correspondingly larger coset space, with two additional moduli for each vertex operator. This is similar to the treatment of vertex operator positions in section 5.4, but more abstract. In the superconformal case, we mod out the superconformal transformations for which $f(z) - z$ and $g(z)$ vanish linearly at each NS vertex operator. At each R vertex operator, $g(z)$ has a branch cut, and so it is appropriate to require $f(z) - z$ to vanish linearly z and $g(z)$ to vanish as $z^{1/2}$. The NS vertex reduces the odd coordinate degrees of freedom by one and so increases the number of inequivalent surfaces: the number of odd moduli increases by one, which we can take to be the θ coordinate of the operator. The condition for the R vertex operator is essentially half as restrictive, so that there is an additional odd modulus for each *pair* of R vertex operators. This has no simple interpretation as a vertex operator position; an R vertex operator produces a branch cut in θ, so there can be no well-defined θ coordinate for the operator. The total number of odd moduli is

$$n_\nu = 2g - 2 + n_B + \frac{n_F}{2} \,. \tag{12.5.23}$$

The measure on supermoduli space

The expression (5.4.19) for the bosonic string S-matrix now generalizes in a natural way,

$$S(1;\ldots;n) = \sum_{\chi,\gamma} \frac{e^{-\lambda\chi}}{n_R} \int_{\chi,\gamma} d^{n_e}t\, d^{n_o}\nu \left\langle \prod_{j=1}^{n_e} B_j \prod_{a=1}^{n_o} \delta(B_a) \prod_{i=1}^{n} \hat{\mathscr{V}}_i \right\rangle \,. \tag{12.5.24}$$

The sum is over topologies χ and spin structures γ. The integral runs over the corresponding supermoduli space. There are n_e even moduli, n_o odd moduli, and n external strings. The quantity B_j in the ghost insertions is

$$B_j = \sum_{(mn)} \int_{C_{mn}} \frac{dz_m d\theta_m}{2\pi i} B(z,\theta) \left[\frac{\partial z_m}{\partial t_j} + \frac{\partial \theta_m}{\partial t_j}\theta_m \right]_{z_n,\theta_n} \,, \tag{12.5.25}$$

plus a right-moving piece of the same form; B_a is given by an identical expression with ν_a replacing t_j. The sum again runs over all pairs of

overlapping patches m and n, clockwise as seen from m, with the z_m integration along a contour between the two patches and the θ_m integral of the usual Berezin form. The B ghost superfield is as in eq. (12.3.25), $B = \beta + \theta b$.

The logic of this expression is the same as for the earlier bosonic expression. First, the number of commuting and anticommuting ghost insertions is correct for a well-defined path integral. Second, the path integral depends only on the superconformal structure and not on the particular choice of patches and transition functions. In particular it is unchanged if we make a superconformal transformation within a single coordinate patch. The combination $\partial z_m + (\partial\theta_m)\theta_m$ transforms as a $(-1,0)$ tensor superfield, so the integrand is a $(\frac{1}{2},0)$ tensor superfield and the integral is invariant. Third, under a change of coordinates in supermoduli space, the product $\prod_{j=1} B_j \prod_a \delta(B_a)$ transforms as a density, inversely to the measure on supermoduli space. Finally, the commutator of the BRST charge with $B_{j,a}$ is $T_{j,a}$, defined in the same way but with B replaced by

$$Q_B \cdot B(z) = T(z) = T_F(z) + \theta T_B(z) \,. \tag{12.5.26}$$

The insertion of $T_{j,a}$ generates a relative coordinate transformation of adjacent patches, which is just the derivative with respect to the supermodulus of the world-sheet.

It is interesting to work out the form of the amplitude more explicitly for a special choice of patches and transition functions. Namely, let patch 1 be contained entirely within patch 2, so that the overlap is an annulus. Let the 1-2 transition functions depend only on a single odd modulus ν, as follows:

$$f_{12}(z_2) = z_2 \,, \quad g_{12}(z_2) = \nu\alpha(z_2) \,, \tag{12.5.27}$$

for some holomorphic function $\alpha(z)$. The ghost factor (12.5.25) reduces to

$$B[\alpha] = \oint \frac{dz_1}{2\pi i} \alpha(z_1)\beta(z,\theta) \,. \tag{12.5.28}$$

Similarly the path integral depends on ν only through the insertion

$$\nu T[\alpha] \,, \tag{12.5.29}$$

where β is replaced by T_F. We can then perform the integration over ν, so that the net effect of the supermodulus is the insertion in the path integral of

$$T[\alpha]\delta(B[\alpha]) = Q_B \cdot \theta(B[\alpha]) \,. \tag{12.5.30}$$

The function $\alpha(z_1)$ is holomorphic in the annular overlap of the patches,

but in general cannot be extended holomorphically into the full inner patch z_1. If it can, the contour integrals $B[\alpha]$ and $T[\alpha]$ vanish. In this case v is not a modulus at all because it can be transformed away. A nontrivial case is

$$\alpha(z_1) = \frac{1}{z_1 - z_0}\,, \tag{12.5.31}$$

for which $B[\alpha] = \beta(z_0)$ and the insertion (12.5.30) just becomes the PCO $X(z_0)$ (the second term in X is from normal ordering). Thus, the PCO is the result of integrating out an odd modulus in this special parameterization of the SRSs. Note that the number (12.5.23) of odd moduli is the same as the number (12.5.14) of PCOs needed in the *ad hoc* approach. This provides the desired geometric derivation of the picture-changing prescription.

The parameterization (12.5.27) is always possible locally on supermoduli space. It can also be used globally, with careful treatment of the modular identification and the limits of moduli space. There is a literature on the 'ambiguity of superstring perturbation theory,' which arose from parameterizations that did not precisely cover supermoduli space. It appears that superstring perturbation theory to arbitrary order is understood in principle, and certain special amplitudes have been calculated at higher orders of perturbation theory. However, the subject is somewhat unfinished — a fully explicit proof of the perturbative consistency of the theory seems to be lacking. With the immense progress in nonperturbative string theory, filling this technical gap does not seem to be a key issue.

We derived the bosonic version (5.4.19) of the measure (12.5.24) by starting with a path integral over the world-sheet metric, whereas in the present case we have written it down directly. One can partly work backwards to an analogous description as follows. Although $\alpha(z_1)$ cannot be extended holomorphically into patch 1 it can be extended smoothly. It can then be removed by a change of variables in the path integral, but not one that leaves the action invariant. The odd modulus v appears in the final action, multiplying T_F and a function that can be regarded as the world-sheet gravitino field. In particular, the PCO can be regarded as coming from a pointlike gravitino, a gauge where the gravitino has delta-function support.

12.6 One-loop amplitudes

We will illustrate one-loop superstring calculations with two examples where the low energy limit can be obtained in closed form.

The first is the heterotic string amplitude with four gauge bosons and one antisymmetric tensor. The Green–Schwarz anomaly cancellation

requires a one-loop Chern–Simons term

$$\int B_2 \mathrm{Tr}_{\mathrm{v}}(F_2^4)\,. \tag{12.6.1}$$

We would like to confirm the appearance of this term by an explicit string calculation.

Note first that this can only arise from the (P,P) path integral. This is because it is odd under spacetime parity: written out in components, it involves the ten-dimensional ϵ-tensor. The heterotic string world-sheet action and constraints are invariant under parity. The parity asymmetry of the theory, the fact that the massless fermions are in a **16** and not a **16′**, comes about from the GSO projection in the right-moving R sector, the choice of $\exp(\pi i F)$ to be $+1$ or -1. The (P,P) path integral produces this term in the projection operator. The path integral is then

$$\left(\frac{2}{\alpha'}\right)^{5/2} g_c^5 \int_F \frac{d\tau d\bar\tau}{8\tau_2}\left[\prod_{i=1}^{5}\int d^2w_i\right]\Big\langle b(0)\tilde b(0)\tilde c(0)c(0)X(0)$$
$$\times\left[\prod_{i=1}^{4}\hat k^{-1/2} j^{a_i}(ie_i\cdot\bar\partial X + \tfrac{1}{2}\alpha' k_i\cdot\tilde\psi\, e_i\cdot\tilde\psi)e^{ik_i\cdot X}(w_i,\bar w_i)\right]$$
$$\times ie_{5\mu\nu}\partial X^\mu\delta(\tilde\gamma)\tilde\psi^\nu e^{ik_5\cdot X}(w_5,\bar w_5)\Big\rangle_{(\mathrm{P,P})} \tag{12.6.2}$$

The bc ghosts and corresponding measure are the same as in the bosonic string, with an extra $\frac{1}{2}$ from the GSO projection operator. For the (P,P) spin structure there is one PCO and one -1 picture vertex operator.

We will consider in order the $\tilde\psi^\mu$, X^μ, bc, $\beta\gamma$, and j^a path integrals. In the vacuum amplitude the $\tilde\psi^\mu$ path integral vanishes in the (P,P) sector. In terms of a trace, this is due to a cancellation between the R sector ground states. In terms of a path integral it is due to the Berezin integration over the zero mode of ψ^μ (which exists only for this spin structure). In the latter form it is clear that we need at least ten factors of $\tilde\psi$ to obtain a nonzero path integral. In fact, the path integral (12.6.2) has a maximum of ten $\tilde\psi$s, including one from the term

$$\delta(\tilde\beta)i(2/\alpha')^{1/2}\tilde\psi^\rho\bar\partial X_\rho \tag{12.6.3}$$

in the PCO. The relevant path integral is easily obtained from a trace, giving

$$\left\langle\prod_{i=1}^{10}\tilde\psi^{\mu_i}\right\rangle_{\tilde\psi(\mathrm{P,P})} = \epsilon^{\mu_1\cdots\mu_{10}}\bar q^{10/24}\prod_{n=1}^{\infty}(1-\bar q^n)^{10}$$
$$= \epsilon^{\mu_1\cdots\mu_{10}}[\eta(\tau)^{10}]^*\,. \tag{12.6.4}$$

The X^μ path integral is then reduced to

$$\left\langle \partial X^\mu(w_5)\bar\partial X^\rho(0)\prod_{i=1}^{5} e^{ik_i\cdot X(w_i,\bar w_i)}\right\rangle_X , \tag{12.6.5}$$

the gradients coming from the tensor vertex operator and the PCO. To make things simple we now take the $k_i \to 0$ limit. Contractions between the gradients and exponentials, and among the exponentials, are then suppressed. Only the contraction between the gradients survives, $-\alpha'/8\pi\tau_2$ from the background charge term $\alpha'(\mathrm{Im}\, w_{ij})^2/4\pi\tau_2$ in the Green's function (7.2.3). The leading term in the expectation value (12.6.5) is then

$$-i(2\pi)^{10}\delta^{10}(\textstyle\sum_i k_i)\dfrac{\eta^{\mu\rho}\alpha'}{8\pi\tau_2(4\pi^2\alpha'\tau_2)^5|\eta(\tau)|^{20}} . \tag{12.6.6}$$

The bc path integral is

$$\left\langle b(0)\tilde b(0)\tilde c(0)c(0)\right\rangle_{bc} = |\eta(\tau)|^4 , \tag{12.6.7}$$

just as in the bosonic string. The $\beta\gamma$ path integral is the reciprocal of the right-moving part of this,

$$\left\langle \delta(\tilde\beta(0))\delta(\tilde\gamma(\bar w_5))\right\rangle_{\beta\gamma} = [\eta(\tau)^{-2}]^* . \tag{12.6.8}$$

Finally for the current algebra, we need

$$\hat k^{-2}\left\langle j^{a_1}(w_1)j^{a_2}(w_2)j^{a_3}(w_3)j^{a_4}(w_4)\right\rangle_g . \tag{12.6.9}$$

We continue to use the convention $\hat k = \frac{1}{2}$ for the rest of the chapter. Note first that all other expectation values are independent of w_i. The integrations over w_i thus have the effect of averaging over $\mathrm{Re}(w_i)$ and so we can replace each current with the corresponding charge, Q^{a_i}. We can then evaluate the expectation value as a trace. However, a careful treatment of the $k \to 0$ limit shows that an additional contact term is needed when two vertex operators coincide,

$$j^a(w)j^b(0) \to \mathrm{T}\left[\hat Q^a(w)\hat Q^b(0)\right] - \pi\delta^2(w,\bar w)\delta^{ab} . \tag{12.6.10}$$

To see this, integrate both sides over the region of world-sheet $-\delta < \sigma^2 < \delta$. On the left we have

$$\frac{\delta^{ab}}{2}\int_{|\sigma^2|<\delta} d^2w\,\frac{1}{w^2}(w\bar w)^{k\cdot k'} , \tag{12.6.11}$$

where we have introduced small k and k'. The $(w\bar w)^{k\cdot k'}$ factor from the X^μ path integral then regulates the integral at the origin. We have kept the leading term in the OPE because this is the only one that contributes at small δ. Writing the integrand as

$$(-1+k\cdot k')^{-1}\partial_w(w^{-1+k\cdot k'}\bar w^{k\cdot k'}) , \tag{12.6.12}$$

we can integrate by parts to convert this to a surface integral, which is then easily evaluated at $k,k' \to 0$ to give -2π. On the right, the first term has no singularity that would allow a nonzero limit as $\delta \to 0$ ($\hat Q^a$ is conserved, so it is a constant except for the time ordering), and so the delta function is needed.

For the product of two currents we would then have

$$\begin{aligned}\langle j^{a_1}(w_1)j^{a_2}(w_2)\rangle &\to \mathrm{Tr}\left\{\exp(2\pi i\tau H)\mathrm{T}\left[\hat Q^{a_1}(w_1)\hat Q^{a_2}(w_2)\right]\right\} \\ &\quad - \delta^{ab}\pi\delta^2(w_{12},\bar w_{12}) \\ &\to \mathrm{Tr}\left\{\exp(2\pi i\tau H)\hat Q^{(a_1}\hat Q^{a_2)}\right\} - \frac{\delta^{ab}}{8\pi\tau_2} .\end{aligned} \tag{12.6.13}$$

In the second line we have used the fact that all other expectation values are independent of the w_i, so that the integrations will have the effect of averaging over w_i. In the first term this symmetrizes the operators as indicated; in the second it allows us to replace the delta function with its average over the torus. For four currents the combinatorics are conveniently summarized in terms of the generating function

$$\begin{aligned}f(q,z) &\equiv \langle \exp(z\cdot\bar j)\rangle \\ &= \exp\left(-\frac{z\cdot z}{16\pi\tau_2}\right)\mathrm{Tr}\left[\exp(2\pi i\tau H)\exp(z\cdot\hat Q)\right] ,\end{aligned} \tag{12.6.14}$$

where $\overline{j^a}$ is the average over the torus and the dot denotes a sum on a. The needed expectation value is the fourth derivative with respect to z^a. The trace is most easily carried out in the bosonic form, where it becomes an oscillator sum plus sum over the $SO(32)$ or $E_8\times E_8$ lattice:

$$f(q,z) = \eta(\tau)^{-16}\exp\left(-\frac{z\cdot z}{16\pi\tau_2}\right)\sum_{l\in\Gamma} q^{l^2/2}\exp(2^{-1/2}z\cdot l) . \tag{12.6.15}$$

Gathering all factors, including $(8\pi\tau_2)^5$ from integrating over the w_i, the amplitude becomes

$$\begin{aligned}&-\frac{ig_c^5}{\pi\alpha'^3}(2\pi)^{10}\delta^{10}(\textstyle\sum_i k_i)\epsilon^{\mu_1\ldots\mu_{10}}k_{1\mu_1}e_{1\mu_2}\ldots k_{4\mu_7}e_{4\mu_8}e_{5\mu_9\mu_{10}} \\ &\qquad\times\int_F\frac{d^2\tau}{\tau_2^2}\left.\frac{\partial^4\hat f(q,z)}{\partial z^{a_1}\cdots\partial z^{a_4}}\right|_{z=0} ,\end{aligned} \tag{12.6.16}$$

where

$$\hat f(q,z) = \eta(\tau)^{-8}f(q,z) . \tag{12.6.17}$$

We leave it as an exercise to show that this is modular-invariant. Substituting $e_{\mu\nu} \to B_{\mu\nu}/2\kappa$ and $k_{[\mu}e_{\nu]} \to -iF_{\mu\nu}/2g_{\mathrm{YM}}$ from the normalization of

the kinetic terms, including a factor 2^5 from converting the ϵ-tensor to form notation, and using the relation between the couplings, the amplitude can be concisely summarized by the effective action

$$-\frac{1}{2^9\pi^6\alpha'}\int B_2\int_F\frac{d^2\tau}{\tau_2^2}\hat{f}(q,F_2)\,. \tag{12.6.18}$$

The form integration picks out the term of order F_2^4. The integral can be written as a surface term and given in closed form, using

$$\frac{\hat{f}(q,F_2)}{\tau_2^2}=-\frac{32\pi i}{F_2\cdot F_2}\frac{\partial\hat{f}(q,F_2)}{\partial\bar{\tau}}\,. \tag{12.6.19}$$

Due to modular invariance, only the limit $\tau_2\to\infty$ contributes. The effective action becomes

$$-\frac{1}{2^4\pi^5\alpha'}\int\frac{B_2}{F_2\cdot F_2}\hat{f}(q,F_2)\Big|_{q^0\text{ term}}\,. \tag{12.6.20}$$

Only the lattice momenta with $l^2=2$ contribute to the q^0 term. These form the adjoint of the gauge group, so the lattice sum reduces to a trace in the adjoint representation and the effective interaction is

$$-\frac{1}{2^4\pi^5 6!\,\alpha'}\int\frac{B_2\mathrm{Tr}_a(F_2^6)}{F_2\cdot F_2}=-\frac{1}{2^4\pi^5 4!\,\alpha'}\int B_2\frac{\mathrm{Tr}_a(F_2^6)}{\mathrm{Tr}_a(F_2^2)}\,. \tag{12.6.21}$$

Ordinarily dividing by a form would make no sense, but we know from the discussion of anomalies that $\mathrm{Tr}_a(F_2^6)\propto F_2\cdot F_2\,X_8(F_2)$, so the effective interaction is proportional to $\int B_2X_8$ as required by anomaly cancellation (and with the correct coefficient). For $SO(32)$ the ratio of forms is $\frac{1}{2}\mathrm{Tr}_v(F_2^4)$; for $E_8\times E_8$ it is

$$\frac{1}{7200}\left\{[\mathrm{Tr}_{a1}(F_2^2)]^2+[\mathrm{Tr}_{a2}(F_2^2)]^2-\mathrm{Tr}_{a1}(F_2^2)\mathrm{Tr}_{a2}(F_2^2)\right\}\,. \tag{12.6.22}$$

With somewhat more effort one can also find the required curvature terms. That we were able with modest effort to bring this string loop amplitude to a closed form is not too surprising, since this is a very special amplitude whose coefficient is determined by symmetry (anomaly cancellation). However, many of the physically interesting corrections to the low energy effective action can be obtained in a closed form.

Next we consider the heterotic string amplitude with four gauge bosons but without the antisymmetric tensor. In contrast to the previous amplitude, which came only from the (P,P) spin structure, the present one comes only from the other three spin structures: with one fewer vertex operator there are not enough insertions of $\tilde{\psi}$ to saturate the zero modes of the

(P,P) path integral. The amplitude is then

$$\frac{4g_c^4}{\alpha'^2}\int_F\frac{d\tau d\bar{\tau}}{8\tau_2}\left[\prod_{i=1}^4\int d^2w_i\right]\sum_{\gamma\neq(\mathrm{P,P})}\Big\langle b(0)\tilde{b}(0)\tilde{c}(0)c(0)$$
$$\times\left[\prod_{i=1}^4\hat{k}^{-1/2}j^{a_i}(ie_i\cdot\bar{\partial}X+\tfrac{1}{2}\alpha' k_i\cdot\tilde{\psi}\,e_i\cdot\tilde{\psi})e^{ik_i\cdot X}(w_i,\bar{w}_i)\right]\Big\rangle_\gamma\,. \tag{12.6.23}$$

For these spin structures, all vertex operators are in the 0 picture and there is no PCO.

We could proceed to calculate in a straightforward way, but the final result simplifies substantially and it would be better to simplify at the start. In fact, this is one amplitude that is much more easily obtained in the light-cone superstring formalism, and so we will effectively convert the calculation to that form.

First, analytically continue the momenta so that $k^0=k^1=0$. To be consistent with the mass-shell condition the momenta must become complex but this will not be a problem. Also, take the polarizations to vanish in the longitudinal directions. The longitudinal degrees of freedom then do not appear in the vertex operators, and so in the (P,A), (A,P), and (A,A) sectors the longitudinal path integrals just give determinants that cancel against the corresponding ghost path integrals. In particular, the combined longitudinal and ghost path integrals for these three sectors are independent of the spin structure, so the net spin structure dependence comes only from the eight transverse $\tilde{\psi}^i$.

Now, we will temporarily change the problem and also add in the (P,P) spin structure for the $\tilde{\psi}^i$, even though in the real amplitude this is multiplied by zero from the $\tilde{\psi}^{0,1}$ zero modes. The sum over four spin structures gives a GSO projection in the transverse $\tilde{\psi}^i$ CFT by itself. Consider the vertex operator $\tilde{\Theta}_\alpha$ for the R ground state, and bosonize:

$$\tilde{\Theta}_\alpha\to\left\{\begin{array}{rcl}\exp[\frac{1}{2}i(\tilde{H}_1+\tilde{H}_2+\tilde{H}_3+\tilde{H}_4)] &=& \exp(i\tilde{H}_1')\\ \exp[\frac{1}{2}i(\tilde{H}_1+\tilde{H}_2-\tilde{H}_3-\tilde{H}_4)] &=& \exp(i\tilde{H}_2')\\ \exp[\frac{1}{2}i(\tilde{H}_1-\tilde{H}_2+\tilde{H}_3-\tilde{H}_4)] &=& \exp(i\tilde{H}_3')\\ \exp[\frac{1}{2}i(\tilde{H}_1-\tilde{H}_2-\tilde{H}_3+\tilde{H}_4)] &=& \exp(i\tilde{H}_4')\end{array}\right\}\to\tilde{\theta}_\alpha\,. \tag{12.6.24}$$

Precisely for eight $\tilde{\psi}^i$, the linear combinations of scalars appearing in the spin field are themselves scalars of canonical normalization

$$\tilde{H}_i'(\bar{z})\tilde{H}_j'(0)\sim-\delta_{ij}\ln\bar{z}\,. \tag{12.6.25}$$

Thus after bosonizing and going to a new basis for the scalars we can refermionize in terms of free $(0,\frac{1}{2})$ fields $\tilde{\theta}_\alpha(\bar{z})$. Note that only for eight $\tilde{\psi}^i$ does the spin field have weight $\frac{1}{2}$. Thus we turn the $\tilde{\psi}^i$ path integral into a $\tilde{\theta}_\alpha$ path integral. Moreover, we claim that the spin structures are related

as follows:

$$\frac{1}{2}\sum_{\gamma}\langle\ \ \rangle_{\tilde{\psi},\gamma} = \langle\ \ \rangle_{\tilde{\theta}(\mathrm{P,P})} \ . \tag{12.6.26}$$

This follows because the spin field $\tilde{\Theta}_\alpha$ survives the GSO projection, implying that it is single-valued on the torus. The periodicity in the time direction implies the insertion of a factor $\exp(\pi i \tilde{F}_\theta)$ in the sum over states, which is appropriate because $\tilde{\theta}_\alpha$ is a spacetime spinor.

Now we come to the payoff: for this spin structure the $\tilde{\theta}_\alpha$ have eight zero modes, and so there must be eight factors of $\tilde{\theta}_\alpha$ to get a nonzero result. We need to refermionize the vertex operators, but this is easy. The spinors appear only in the combination

$$k_i e_j \tilde{\psi}^{[i}\tilde{\psi}^{j]} \ , \tag{12.6.27}$$

where we can antisymmetrize because $e \cdot k = 0$. The product of fermions is just an $SO(8)$ rotation current, so we can immediately write

$$\tilde{\psi}^{[i}\tilde{\psi}^{j]} \to \frac{1}{4}\tilde{\theta}^T \Gamma^{ij}\tilde{\theta} \ . \tag{12.6.28}$$

One can check this by taking the OPE of the two sides with $\tilde{\Theta}_\alpha$ and $\tilde{\theta}_\alpha$ respectively. The fermionic terms in the vertex operators then provide precisely the eight $\tilde{\theta}$s needed to saturate the zero modes, with the result

$$\begin{aligned}\left\langle \prod_{a=1}^{4} \frac{1}{4}\tilde{\theta}^T \Gamma^{i_a j_a}\tilde{\theta} \right\rangle_{\tilde{\theta}(\mathrm{P,P})} &= \frac{1}{2^8}\epsilon^{\alpha_1\ldots\alpha_8}\Gamma^{i_1 j_1}_{\alpha_1\alpha_2}\cdots\Gamma^{i_4 j_4}_{\alpha_7\alpha_8} \\ &= t^{i_1 j_1 \ldots i_4 j_4} + \epsilon^{i_1 j_1 \ldots i_4 j_4} \ .\end{aligned} \tag{12.6.29}$$

Here t is the same tensor (12.4.25) that appears in the tree-level amplitudes.

It remains to separate out the unwanted (P,P) sector of the $\tilde{\psi}$ path integral, but this is easy. It is the only one that is odd under a reflection of one of the transverse directions, so it is responsible for the term $\epsilon^{i_1 j_1 \ldots i_4 j_4}$. Thus we omit this term, which in any case does not contribute because the momenta with which it contracts are not linearly independent. Further, the tensor t has a unique covariant extension.

The remaining factors are much as in the previous amplitude, leading for $SO(32)$ to the effective interaction

$$\frac{1}{2^8\pi^5 4!\,\alpha'} t^{\mu\nu\sigma\rho\alpha\beta\gamma\delta}\mathrm{Tr}_{\mathrm{v}}(F_{\mu\nu}F_{\sigma\rho}F_{\alpha\beta}F_{\gamma\delta}) \ . \tag{12.6.30}$$

For $E_8 \times E_8$ one has instead the group theory structure (12.6.22). Given the similarity of the F^4 amplitude to the BF^4 amplitude, the reader may not be surprised that they are in fact related by supersymmetry.

By this same method several other amplitudes can be obtained, including the type I cylinder with four open string gauge bosons and the type II torus with four gravitons. In the latter case both the ψ^i and the $\tilde{\psi}^i$ are refermionized, and the tensor structure is the same as for the tree-level amplitude (12.4.34), with two t tensors.

The $\tilde{\theta}_\alpha$ are the first free fields carrying a spacetime spinor index that we have encountered. One might have expected these to arise at some earlier stage. In fact the covariant Green–Schwarz superstring theory, with manifest spacetime supersymmetry, has such fields. It is equivalent to the RNS superstring: after putting each theory in the light cone they are related by the refermionization above. However, the constraints and gauge fixing in the Green–Schwarz description are rather more complicated, and so we have chosen not to emphasize this subject.

Nonrenormalization theorems

It follows from the preceding calculations that any amplitude with three or fewer massless particles vanishes because there are too few factors of θ_α to saturate the zero-mode integrations. One consequence is that there is no renormalization of Newton's constant, which can be measured in the three-graviton amplitude.

It also follows that all amplitudes vanish at least as k^4 when $k \to 0$, from the explicit momentum factors in the vertex operators.[5] This has the important physical consequence that the constant background

$$G_{\mu\nu}(x) = \eta_{\mu\nu} \ , \quad \Phi(x) = \Phi_0 \tag{12.6.31}$$

around which we are expanding remains a solution of the field equations to one-loop order. No interaction

$$\int d^{10}x\,(-G)^{1/2}V(\Phi) \tag{12.6.32}$$

is generated. Actually we already knew this from the calculation of the one-loop vacuum amplitude in chapter 10, which vanished by cancellation between bosons and fermions. These nonrenormalization theorems have been argued to extend to all orders of string perturbation theory; the details are left to the references.

Nonrenormalization can also be understood from a spacetime point of view. The tree-level action has local supersymmetry. Therefore the loop corrections must respect this symmetry or else the unphysical polarizations of the gravitino will not decouple. However, no interaction of the

[5] This kind of argument is subtle because one can obtain offsetting poles from $\int d^2w\,(w\bar{w})^{-1+k\cdot k'}$, but the necessary singularity in w does not appear here because the zero modes are independent of $\bar{w}$.

form (12.6.32) is allowed by $d = 10$, $N = 1$ or $N = 2$ supersymmetry.[6] This argument applies to all orders of perturbation theory and in fact it is a nonperturbative (exact) result. The latter fact is very striking, because with the string technology developed so far we have no direct way to understand strings beyond perturbation theory. It should be noted that the leap from 'all orders of perturbation theory' to 'exact' is quite nontrivial, because in theories with less symmetry there are many examples of corrections that arise only from nonperturbative effects. We will see some of these later.

Exercises

12.1 Derive the $SO(n)$ trace identities (12.2.19). You can assume a basis in which the generator t is a linear combination of the commuting generators H^i.

12.2 Obtain the trace relations (12.2.20) for E_8, and show that $\mathrm{Tr}_a(t^6)$ can be reduced to lower order traces for $E_8 \times E_8$.

12.3 Show that the anomaly factorizes for the massless spectrum of the $SO(16) \times SO(16)$ nonsupersymmetric heterotic string.

12.4 Show that the superfield forms for the superconformal transformation δX^μ, the OPEs (12.3.22) and (12.3.26), and the action (12.3.23) reduce to the correct component forms.

12.5 Show that the superfield form of the sigma model action reduces to components as shown in eq. (12.3.27).

12.6 Using the contour method from sections 6.2 and 6.3, show that the sphere amplitude must have total ϕ charge -2 as discussed at the beginning of section 12.4.

12.7 (a) Calculate the tree-level heterotic string amplitude with two gauginos and a gauge boson.
(b) Calculate the tree-level heterotic string amplitude with two gauginos and a massless tensor.

12.8 Calculate the tree-level type II amplitude with one NS–NS boson and two R–R bosons.

12.9 Calculate the tree-level heterotic string amplitude with four gauginos. You can either do this directly, or by first calculating an appropriate bosonic open string amplitude and then using the open–closed relation (12.4.29).

[6] The massive IIA supergravity theory (12.1.24) effectively has such a term after setting the fields M and F_{10} to fixed background values, but the dilaton dependence is fixed and corresponds to a tree-level effect.

12.10 Calculate the $k_i \to 0$ limit of the type I cylinder amplitude with four gauge bosons, where two open string vertex operators are on each boundary. This is often referred to as the *nonplanar amplitude*.

12.11 Calculate the same amplitude as in the previous problem but with all four vertex operators on one boundary. This *planar amplitude* has a divergence; calculate the canceling Möbius amplitude.

13
D-branes

In chapter 8 we found that a number of new phenomena, unique to string theory, emerged when the theory was toroidally compactified. Most notable were the T-duality of the closed oriented theory and the appearance of D-branes in the $R \to 0$ limit of the open string theory. These subjects become richer still with the introduction of supersymmetry. We will see that the D-branes are BPS states and carry R–R charges. We will argue that the type I, IIA, and IIB string theories are actually different states in a single theory, which also includes states containing general configurations of D-branes. Whereas previously we considered only parallel D-branes all of the same dimension, we now wish to study more general configurations. We will be concerned with the breaking of supersymmetry, the spectrum and effective action of strings stretched between different D-branes, and scattering and bound states of D-branes. In the present chapter we are still in the realm of string perturbation theory, but many of the results will be used in the next chapter to understand the strongly coupled theory.

13.1 T-duality of type II strings

Even in the closed oriented type II theories T-duality has an interesting new effect. Compactify a single coordinate X^9 in either type II theory and take the $R \to 0$ limit. This is equivalent to the $R \to \infty$ limit in the dual coordinate, whose right-moving part is reflected

$$X_R'^9(\bar{z}) = -X_R^9(\bar{z}) \tag{13.1.1}$$

just as in the bosonic string. By superconformal invariance we must also reflect $\tilde{\psi}^9(\bar{z})$,

$$\tilde{\psi}'^9(\bar{z}) = -\tilde{\psi}^9(\bar{z}) \,. \tag{13.1.2}$$

However, this implies that the chirality of the right-moving R sector ground state is reversed: the raising and lowering operators $\tilde{\psi}_0^8 \pm i\tilde{\psi}_0^9$ are interchanged. Simply put, T-duality is a spacetime parity operation on just one side of the world-sheet, and so reverses the relative chiralities of the right- and left-moving ground states. If we begin with the IIA theory and take the compactification radius to be small, we obtain the IIB theory at large radius, and vice versa. The same is true if we T-dualize — that is, carry out the change of variables (13.1.1) and (13.1.2) — on any odd number of dimensions, while T-dualizing on an even number returns one to the type II theory with which one began. Thus the two type II theories are related in the same way as the two heterotic theories: in each case the two noncompact theories are different limits of a single space of compactified theories. The type II relation is even simpler than the heterotic relation, in that one takes the radius to zero without having also to include a Wilson line.

Since the IIA and IIB theories have different R–R fields, T-duality must transform one set into the other. Again focus on T-duality in just the 9-direction. In order to preserve the OPE between $\tilde{\psi}^\mu$ and the spin field, this must act as

$$\mathscr{S}'_\alpha(z) = \mathscr{S}_\alpha(z) \,, \quad \tilde{\mathscr{S}}'_\alpha(\bar{z}) = \beta^9_{\alpha\beta}\tilde{\mathscr{S}}_\beta(\bar{z}) \,, \tag{13.1.3}$$

where β^9 is the parity transformation (9-reflection) on the spinors. It anticommutes with Γ^9 and commutes with the remaining Γ^μ, so $\beta^9 = \Gamma^9\Gamma$. Now consider the effect on the R–R vertex operators

$$\overline{\mathscr{S}}\Gamma^{\mu_1\ldots\mu_p}\tilde{\mathscr{S}} \,. \tag{13.1.4}$$

The T-duality multiplies the product of Γ matrices by $\Gamma^9\Gamma$ on the right. The Γ just gives ± 1 because the R ground states have definite chirality. The Γ^9 adds a '9' to the set $\mu_1 \ldots \mu_p$ if none is present, or removes one if it is present via $(\Gamma^9)^2 = 1$. This is how T-duality acts on the R–R field strengths and potentials, adding or subtracting the index for the dualized dimensions. Thus, if we start from the IIA string we get the IIB R–R fields as follows (up to signs),

$$C_9 \to C \,, \tag{13.1.5a}$$

$$C_\mu, C_{\mu\nu 9} \to C_{\mu 9}, C_{\mu\nu} \,, \tag{13.1.5b}$$

$$C_{\mu\nu\lambda} \to C_{\mu\nu\lambda 9} \,, \tag{13.1.5c}$$

where here μ stands for a nondualized dimension. We could go on, getting $C_{\mu\nu\lambda\omega}$ from $C_{\mu\nu\lambda\omega 9}$ and so on, but these are not independent fields, and give rather the Poincaré dual description of the fields listed.

For T-duality on multiple dimensions replace β^9 with

$$\prod_m \beta^m , \tag{13.1.6}$$

where $\beta^m = \Gamma\Gamma^m$ and the product runs over the dualized directions. There are some signs which should be noted but should not distract attention from the main physical point. Since $\beta^m\beta^n = -\beta^n\beta^m$ for $m \neq n$, T-dualities in different directions do not quite commute but differ by a sign in the right-moving R sector. We can write this as

$$\beta^m\beta^n = \exp(\pi i\tilde{\mathbf{F}})\beta^n\beta^m , \tag{13.1.7}$$

where $\tilde{\mathbf{F}}$ is the spacetime fermion number of the right-moving state of the string; this is a symmetry that flips the sign of all right-moving R states. Also, we have defined β^m so as to preserve the Hermiticity of $\tilde{\mathscr{V}}_\alpha$ (that is, it is real in a Majorana basis), but then $\beta^m\beta^m = -1$ and so acting twice with T-duality gives $\exp(\pi i\tilde{\mathbf{F}})$.

13.2 T-duality of type I strings

Taking the $R \to 0$ limit of the open and unoriented type I $SO(32)$ theory leads to D-branes and orientifold planes by the same arguments as for the bosonic string in chapter 8, which the reader should review. In particular, taking the T-dual on a single dimension leads to a space with 16 D8-branes between two orientifold hyperplanes.

Let us first consider the bulk physics of the T-dual theory, obtained by taking $R \to 0$ and concentrating on a region of the dual spacetime that is far away from the fixed planes and D-branes, as illustrated in figure 13.1. The local physics is that of a closed oriented superstring theory: closed because the open strings live far away on the D-branes; oriented because the orientation projection relates the state of any string to that of its image behind the fixed plane, but does not locally constrain the space of states. Thus the local physics must be that of a type II theory. In particular there are two gravitinos, and any closed string scattering process will be invariant under the 32 supersymmetries of the type II theory. Since the type I theory with which we started has equal left- and right-moving chiralities, taking the T-dual in one dimension makes them opposite: the local physics is the IIA superstring. Taking the T-dual on any odd number of dimensions has the same effect; taking the T-dual on any even number of dimensions gives the IIB theory in the bulk.

Now take the $R \to 0$ limit while concentrating on the neighborhood of one D-brane in the T-dual theory, adjusting the Wilson lines so that again the fixed plane and other D-branes move away in the T-dual spacetime. The low energy degrees of freedom on the D-brane are the massless open



Fig. 13.1. A D-brane, with one attached open string and one closed string moving in the bulk. The physics away from the D-brane is described by a type II string theory, so the string theory with the D-brane has the physical properties of a state of the *type II* theory containing an extended object.

string states

$$\psi^\mu_{-1/2}|k\rangle_{\mathrm{NS}} , \quad \psi^9_{-1/2}|k\rangle_{\mathrm{NS}} , \quad |\alpha;k\rangle_{\mathrm{R}} . \tag{13.2.1}$$

As in the bosonic theory, the bosonic states are a gauge field living on the D-brane and the collective coordinates for the D-brane. The fermionic states are the superpartners of these.

Consider now a process where closed strings scatter from the D-brane; this necessarily involves a world-sheet with boundary. Now, the open string boundary conditions are invariant only under $d = 10$, $N = 1$ supersymmetry. In the original type I theory, the left-moving world-sheet current for spacetime supersymmetry $\mathscr{V}_\alpha(z)$ flows into the boundary and the right-moving current $\tilde{\mathscr{V}}_\alpha(\bar{z})$ flows out, so only the total charge $Q_\alpha + \tilde{Q}_\alpha$ of the left- and right-movers is conserved. Under T-duality this becomes

$$Q'_\alpha + (\beta^9\tilde{Q}')_\alpha . \tag{13.2.2}$$

The scattering amplitudes of closed strings from the D-brane are invariant only under these 16 supersymmetries.

To see the significance of this, consider first the conservation of momentum. There is a nonzero amplitude for a closed string to reflect backwards from the D-brane, which clearly does not conserve momentum in the direction orthogonal to the D-brane. This occurs because the Dirichlet boundary conditions explicitly break translational invariance. However,

from the spacetime point of view the breaking is spontaneous: we are expanding around a D-brane in some definite location, but there are degenerate states with the D-brane translated by any amount.[1] For a spontaneously broken symmetry the consequences are more subtle than for an unbroken symmetry: the apparent violation of the conservation law is related to the amplitude to emit a long-wavelength Goldstone boson. For the D-brane, as for any extended object, the Goldstone bosons are the collective coordinates for its motion. In fact, the nonconservation of momentum is measured by the integral of the corresponding current over the world-sheet boundary,

$$\frac{1}{2\pi\alpha'}\int_{\partial M} ds\, \partial_n X'^9 \,, \tag{13.2.3}$$

which up to normalization is just the (0 picture) vertex operator for the collective coordinate, with zero momentum in the Neumann directions.

We conclude by analogy that the D-brane also spontaneously breaks 16 of the 32 spacetime supersymmetries, the ones that are explicitly broken by the open string boundary conditions. The integrals

$$\int_{\partial M} ds\, \mathscr{V}'_\alpha = -\int_{\partial M} ds\, (\beta^9 \tilde{\mathscr{V}}')_\alpha \,, \tag{13.2.4}$$

which measure the breaking of supersymmetry, are just the vertex operators for the fermionic open string state (13.2.1). Thus this state is a *goldstino*, the Goldstone state associated with spontaneously broken supersymmetry.

It is not surprising that the D-brane breaks some supersymmetry. The only state invariant under all supersymmetries is the vacuum. Rather, what is striking is that it leaves half the supersymmetries unbroken: *it is a BPS state.* This same argument holds for any number of dualized dimensions, and so for Dp-branes for all p. The unbroken supersymmetry is

$$Q'_\alpha + (\beta^\perp \tilde{Q}')_\alpha \,, \tag{13.2.5}$$

where

$$\beta^\perp = \prod_m \beta^m \,, \tag{13.2.6}$$

the product running over all the dimensions perpendicular to the D-brane.

[1] The Mermin–Wagner–Coleman theorem from quantum field theory implies that if the D-brane has two or more noncompact directions there will indeed be an infinite number of degenerate states. If it has one or zero noncompact directions, quantum fluctuations force it into a unique translationally invariant state. The latter effect shows up in perturbation theory through IR divergences. For a spontaneously broken *supersymmetry* the fluctuations are less effective: supersymmetry can be broken even by a zero-dimensional object.

BPS states, which are discussed in section B.2, must carry conserved charges. In the present case there is a natural set of charges with the correct Lorentz properties, namely the antisymmetric R–R charges. The world-volume of a p-brane naturally couples to a $(p+1)$-form potential

$$\int C_{p+1} \,, \tag{13.2.7}$$

the integral running over the D-brane world-volume. By T-duality we can reach the IIA theory with a Dp-brane of any even p. Thus we need 1-, 3-, 5-, 7-, and 9-form potentials. Indeed, the 1-form and 3-form are in the IIA theory and the 5-form and 7-form give equivalent descriptions of the same physics. The 9-form potential we have discussed in section 12.1 in the context of massive IIA supergravity. Although it is not associated with propagating states, and so was not detected in the quantization of the IIA string, the existence of D8-branes shows that it must be included.

By analogy with electromagnetism in four dimensions, where the 1-form electric potential can be replaced with a 1-form magnetic potential, a Dirichlet p-brane and $(6-p)$-brane are like electric and magnetic sources for the same field strength. For example, the free field equation and Bianchi identity for a 2-form field strength, $d*F_2 = dF_2 = 0$, are symmetric between F_2 and $(*F)_8$, and can be written either in terms of a 1-form or 7-form potential:

$$F_2 = dC_1 \,, \quad d\wedge *dC_1 = 0 \,, \tag{13.2.8a}$$

$$*F_2 = (*F)_8 = dC_7 \,, \quad d\wedge *dC_7 = 0 \,. \tag{13.2.8b}$$

At an electric source, which would be a D0-brane for C_1 or a D6-brane for C_7, the field equation has a source term. At a magnetic source, a D6-brane for C_1 or a D0-brane for C_7, the Bianchi identity breaks down, and the potential cannot be globally defined: one must introduce a Dirac string, or use different potentials in different patches.[2]

For the IIB theory we need 2-, 4-, 6-, 8-, and 10-form potentials. The first four arise in either the electric or magnetic description of the propagating R–R states. The existence of the 10-form was deduced in section 10.8, from the study of type I divergences. Indeed, we argued there for the coupling (13.2.7) for the 10-form, where the integral runs over all spacetime. This fits with a point made in chapter 8, that it is natural to interpret each Chan–Paton degree of freedom in the fully Neumann theory as a 9-brane filling spacetime. All the other R–R couplings follow from this one by T-duality, since each time we T-dualize in an additional

[2] It should be mentioned that there is no local covariant action for a system with both electric and magnetic charges, even though the physics is covariant and presumably satisfies the axioms of local quantum field theory.

direction the dimension of the p-branes goes down by one and the R–R form loses an index.

The IIB theory also has a 0-form potential C_0, the R–R scalar. This should couple to a '(-1)-brane.' Indeed, there is a natural interpretation to this: it is defined by Dirichlet boundary conditions in all directions, time as well as space, so its world-sheet is zero-dimensional and the integral (13.2.7) reduces to the value of C_0 at that point. An object that is localized in time as well as space is an instanton. Instantons in Euclidean path integrals correspond to tunneling events, and we will argue shortly that these must be present in string theory.

We will verify the R–R couplings of D-branes in the next section; for the remainder of this section we will discuss some of the consequences. The discovery that D-branes carry R–R charges neatly ties together two loose ends. On the one hand, it was argued in section 12.1 that the ordinary string states do not have R–R charges, but now we see that string theory does have a source for every gauge field.[3] This extends the result from chapter 8, that the gauge field from compactification of the antisymmetric tensor (under which all states in quantum field theory are neutral) couples to winding strings. On the other hand, the existence of so many different kinds of extended object, Dp-branes for every p, might have seemed excessive, but we now see that these are in one-to-one correspondence with the R–R potentials of the respective type II theories.

The divergence of the type I theory for groups other than $SO(32)$ arose from the R–R 10-form field equation. This divergence is unaffected by toroidal compactification and again cancels only for $SO(32)$. It would have been surprising if toroidal compactification made a consistent theory inconsistent, or the reverse, and it is not hard to verify explicitly that this does not happen. The effect of toroidal compactification is to add world-sheets that wrap around the periodic directions of spacetime. These correspond to exchange of closed strings with winding number, which are massive and so do not have dangerous tadpoles.

The spacetime interpretation of the divergence in the T-dual picture with D-branes is again an inconsistency in the R–R field equations. One can picture field lines emerging from each D-brane, orthogonal to the noncompact dimensions, and these field lines must end somewhere. Further, all D-branes must have the same sign of the charge: the full set of D-branes is still a BPS state, being T-dual to the type I theory, and the total mass is linear in the total charge for a BPS state. We know that the disk tadpole is canceled by the unoriented cross-cap. In the T-dual spacetime the cross-cap must be localized near one of the orientifold

[3] In the next chapter we will discuss a seemingly different kind of R–R source, the *black p-brane*.

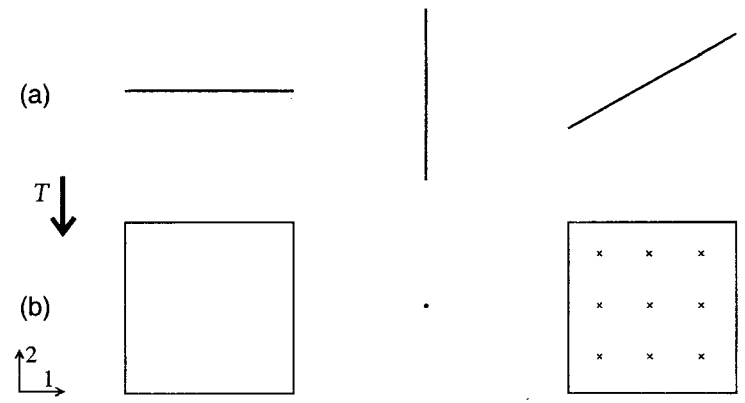


Fig. 13.2. Effect of a T-duality in the 2-direction on D1-branes at various angles in the (1,2) plane: (a) before T-duality; (b) after T-duality. The $\times$s indicate a magnetic field on the D2-brane.

planes, because the string theory in the bulk is oriented. Thus we deduce that the orientifold planes are sinks for R–R charge. If we T-dualize on k dimensions there are 2^k orientifold planes but still 16 D-branes, so the charge of an orientifold plane must be -2^{4-k} times that of a D-brane of the same dimension.

New connections between string theories

Starting from the toroidally compactified type I theory, we can reach either $d = 10$ type II theory. Simply take an odd or even number of radii to zero, while moving the D-branes and fixed planes off to infinity as the dual spacetime expands. Thus, just as for the two heterotic theories, these should be thought of as limits of a single theory. The theory has many other states as well: we can take the limit while keeping some of the D-branes in fixed positions, so that we obtain the compact theory in a state with D-branes. The simple T-duality leads only to parallel D-branes of equal dimension, but since the D-branes are dynamical we can continuously vary their configurations. We can then reach states with p-branes of different dimension as follows. Consider two D1-branes (D-strings) in the IIB theory, from dualizing in eight directions. Let one be along the 1-direction and the other be rotated to lie along the 2-direction. As illustrated in figure 13.2, a further T-duality in the 2-direction reverses Dirichlet and Neumann boundary conditions in this direction and so turns

the first D-string into a D2-brane extended in the 1- and 2-directions but the second into a D0-brane. Thus these can coexist in the IIA theory. Of course T-duality leads only to states with 16 D-branes, but we understand now that this is due to the R–R flux conservation in the compact space. In a noncompact space the R–R flux can run to infinity and so any number of D-branes should be allowed.

Thus, starting from the type I theory we can reach states that look like the type IIA theory with any collection of even Dp-branes or the type IIB theory with any collection of odd Dp-branes. Of course if we start with an ordinary type II theory, T-duality will never give us open strings or D-branes, so one might imagine that there is a different type II theory in which D-branes are not allowed. This seems unlikely, however: everything we know points to the uniqueness of the theory, so we do not have such alternatives. Also, we will see in the next chapter that the inclusion of D-branes leads to a much more elegant and symmetric theory.

In summary, we are now considering a single theory, which has a state that contains no D-branes and looks like the ordinary IIA theory, a second state (T-dual to the first) that contains no D-branes and looks like the ordinary IIB theory, and a third state that contains 16 D9-branes (and an orientifold 9-plane) that looks like the type I theory. It also contains an infinite number of other states with very general configurations of D-branes.

We can now write down the supersymmetry algebra for this theory:

$$\{Q_\alpha, \bar{Q}_\beta\} = -2\left[P_M + (2\pi\alpha')^{-1} Q_M^{\mathrm{NS}}\right]\Gamma^M_{\alpha\beta}\,, \tag{13.2.9a}$$

$$\{\tilde{Q}_\alpha, \tilde{\bar{Q}}_\beta\} = -2\left[P_M - (2\pi\alpha')^{-1} Q_M^{\mathrm{NS}}\right]\Gamma^M_{\alpha\beta}\,, \tag{13.2.9b}$$

$$\{Q_\alpha, \tilde{\bar{Q}}_\beta\} = -2\sum_p \frac{\tau_p}{p!} Q^{\mathrm{R}}_{M_1\ldots M_p}(\beta^{M_1}\cdots\beta^{M_p})_{\alpha\beta}\,. \tag{13.2.9c}$$

The spacetime supersymmetries Q_α and $\tilde{Q}_\alpha$ act respectively on the right- and left-movers.

The anticommutator (13.2.9b) of two right-moving supersymmetries is the same as the heterotic string anticommutator (11.6.32), containing the charge that couples to the NS–NS 2-form. The argument for the appearance of this term is the same as before: the $\tilde{\mathscr{V}}_\alpha\tilde{\mathscr{V}}_\beta$ OPE contains the right-moving momentum $\bar{\partial}X^\mu$, which involves both ordinary momentum and winding number. Similarly, the $\mathscr{V}_\alpha\mathscr{V}_\beta$ OPE contains the left-moving momentum ∂X^μ, so the NS–NS charge appears in the left-moving anticommutator with the opposite sign. We have added a superscript NS to distinguish this charge from the charges Q^{R} that couple to R–R forms. Also, we have changed conventions so that all charges are now normalized to one per unit world-volume of the respective extended object, and so the string tension $(2\pi\alpha')^{-1}$ appears explicitly. As discussed in section 11.6,

Q_M^{NS} is the charge carried by the *fundamental string*, meaning the original quantized string. Henceforth we use this term, or *F-string*, to distinguish it from the D1-brane.

From the argument that D-branes are BPS states, we expect the R–R charges to appear in the algebra as well, and the natural place for an R–R charge to appear is in the anticommutator of a left- and a right-moving supersymmetry. The sum on p runs over even values in the IIA theory and odd values in the IIB theory. By analogy with the NS–NS case we have included the D-brane tensions τ_p, whose values will be obtained in the next section; the factor of $1/p!$ offsets the sum over permutations of indices. To see that the algebra is correct, focus on a state that contains a single static Dp-brane. The nonzero charge is $Q_{\mu_1\cdots\mu_p}$, where the indices run over the directions tangent to the Dp-brane. Note that

$$\beta^{\mu_1}\cdots\beta^{\mu_p} = \beta^\perp\Gamma^0\,, \tag{13.2.10}$$

up to a possible overall sign that can be reabsorbed in the definition of $\tilde{Q}$; $\beta^\perp$ is the same as in eq. (13.2.5). It then follows that the anticommutator of $Q+\beta^\perp\tilde{Q}$ with any supercharge vanishes in this state, as required by the BPS property. (In eq. (13.2.5) we included primes on the supercharges to indicate that we were working in a T-dual description to the type I theory with which we began. In writing the algebra (13.2.9) we are considering an arbitrary state with D-branes, without necessarily obtaining it from T-duality, so there are no primes.)

Incidentally, the central charge (13.2.9) is still not complete: the *magnetic NS–NS* charge is missing. This is not carried by D-branes or F-strings. We will discuss this further in the next chapter.

Finally, let us also explain the necessity of D-instantons, localized in time. We could try to use T-duality in the time direction, but it is not clear whether this is meaningful. Rather, consider D0-branes, whose world-lines are one-dimensional, in a space with one compact spatial dimension. For an ordinary quantized particle in a path sum description we would have to include closed paths that wind around the compact direction. Such paths are responsible for Casimir energies and other effects of compactification. Presumably we must do the same for the D0-branes as well. The shortest such winding path is a straight line in the compact spatial direction. This is localized in time and so is essentially an instanton: Casimir energies, in the path sum description, are essentially instanton effects. Further, by a T-duality in the compact dimension we obtain a D-instanton that is localized in all directions. We know from chapter 8 that the D-brane action depends on the closed string coupling as $O(1/g)$, so the D-instanton amplitude is of order $e^{-O(1/g)}$. Thus we have found an example of the enhanced nonperturbative effects that were inferred in section 9.7 from the high order behavior of string perturbation theory.

On the heterotic world-sheet there are no boundary conditions that preserve the world-sheet gauge symmetries, and there is no indication that D-branes exist. Nevertheless, we will see in the next chapter that the D-branes of the type I/II theory enable us to learn a great deal about the heterotic string as well.

13.3 The D-brane charge and action

There is no force between static BPS objects of like charge. The multi-object state is still supersymmetric and so its energy is determined only by its charge and is independent of the separations. For parallel Dp-branes, the unbroken supersymmetry (13.2.5) is the same as for a single Dp-brane.

The vanishing of the force comes about from a cancellation between attraction due to the graviton and dilaton and repulsion due to the R–R tensor. We can calculate these forces explicitly from the usual cylinder vacuum amplitude. The exchange of light NS–NS closed strings was isolated in eq. (10.8.4). Modify this expression by removing the factors for the momentum integrations in the Dirichlet directions and introducing a term for the tension of a string stretched over a separation y^μ:

$$\begin{aligned}\mathscr{A}_{\text{NS-NS}} &\approx \frac{iV_{p+1}4\times 16}{8\pi(8\pi^2\alpha')^5}\int_0^\infty \frac{\pi dt}{t^2}(8\pi^2\alpha' t)^{(9-p)/2}\exp\left(-\frac{ty^2}{2\pi\alpha'}\right)\\ &= iV_{p+1}2\pi(4\pi^2\alpha')^{3-p}G_{9-p}(y)\end{aligned} \tag{13.3.1}$$

with $G_d(y) = 2^{-2}\pi^{-d/2}\Gamma(\frac{1}{2}d-1)y^{2-d}$ the scalar Green's function. The Chan–Paton weight is 2 here, from the two orientations of the open string, and there is no factor of $\frac{1}{2}$ from the orientation projection because the physics is locally oriented. Due to supersymmetric cancellation in the trace, the R–R exchange amplitude is

$$\mathscr{A}_{\text{R-R}} = -\mathscr{A}_{\text{NS-NS}} \tag{13.3.2}$$

and so the total force vanishes as expected.

The field theory calculation (8.7.25) for the dilaton–graviton potential changes only by the substitution $6 = (D-2)/4 \to 2$, and so is

$$2i\kappa^2\tau_p^2 G_{9-p}(y)\,. \tag{13.3.3}$$

Thus

$$\tau_p^2 = \frac{\pi}{\kappa^2}(4\pi^2\alpha')^{3-p}\,. \tag{13.3.4}$$

This satisfies the same T-duality relation as in the bosonic string. For the R–R exchange, the low-energy action is

$$-\frac{1}{4\kappa_{10}^2}\int d^{10}x\,(-G)^{1/2}|F_{p+2}|^2 + \mu_p\int C_{p+1}\,. \tag{13.3.5}$$

The kinetic term is canonically normalized, so the propagator for any given component (such as the one parallel to the D-brane) is i/k^2, and the field theory amplitude is

$$-i\mu_p^2 G_{9-p}(y)\,. \tag{13.3.6}$$

Hence

$$\mu_p^2 = \frac{\pi}{\kappa_{10}^2}(4\pi^2\alpha')^{3-p} = e^{2\Phi}\tau_p^2\,. \tag{13.3.7}$$

The reader can carry out a similar calculation of the force between a D-brane and an orientifold plane and show that it has an additional $-(2^{5-k})$. We deduced from the cancellation of divergences that the charge of the orientifold plane should have a factor of $-(2^{4-k})$; the extra factor of 2 in the force arises because the orientifold geometry squeezes the flux lines into half the solid angle.

The calculation of the interaction confirms our earlier deduction that D-branes carry the R–R charges. It is interesting to see how this is consistent with our earlier discussion of string vertex operators. The R–R vertex operator (12.1.14) is in the $(-\frac{1}{2},-\frac{1}{2})$ picture, which can be used in almost all processes. On the disk, however, the total left- plus right-moving ghost number must be -2. With two or more R–R vertex operators, all can be in the $(-\frac{1}{2},-\frac{1}{2})$ picture (with PCOs included as well), but a single vertex operator must be in either the $(-\frac{3}{2},-\frac{1}{2})$ or the $(-\frac{1}{2},-\frac{3}{2})$ picture. The $(-\frac{1}{2},-\frac{1}{2})$ vertex operator is essentially $e^{-\phi}G_0$ times the $(-\frac{3}{2},-\frac{1}{2})$ operator, so besides the shift in the ghost number the latter has one less power of momentum and one less Γ-matrix. The missing factor of momentum turns F into C, and the missing Γ-matrix gives the correct Lorentz representations for the potential rather than the field strength.

Dirac quantization condition

There is an important consistency check on the value of the R–R charge, which generalizes the Dirac quantization condition for magnetic monopole charge. Let us review the Dirac condition, shown in figure 13.3. Consider a magnetic charge μ_{m} at the origin. The integrated flux is

$$\int_{S_2} F_2 = \mu_{\text{m}}\,. \tag{13.3.8}$$

Because the integral over a closed surface is nonzero, we cannot write $F_2 = dA_1$ for any vector potential. However, we can write $F_2 = dA_1$ except along a Dirac string ending on the monopole. Now consider an electric charge μ_{e} moving in this field. Its coupling to the field produces a

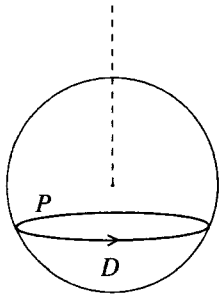


Fig. 13.3. Sphere surrounding monopole, with a Dirac string running upward. The particle path P is bounded by the lower cap D.

phase

$$\exp\left(i\mu_e \oint_P A_1\right) = \exp\left(i\mu_e \int_D F_2\right) \tag{13.3.9}$$

when the particle moves on a closed path P. The surface D spans P and does not intersect the Dirac string. Now consider the limit as the path is contracted to a small circle around the Dirac string. The phase becomes

$$\exp\left(i\mu_e \int_{S_2} F_2\right) = \exp(i\mu_e\mu_m) \,. \tag{13.3.10}$$

The Dirac string must be invisible, so this phase must be 1. Equivalently, this is the condition that the phase (13.3.9) is unchanged if we instead take the surface $D' = S_2 - D$ spanning P in the upper hemisphere. The result is the Dirac quantization condition,

$$\mu_e\mu_m = 2\pi n \tag{13.3.11}$$

for some integer n.

A p-brane and $(6-p)$-brane are sources for F_{p+2} and F_{8-p} respectively. These two field strengths are Poincaré dual to one another, so again there is a Dirac quantization condition that must be satisfied by the product of their charges. Let us think about F_{p+2} as the field strength, so that the p-brane is an electric source and the $(6-p)$-brane a magnetic source. In nine dimensions a $(6-p)$-dimensional object is surrounded by a $(p+2)$-sphere, so by analogy to the magnetic flux (13.3.8),

$$\int_{S_{p+2}} F_{p+2} = \mu_{6-p} \,. \tag{13.3.12}$$

One can then repeat the same argument. For example, let the p-brane be extended in the directions $4 \le \mu \le p+3$ and the $(6-p)$-brane in the

directions $p+4 \le \mu \le 9$. The system essentially reduces to the three-dimensional situation of figure 13.3 in the directions $\mu = 1,2,3$, and the charges must satisfy

$$\mu_p\mu_{6-p} = 2\pi n \,. \tag{13.3.13}$$

Remarkably, the charges (13.3.7), arrived at in an entirely different way, satisfy this relation with the minimum quantum $n = 1$.

D-brane actions

The coupling of a D-brane to NS–NS closed string fields is the same Dirac–Born–Infeld action as in the bosonic string,

$$S_{Dp} = -\mu_p \int d^{p+1}\xi \, \mathrm{Tr}\left\{e^{-\Phi}[-\det(G_{ab} + B_{ab} + 2\pi\alpha' F_{ab})]^{1/2}\right\} , \tag{13.3.14}$$

where G_{ab} and B_{ab} are the components of the spacetime NS–NS fields parallel to the brane and F_{ab} is the gauge field living on the brane. The argument leading to this form is exactly as in the bosonic case, section 8.7. Recall that for n D-branes at small separation, where the strings stretched between them are light enough to be included in the low energy action, the collective coordinates $X^\mu(\xi)$, gauge fields $A_a(\xi)$, and their fermionic partners $\lambda(\xi)$ all become $n \times n$ matrices. The trace in the action is in this $n \times n$ space. In addition there is a term

$$O([X^m, X^n]^2) \tag{13.3.15}$$

in the potential. As discussed in chapter 8, the effect of this potential is that in the flat directions the collective coordinates become diagonal. They can then be interpreted as n ordinary collective coordinates for n objects. At small separation the full matrix dynamics is crucial, as we will see.

The coupling to the R–R background also includes corrections involving the gauge field on the brane. Like the Born–Infeld action, these can be deduced via T-duality. Consider, as an example, a 1-brane in the (1,2) plane. The action is

$$\int C_1 = \int (dx^1 C_1 + dx^2 C_2) = \int dx^1 \left(C_1 + \partial_1 X^2 C_2\right) . \tag{13.3.16}$$

Under a T-duality in the 2-direction this becomes

$$\int dx^1 \, (C_{12} + 2\pi\alpha' F_{12} C) \,. \tag{13.3.17}$$

We have used the T-transformation of the C fields, eq. (13.1.5). A D-brane at an angle is T-dual to one with a magnetic field, as in figure 13.2. We are not keeping track of the normalization but one could, with the result $\mu_p = \mu_{p-1}/2\pi\alpha'^{1/2}$ in agreement with the explicit calculation. The

generalization of (13.3.17) to an arbitrary configuration, and to multiple D-branes, gives the Chern–Simons-like result

$$i\mu_p \int_{p+1} \mathrm{Tr}\left[\exp(2\pi\alpha' F_2 + B_2) \wedge \sum_q C_q\right] . \tag{13.3.18}$$

The expansion of the integrand (13.3.18) involves forms of various rank; the integral picks out precisely the terms that are proportional to the volume form of the p-brane world-volume. There are similar couplings with the spacetime curvature in addition to the field strength; these can be obtained from a string calculation.

Thus far we have given only the action for the bosonic fields on the brane. For the leading fluctuations around a flat D-brane in flat spacetime the fermionic action is of the usual Dirac form

$$-i \int d^{p+1}\xi \, \mathrm{Tr}(\bar{\lambda}\Gamma^a D_a \lambda) . \tag{13.3.19}$$

The full nonlinear supersymmetric form is left to the references.

Coupling constants

The ratio of the F-string tension to the D-string tension is

$$\frac{\tau_{\mathrm{F1}}}{\tau_{\mathrm{D1}}} = \frac{1}{2\pi\alpha'} \frac{\kappa}{4\pi^{5/2}\alpha'} = \frac{\kappa}{8\pi^{7/2}\alpha'^2} . \tag{13.3.20}$$

Up to now there has been no natural convention for defining the additive normalization of the dilaton field or the multiplicative normalization of the closed string coupling $g = e^{\Phi}$. The dimensionless ratio (13.3.20) is proportional to the closed string coupling, and it turns out to be very convenient to take it as the definition of the coupling,

$$g = \frac{\tau_{\mathrm{F1}}}{\tau_{\mathrm{D1}}} . \tag{13.3.21}$$

Then the gravitational coupling is

$$\kappa^2 = \tfrac{1}{2}(2\pi)^7 g^2 \alpha'^4 \tag{13.3.22}$$

and the D-brane tension is

$$\tau_p = \frac{1}{g(2\pi)^p \alpha'^{(p+1)/2}} = (2\kappa^2)^{-1/2}(2\pi)^{(14-p)/2}\alpha'^{(3-p)/2} . \tag{13.3.23}$$

Also, the constant appearing in the low energy actions of section 12.1 is

$$\kappa_{10}^2 = \tfrac{1}{2}(2\pi)^7 \alpha'^4 ; \tag{13.3.24}$$

this differs from the physically measured κ because the latter depends on the dilaton background.

Expanding the action (13.3.14) gives the coupling of the Yang–Mills theory on the Dp-brane,

$$g_{\mathrm{D}p}^2 = \frac{1}{(2\pi\alpha')^2 \tau_p} = (2\pi)^{p-2} g \alpha'^{(p-3)/2} . \tag{13.3.25}$$

Notice that for $p = 3$ this coupling is dimensionless, as expected in a $(3+1)$-dimensional gauge theory. For $p < 3$ the coupling has units of length to a negative power, and for $p > 3$ length to a positive power.

We now wish to obtain the relation among κ, g_{YM}, and α' in the type I theory. We cannot quite identify $g_{\mathrm{D}p}$ for $p = 9$ with g_{YM}, because the former has been obtained in a locally oriented theory and there are some additional factors of 2 in the type I case. Rather than repeat the string calculation we will make a more roundabout but possibly instructive argument using T-duality.

First, we should note that the coupling (13.3.25) is for the $U(n)$ gauge theory of coincident branes in the oriented theory: it appears in the form

$$\frac{1}{4g_{\mathrm{D}p}^2} \mathrm{Tr}_{\mathrm{f}} , \tag{13.3.26}$$

where the trace is in the $n \times n$ fundamental representation. Now let us consider moving the branes to an orientifold plane so that the gauge symmetry is enlarged to $SO(2n)$. An $SU(n)$ generator t is embedded in $SO(2n)$ as

$$\tilde{t} = \begin{bmatrix} t & 0 \\ 0 & -t^T \end{bmatrix} , \tag{13.3.27}$$

because the orientation projection reverses the order of the Chan–Paton factors and the sign of the gauge field. Comparing the low energy actions gives

$$\frac{1}{4g_{\mathrm{D}p}^2} \mathrm{Tr}_{\mathrm{f}}(t^2) = \frac{1}{4g_{\mathrm{D}p,SO(2n)}^2} \mathrm{Tr}_{\mathrm{v}}(\tilde{t}^2) \tag{13.3.28}$$

and so $g_{\mathrm{D}p,SO(2n)}^2 = 2g_{\mathrm{D}p}^2$.

Now consider the type I theory compactified on a k-torus with all radii equal to R. The couplings in the lower-dimensional $SO(32)$ theory are related to those in the type I theory by

$$\kappa_{9-k}^2 = (2\pi R)^{-k}\kappa^2 \text{ (type I)} , \quad g_{9-k,\mathrm{YM}}^2 = (2\pi R)^{-k} g_{\mathrm{YM}}^2 \text{ (type I)} . \tag{13.3.29}$$

In the T-dual picture, the bulk theory is of type II and the gauge fields live on a D$(p-k)$-brane, and

$$\kappa_{9-k}^2 = 2(2\pi R')^{-k}\kappa^2 , \quad g_{9-k,\mathrm{YM}}^2 = g_{9-k,SO(32)}^2 . \tag{13.3.30}$$

The dimensional reduction for κ_{9-k}^2 has an extra factor of 2 because the compact space is an orientifold, its volume halved. The gauge coupling is

independent of the volume because the fields are localized on the D-brane. Combining these results with the relations (13.3.22) and (13.3.25) gives, independent of k, the type I relation

$$\frac{g_{\rm YM}^2}{\kappa} = 2(2\pi)^{7/2}\alpha' \quad \text{(type I)} . \tag{13.3.31}$$

As one final remark, the Born–Infeld form for the gauge action applies by T-duality to the type I theory:

$$S = -\frac{1}{(2\pi\alpha')^2 g_{\rm YM}^2}\int d^{10}x\,{\rm Tr}\left\{[-\det(\eta_{\mu\nu} + 2\pi\alpha' F_{\mu\nu})]^{1/2}\right\} , \tag{13.3.32}$$

whose normalization is fixed by the quadratic term in F. In the previous chapter we obtained the tree-level string correction (12.4.28) to the type I effective action. If the gauge field lies in an Abelian subgroup, the tensor structure simplifies to

$$\frac{(2\pi\alpha')^2}{32 g_{\rm YM}^2}{\rm Tr}_{\rm v}\left(4F_{\mu\nu}F^{\nu\sigma}F_{\sigma\rho}F^{\rho\mu} - F_{\mu\nu}F^{\nu\mu}F_{\sigma\rho}F^{\rho\sigma}\right) . \tag{13.3.33}$$

This is indeed the quartic term in the expansion of the Born–Infeld action, as one finds by using

$$\det{}^{1/2}(1+M) = \exp\left[\frac{1}{2}{\rm tr}\left(M - \frac{1}{2}M^2 + \frac{1}{3}M^3 - \frac{1}{4}M^4 + \ldots\right)\right] \tag{13.3.34}$$

with $M_\mu^{\ \nu} = 2\pi\alpha' F_{\mu\sigma}\eta^{\sigma\nu}$. The trace here is on the Lorentz indices, and ${\rm tr}(x^{2k+1}) = 0$ for antisymmetric x. Note that only when the gauge field can be diagonalized can we give a geometric interpretation to the T-dual configuration and so derive the Born–Infeld form.

13.4 D-brane interactions: statics

Many interesting new issues arise with D-branes that are not parallel, or are of different dimensions. In this section we focus on static questions. The first of these concerns the breaking of supersymmetry. Let us consider a Dp-brane and a Dp'-brane, which we take first to be aligned along the coordinate axes. That is, we can partition the spacetime directions μ into two sets $S_{\rm D}$ and $S_{\rm N}$ according to whether the coordinate X^μ has Dirichlet or Neumann boundary conditions on the first D-brane, and similarly into two sets $S'_{\rm D}$ and $S'_{\rm N}$ depending on the alignment of the second D-brane. The *DD coordinates* are $S_{\rm D}\cap S'_{\rm D}$, the *ND coordinates* are $S_{\rm N}\cap S'_{\rm D}$, and so on.

The first D-brane leaves unbroken the supersymmetries

$$Q_\alpha + (\beta^\perp\tilde{Q})_\alpha , \quad \beta^\perp = \prod_{m\in S_{\rm D}}\beta^m . \tag{13.4.1}$$

Similarly the second D-brane leaves unbroken

$$Q_\alpha + (\beta^{\perp\prime}\tilde{Q})_\alpha = Q_\alpha + [\beta^\perp(\beta^{\perp-1}\beta^{\perp\prime})\tilde{Q}]_\alpha , \quad \beta^{\perp\prime} = \prod_{m\in S'_{\rm D}}\beta^m . \tag{13.4.2}$$

The complete state is invariant only under supersymmetries that are of both forms (13.4.1) and (13.4.2). These are in one-to-one correspondence with those spinors left invariant by $\beta^{\perp-1}\beta^{\perp\prime}$. The operator $\beta^{\perp-1}\beta^{\perp\prime}$ is a reflection in the DN and ND directions. Let us denote the total number of DN and ND directions $\#_{\rm ND}$. Since $p-p'$ is always even the number $\#_{\rm ND} = 2j$ is also even. We can then pair these dimensions and write $(\beta^\perp)^{-1}\beta^{\perp\prime}$ as a product of rotations by π,

$$\beta \equiv (\beta^\perp)^{-1}\beta^{\perp\prime} = \exp[\pi i(J_1 + \ldots + J_j)] . \tag{13.4.3}$$

In a spinor representation, each $\exp(i\pi J)$ has eigenvalues $\pm i$, so there will be unbroken supersymmetries only if j is even and so $\#_{\rm ND}$ is a multiple of 4. In this case there are 8 unbroken supersymmetries, one quarter of the original 32. Note that T-duality switches NN↔DD and ND↔DN and so leaves $\#_{\rm ND}$ invariant. When $\#_{\rm ND} = 0$, then $(\beta^\perp)^{-1}\beta^{\perp\prime} = 1$ identically and there are 16 unbroken supersymmetries. This is the same as for the original type I theory, to which it is T-dual.

An open string can have both ends on the same D-brane or one on each. The p-p and p'-p' spectra are the same as obtained before by T-duality from the type I string, but the p-p' strings are new. Each of the four possible boundary conditions can be written with the doubling trick

$$X^\mu(w,\bar{w}) = X^\mu(w) + \tilde{X}^\mu(\bar{w}) \tag{13.4.4}$$

in terms of one of two mode expansions,

$$X^\mu(w) = x^\mu + \left(\frac{\alpha'}{2}\right)^{1/2}\left[-\alpha_0^\mu w + i\sum_{\substack{m\in{\bf Z}\\ m\neq 0}}\frac{\alpha_m^\mu}{m}\exp(imw)\right] , \tag{13.4.5a}$$

$$X^\mu(w) = i\left(\frac{\alpha'}{2}\right)^{1/2}\sum_{r\in{\bf Z}+1/2}\frac{\alpha_r^\mu}{r}\exp(irw) . \tag{13.4.5b}$$

The periodic expansion (13.4.5a) describes NN strings for

$$\tilde{X}^\mu(\bar{w}) = X^\mu(2\pi - \bar{w}) \tag{13.4.6}$$

and DD strings for

$$\tilde{X}^\mu(\bar{w}) = -X^\mu(2\pi - \bar{w}) . \tag{13.4.7}$$

The antiperiodic expansion (13.4.5b) similarly defines DN and ND strings, with $\tilde{X}^\mu(\bar{w}) = \pm X^\mu(2\pi - w)$. For ψ^μ, the periodicity in the R sector is the same as for X^μ because T_F is periodic. In the NS sector it is the opposite.

The string zero-point energy is zero in the R sector as always, because bosons and fermions with the same periodicity cancel. In the NS sector it is

$$(8-\#_{\rm ND})\left(-\frac{1}{24}-\frac{1}{48}\right)+\#_{\rm ND}\left(\frac{1}{48}+\frac{1}{24}\right)=-\frac{1}{2}+\frac{\#_{\rm ND}}{8}\,. \qquad (13.4.8)$$

The oscillators can raise the level in half-integer units, so only for $\#_{\rm ND}$ a multiple of 4 is degeneracy between the R and NS sectors possible. This agrees with the analysis above: the $\#_{\rm ND}=2$ and $\#_{\rm ND}=6$ systems cannot be supersymmetric. Later we will see that there are supersymmetric *bound states* when $\#_{\rm ND}=2$, but their description is rather different.

Branes at general angles

It is interesting to consider the case of D-branes at general angles to one another. To be specific consider two D4-branes. Let both initially be extended in the (2,4,6,8)-directions, and separated by some distance y_1 in the 1-direction. Now rotate one of them by an angle ϕ_1 in the (2,3) plane, ϕ_2 in the (4,5) plane, and so on; call this rotation ρ. The supersymmetry unbroken by the rotated 4-brane is

$$Q_\alpha+(\rho^{-1}\beta^\perp\rho\tilde{Q})_\alpha\,. \qquad (13.4.9)$$

Supersymmetries left unbroken by both branes then correspond to spinors left invariant by

$$(\beta^\perp)^{-1}\rho^{-1}\beta^\perp\rho=(\beta^\perp)^{-1}\beta^\perp\rho^2=\rho^2\,. \qquad (13.4.10)$$

In the usual s-basis the eigenvalues of ρ^2 are

$$\exp\left(2i\sum_{a=1}^{4}s_a\phi_a\right)\,. \qquad (13.4.11)$$

In the **16** the $(2s_1,2s_2,2s_3,2s_4)$ run over all 16 combinations of ±1s; each combination such that the phase (13.4.11) is 1 gives an unbroken supersymmetry. There are many possibilities — for example:

- For generic ϕ_a there are no unbroken supersymmetries.
- For angles $\phi_1+\phi_2+\phi_3+\phi_4=0$ mod 2π (but otherwise generic) there are two unbroken supersymmetries, namely those with $s_1=s_2=s_3=s_4$. The rotated D4-brane breaks seven-eighths of the supersymmetry of the first.
- For $\phi_1+\phi_2+\phi_3=\phi_4=0$ mod 2π there are four unbroken supersymmetries.
- For $\phi_1+\phi_2=\phi_3+\phi_4=0$ mod 2π there are four unbroken supersymmetries.

- For $\phi_1+\phi_2=\phi_3=\phi_4=0$ mod 2π there are eight unbroken supersymmetries.

Also, when k angles are $\pi/2$ and the rest are zero this reduces to the earlier analysis with $\#_{\rm ND}=2k$.

For later reference let us also describe these results as follows. Join the coordinates into complex pairs, $Z^1=X^2+iX^3$ and so on, with the conjugate $\overline{Z^a}$ denoted $Z^{\bar a}$. Then ρ takes Z^a to $\exp(i\phi_a)Z^a$. The $SO(8)$ rotation group on the transverse dimensions has a $U(4)$ subgroup that preserves the complex structure. That is, it rotates $Z'^a=U^{ab}Z^b$, whereas a general $SO(8)$ rotation would mix in $Z^{\bar b}$ as well. The rotation ρ in particular is the $U(4)$ matrix

$$\mathrm{diag}\left[\exp(i\phi_1),\exp(i\phi_2),\exp(i\phi_3),\exp(i\phi_4)\right]\,. \qquad (13.4.12)$$

When $\phi_1+\phi_2+\phi_3+\phi_4=0$ mod 2π, which is the condition for two supersymmetries to be unbroken, the determinant of ρ is 1 and so it actually lies in the $SU(4)$ subgroup of $U(4)$. Then we can summarize the above by saying that a general $U(4)$ rotation breaks all the supersymmetry, an $SU(4)$ rotation breaks seven-eighths, an $SU(3)$ or $SU(2)\times SU(2)$ rotation breaks three-quarters, and an $SU(2)$ rotation half. Further, if we consider several branes, so that in general the rotations ρ_i cannot be simultaneously diagonalized, then as long as all of them lie within a given subgroup the number of unbroken supersymmetries is as above.

Now let us calculate the force between these rotated branes. The cylinder graph involves traces over the p-p' strings, so we need to generalize the mode expansion to the rotated case. Letting the $\sigma^1=0$ endpoint be on the unrotated brane and the $\sigma^1=\pi$ endpoint on the rotated brane, it follows that the boundary conditions are

$$\sigma_1=0:\quad \partial_1\mathrm{Re}(Z^a)=\mathrm{Im}(Z^a)=0\,, \qquad (13.4.13a)$$

$$\sigma_1=\pi:\quad \partial_1\mathrm{Re}[\exp(i\phi_a)Z^a]=\mathrm{Im}[\exp(i\phi_a)Z^a]=0\,. \qquad (13.4.13b)$$

These are satisfied by

$$\begin{aligned}Z^a(w,\bar w)&=\mathscr{Z}^a(w)+\overline{\mathscr{Z}^a}(-\bar w)\,,\\&=\exp(-2i\phi_a)\mathscr{Z}^a(w+2\pi)+\overline{\mathscr{Z}^a}(-\bar w)\,,\end{aligned} \qquad (13.4.14)$$

where $w=\sigma^1+i\sigma^2$. This implies the mode expansion

$$\mathscr{Z}^a(w)=i\left(\frac{\alpha'}{2}\right)^{1/2}\sum_{r\in\mathbf{Z}+\nu_a}\frac{\alpha^a_r}{r}\exp(irw)\,, \qquad (13.4.15)$$

with $\nu_a=\phi_a/\pi$. The modes $(\alpha^a_r)^\dagger$ are linearly independent.

The partition function for one such complex scalar is

$$q^{E_0}\prod_{m=0}^{\infty}\left[1-q^{m+(\phi/\pi)}\right]^{-1}\left[1-q^{m+1-(\phi/\pi)}\right]^{-1}=-i\frac{\exp(\phi^2 t/\pi)\eta(it)}{\vartheta_{11}(i\phi t/\pi,it)} \quad (13.4.16)$$

with $q=\exp(-2\pi t)$ and

$$E_0=\frac{1}{24}-\frac{1}{2}\left(\frac{\phi}{\pi}-\frac{1}{2}\right)^2 . \quad (13.4.17)$$

The definitions and properties of theta functions are collected in section 7.2, but we reproduce here the results that will be most useful:

$$\vartheta_{11}(v,it)=-2q^{1/8}\sin\pi v\prod_{m=1}^{\infty}(1-q^m)(1-zq^m)(1-z^{-1}q^m) , \quad (13.4.18a)$$

$$\vartheta_{11}(-iv/t,i/t)=-it^{1/2}\exp(\pi v^2/t)\vartheta_{11}(v,it) , \quad (13.4.18b)$$

where $z=\exp(2\pi iv)$. Similarly in each of the sectors of the fermionic path integral one replaces the $Z^{\alpha}{}_{\beta}(it)$ that appears for parallel D-branes with[4]

$$Z^{\alpha}{}_{\beta}(\phi,it)=\frac{\vartheta_{\alpha\beta}(i\phi t/\pi,it)}{\exp(\phi^2 t/\pi)\eta(it)} . \quad (13.4.19)$$

The full fermionic partition function is

$$\frac{1}{2}\left[\prod_{a=1}^{4}Z^0{}_0(\phi_a,it)-\prod_{a=1}^{4}Z^0{}_1(\phi_a,it)-\prod_{a=1}^{4}Z^1{}_0(\phi_a,it)-\prod_{a=1}^{4}Z^1{}_1(\phi_a,it)\right] , \quad (13.4.20)$$

generalizing the earlier $Z^{+}_{\psi}(it)$. By a generalization of the abstruse identity (7.2.41), the fermionic partition function can be rewritten

$$\prod_{a=1}^{4}Z^1{}_1(\phi'_a,it) , \quad (13.4.21)$$

where

$$\phi'_1=\frac{1}{2}(\phi_1+\phi_2+\phi_3+\phi_4) , \quad \phi'_2=\frac{1}{2}(\phi_1+\phi_2-\phi_3-\phi_4) , \quad (13.4.22a)$$

$$\phi'_3=\frac{1}{2}(\phi_1-\phi_2+\phi_3-\phi_4) , \quad \phi'_4=\frac{1}{2}(\phi_1-\phi_2-\phi_3+\phi_4) . \quad (13.4.22b)$$

This identity has a simple physical origin. If we refermionize, writing the theory in terms of the free fields θ_α as in eq. (12.6.24), we get the form (13.4.21) directly. In particular, the $\exp(2i\phi'_a)$ are the eigenvalues of ρ in the **8** of $SO(8)$.

Collecting all factors, the potential is

$$V=-\int_0^{\infty}\frac{dt}{t}(8\pi^2\alpha' t)^{-1/2}\exp\left(-\frac{ty_1^2}{2\pi\alpha'}\right)\prod_{a=1}^{4}\frac{\vartheta_{11}(i\phi'_a t/\pi,it)}{\vartheta_{11}(i\phi_a t/\pi,it)} . \quad (13.4.23)$$

Note that for nonzero angles the stretched strings are confined near the point of closest approach of the two 4-branes. The function $\vartheta_{11}(v,it)$ is odd in v and so vanishes when $v=0$. If any of the ϕ_a vanish the denominator has a zero. This is because the 4-branes become parallel in one direction and the strings are then free to move in that direction. One must replace

$$i\vartheta_{11}(i\phi_a t/\pi,it)^{-1}\to L\eta(it)^{-3}(8\pi^2\alpha' t)^{-1/2} . \quad (13.4.24)$$

This gives the usual factors for a noncompact direction, L being the length of the spatial box. Taking $\phi_4\to 0$ so the 4-branes both run in the 8-direction, one can T-dualize in this direction to get a pair of 3-branes with relative rotations in three planes. The fermionic partition function is unaffected, while the factors (13.4.24) are instead replaced by

$$\eta(it)^{-3}\exp\left[-\frac{t(y_8^2+y_9^2)}{2\pi\alpha'}\right] , \quad (13.4.25)$$

allowing for the possibility of a separation in the (8,9) plane. Taking the T-dual in the 9-direction instead one obtains 5-branes that are separated in the 1-direction, extended in the (8,9)-directions, and with relative rotations in the other three planes. The effect is an additional factor of $L_9(8\pi^2\alpha' t)^{-1/2}$. The extension to other p is straightforward.

If instead any of the ϕ'_a vanishes, the potential is zero. The reason is that there is unbroken supersymmetry: the phases (13.4.11) include $\exp(\pm 2i\phi'_a)$. Curiously this covers only eight of the sixteen phases (13.4.11), so that if some phases (13.4.11) are unity but not those of the form $\exp(\pm 2i\phi'_a)$, then supersymmetry is unbroken but the potential is nonzero. This is an exception to the usual rule that the vacuum loop amplitudes vanish by Bose–Fermi cancellation. The rotated D-branes leave only two supersymmetries unbroken, so that BPS multiplets of open strings contain a single bosonic or fermionic state.

The potential is a complicated function of position, but at long distance it simplifies. The exponential factor in the integral (13.4.23) forces t to be small, and then the ϑ-functions simplify,

$$\prod_{a=1}^{4}\frac{\vartheta_{11}(i\phi'_a t/\pi,it)}{\vartheta_{11}(i\phi_a t/\pi,it)}\to\prod_{a=1}^{4}\frac{\sin\phi'_a}{\sin\phi_a} , \quad (13.4.26)$$

by using the modular transformation of ϑ_{11}. The t-integral then gives a power of the separation y_1. The result agrees with the low energy field

[4] If one applies the formalism of the previous chapter, in the (P,P) spin structure there are two $\beta\gamma$ zero modes and two longitudinal ψ zero modes for a net $0^2/0^2$. One can define this by a more careful gauge fixing, or equivalently by adding a graviton vertex operator (which allows all the zero modes to be soaked up) and relating the zero-momentum graviton coupling to the potential. However, we simply rely on the physical input of the Coleman–Weinberg formula.

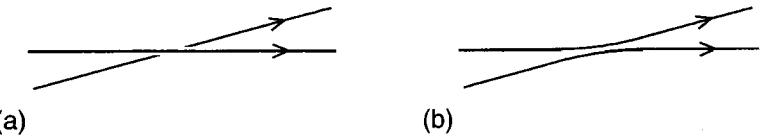


Fig. 13.4. (a) D-branes at relative angle. (b) Lower energy configuration.

theory calculation, including the angular factor. For 4-branes with all ϕ_a nonzero the potential grows linearly with y_1 at large distance, for 3-branes with all ϕ_a nonzero it falls as $1/y_1$, and so on.

In nonsupersymmetric configurations a tachyon can appear. For simplicity let only ϕ_1 be nonzero, with $0 \le \phi_1 \le \pi$. The NS ground state energy is $-(1/2)+(\phi_1/2\pi)$, and the first excited state $\psi_{-(1/2)+(\phi_1/\pi)}|0\rangle_{\rm NS}$, which survives the GSO projection, has weight $-\phi_1/2\pi$. Including the energy from tension, the lightest state has

$$m^2 = \frac{y_1^2}{4\pi^2\alpha'^2} - \frac{2\phi_1}{\pi\alpha'}\,, \quad 0 \le \phi_1 \le \pi\,. \tag{13.4.27}$$

This is negative if the separation is small enough. A special case is $\phi_1 = \pi$, when the 4-branes are antiparallel rather than parallel. The NS–NS and R–R exchanges are then both attractive, and below the critical separation $y_1^2 = 8\pi^2\alpha'$ the cylinder amplitude diverges as $t \to \infty$. This is where the tachyon appears — evidently it represents D4-brane/anti-D4-brane annihilation. Even when the D-branes are nearly parallel they can lower their energy by reconnecting as in figure 13.4(b), and this is the origin of the instability. This is one example where the tachyon has a simple physical interpretation and we can see that the decay has no end: the reconnected strings move apart indefinitely. On the other hand, for the same instability but with the strings wound on a two-torus there is a lower bound to the energy.

13.5 D-brane interactions: dynamics

D-brane scattering

For parallel static D-branes the potential energy is zero, but if they are in relative motion all supersymmetry is broken and there is a velocity-dependent force. This can be obtained by an analytic continuation of the static potential for rotated branes. Consider the case that only ϕ_1 is nonzero, so the rotated brane satisfies $X^3 = X^2 \tan\phi_1$. Analytically

continue $X^2 \to iX'^0$ and let $\phi_1 = -iu$. Then

$$X^3 = X'^0 \tanh u\,, \tag{13.5.1}$$

which describes a D-brane moving with constant velocity. Continue also $X^0 \to -iX'^2$ to eliminate the spurious extra time coordinate. The interaction amplitude (13.4.23) between the D-branes becomes

$$\mathscr{A} = -iV_p \int_0^\infty \frac{dt}{t}(8\pi^2\alpha' t)^{-p/2} \exp\left(-\frac{ty_1^2}{2\pi\alpha'}\right) \frac{\vartheta_{11}(ut/2\pi, it)^4}{\eta(it)^9\vartheta_{11}(ut/\pi, it)}\,, \tag{13.5.2}$$

where we have extended the result to general p by using T-duality.[5] It is also useful to give the modular transformation

$$\mathscr{A} = \frac{V_p}{(8\pi^2\alpha')^{p/2}} \int_0^\infty \frac{dt}{t} t^{(6-p)/2} \exp\left(-\frac{ty_1^2}{2\pi\alpha'}\right) \frac{\vartheta_{11}(iu/2\pi, i/t)^4}{\eta(i/t)^9\vartheta_{11}(iu/\pi, i/t)}\,. \tag{13.5.3}$$

We can write this as an integral over the world-line,

$$\mathscr{A} = -i\int_{-\infty}^{\infty} d\tau\, V(r(\tau), v)\,, \tag{13.5.4}$$

where

$$r(\tau)^2 = y_1^2 + v^2\tau^2\,, \quad v = \tanh u\,, \tag{13.5.5}$$

and

$$V(r,v) = i\frac{2V_p}{(8\pi^2\alpha')^{(p+1)/2}} \int_0^\infty dt\, t^{(5-p)/2} \times \exp\left(-\frac{tr^2}{2\pi\alpha'}\right) \frac{(\tanh u)\vartheta_{11}(iu/2\pi, i/t)^4}{\eta(i/t)^9\vartheta_{11}(iu/\pi, i/t)}\,. \tag{13.5.6}$$

The interaction has a number of interesting properties. The first is that as $v \to 0$ (so that $u \to 0$), it vanishes as v^4 from the zeros of the theta functions. We expect only even powers of v by time-reversal invariance. The vanishing of the v^2 interaction, like the vanishing of the static interaction, is a consequence of supersymmetry. The low energy field theory of the D-branes is a $U(1) \times U(1)$ supersymmetric gauge theory with 16 supersymmetries. What we are calculating is a correction to the effective action from integrating out massive states, strings stretched between the D-branes. The vanishing of the v^2 term is then consistent with the assertion in section B.6 that with 16 supersymmetries corrections to the kinetic term are forbidden — the moduli space is flat. If we had instead taken $\phi_3 = \phi_4 = \pi/2$ so that $\#_{\rm ND} = 4$, there would only be two zeros in the numerator and thus a v^2 interaction. This is consistent with

[5] We find it difficult to keep track of the sign during the continuation, but it is easily checked by looking at the contribution of NS–NS exchange in the static limit. Note that the ϑ_{11} are negative for small positive u.

the result that corrections to the kinetic term are allowed when there are eight unbroken supersymmetries.

The interaction (13.5.6) is in general a complicated function of the separation, but in an expansion in powers of the velocity the leading $O(v^4)$ term is simple,

$$V(r,v) = -v^4 \frac{2V_p}{(8\pi^2\alpha')^{(p+1)/2}} \int_0^\infty dt\, t^{(5-p)/2} \exp\left(-\frac{tr^2}{2\pi\alpha'}\right) + O(v^6)$$
$$= -\frac{v^4}{r^{7-p}} \frac{V_p}{\alpha'^{p-3}} 2^{2-2p} \pi^{(5-3p)/2} \Gamma\left(\frac{7-p}{2}\right) + O(v^6) \,. \qquad (13.5.7)$$

At long distances this is in agreement with low energy supergravity. It is also the leading behavior if we expand in powers of $1/r$ rather than v.

In general the behavior of $V(r,v)$ as $r \to 0$ is quite different from the behavior as $r \to \infty$. The r-dependence of the integral (13.5.6) arises from the factor $\exp(-tr^2/2\pi\alpha')$, so that $t \approx 2\pi\alpha'/r^2$ governs the behavior at given r. Large r corresponds to small t, where the asymptotic behavior is given by tree-level exchange of light closed strings — hence the agreement with classical supergravity. Small r corresponds to large t, where the asymptotic behavior is given by a loop of the light open strings. The cross-over is at $r^2 \sim 2\pi\alpha'$. This is as we expect: string theory modifies gravity at distances below the string scale.

This simple r-dependence of the v^4 term is another consequence of supersymmetry. The fact that this term is singular as $r \to 0$ might seem to conflict with the assertion that string theory provides a short-distance cutoff. However, one must look more carefully. To obtain the small-r behavior of the scattering amplitude (13.5.6), take the large-t limit without expanding in v to obtain

$$V(r,v) \approx -2V_{p+1} \int_0^\infty \frac{dt}{(8\pi^2\alpha' t)^{(p+1)/2}} \exp\left(-\frac{tr^2}{2\pi\alpha'}\right) \frac{\tanh u \, \sin^4 ut/2}{\sin ut} \,. \qquad (13.5.8)$$

Since $t \approx 2\pi\alpha'/r^2$ and $v \approx u$, the arguments of the sines are $ut \approx 2\pi\alpha' v/r^2$. No matter how small v is the v^4 term will cease to dominate at small enough r. The oscillations of the integrand then smooth the small-r behavior on a scale $ut \approx 1$. The effective scale probed by the scattering is

$$r \approx \alpha'^{1/2} v^{1/2} \,. \qquad (13.5.9)$$

A small-velocity D-brane probe is thus sensitive to distances shorter than the string scale. This is in contrast to the behavior we have seen in string scattering at weak coupling, but fits nicely with the understanding of strongly coupled strings in the next chapter.

Let us expand on this result. A slower D-brane probes shorter distances,

but the scattering process takes longer, $\delta t \approx r/v$. Then

$$\delta x \, \delta t \gtrsim \alpha' \,. \qquad (13.5.10)$$

This is a suggestion for a new uncertainty relation involving only the coordinates. It is another indication of 'noncommutative geometry,' perhaps connected with the promotion of D-brane collective coordinates to matrices.

For a pointlike D0-brane probe there is a minimum distance that can be measured by scattering. The wavepacket in which it is prepared satisfies

$$\delta x \gtrsim \frac{1}{m\delta v} = \frac{g\alpha'^{1/2}}{\delta v} \,. \qquad (13.5.11)$$

The combined uncertainties (13.5.9) and (13.5.11) are minimized by $v \approx g^{2/3}$, for which

$$\delta x \gtrsim g^{1/3} \alpha'^{1/2} \,. \qquad (13.5.12)$$

We will see the significance of this scale in the next chapter.

D0-brane quantum mechanics

The nonrelativistic effective Lagrangian for n D0-branes is

$$L = \mathrm{Tr}\left\{ \frac{1}{2g\alpha'^{1/2}} D_0 X^i D_0 X^i + \frac{1}{4g\alpha'^{1/2}(2\pi\alpha')^2} [X^i, X^j]^2 \right.$$
$$\left. - \frac{i}{2} \lambda D_0 \lambda + \frac{1}{4\pi\alpha'} \lambda \Gamma^0 \Gamma^i [X^i, \lambda] \right\} \,. \qquad (13.5.13)$$

The first term is the usual nonrelativistic kinetic energy with $m = \tau_0 = 1/g\alpha'^{1/2}$, dropping the constant rest mass $n\tau_0$. The coefficients of the other terms are most easily obtained by T-duality from the ten-dimensional super-Yang–Mills action (B.6.13), with $A^i \to X^i/2\pi\alpha'$. We have taken a basis in which the fermionic field λ is Hermitean, and rescaled λ to obtain a canonical kinetic term. The index i runs over the nine spatial dimensions. The gauge field A^0 has no kinetic term but remains in the covariant derivatives. It couples to the $U(n)$ charges, so its equation of motion amounts to the constraint that only $U(n)$-invariant states are allowed. Only terms with at most two powers of the velocity have been kept, not the full Born–Infeld action.

The Hamiltonian is

$$H = \mathrm{Tr}\left\{ \frac{g\alpha'^{1/2}}{2} p_i p_i - \frac{1}{16\pi^2 g\alpha'^{5/2}} [X^i, X^j]^2 - \frac{1}{4\pi\alpha'} \lambda \Gamma^0 \Gamma^i [X^i, \lambda] \right\} \,. \qquad (13.5.14)$$

Note that the potential is positive because $[X^i, X^j]$ is anti-Hermitean. The canonical momentum, like the coordinate, is a matrix,

$$[p_{iab}, X^j_{cd}] = -i\delta^j_i \delta_{ad} \delta_{bc} \,. \qquad (13.5.15)$$

Now we define

$$X^i = g^{1/3}(2\pi\alpha')^{1/2}Y^i \,, \tag{13.5.16}$$

so that also $p_i = p_{Yi}/g^{1/3}\alpha'^{1/2}$. The Hamiltonian becomes

$$H = \frac{g^{1/3}}{\alpha'^{1/2}}\mathrm{Tr}\left\{\frac{1}{2}p_{Yi}p_{Yi} - \frac{1}{16\pi^2}[Y^i, Y^j]^2 - \frac{1}{4\pi}\lambda\Gamma^0\Gamma^i[Y^i,\lambda]\right\} . \tag{13.5.17}$$

The parameters g and α' now appear only in the overall normalization. It follows that the wavefunctions are independent of the parameters when expressed in terms of the variables Y^i. In terms of the original coordinates X^i their characteristic size scales as $g^{1/3}\alpha'^{1/2}$, the same scale (13.5.12) found above. The energies scale as $g^{1/3}/\alpha'^{1/2}$ from the overall normalization of H, and the characteristic time scale as the inverse of this, so we find again the relation (13.5.10).

Recall from the discussion of D-brane scattering that at distances less than the string scale only the lightest open string states (those which become massless when the D-branes are coincident) contribute. In this regime the cylinder amplitude reduces to a loop amplitude in the low energy field theory (13.5.13).

The $\#_{\mathrm{ND}} = 4$ system

Another low energy action with many applications is that for a Dp-brane and Dp'-brane with relative $\#_{\mathrm{ND}} = 4$. There are three kinds of light strings: p-p, p-p', and p'-p', with ends on the respective D-branes. We will consider explicitly the case $p = 5$ and $p' = 9$, where we can take advantage of the $SO(5,1) \times SO(4)$ spacetime symmetry; all other cases are related to this by T-duality.

The 5-5 and 9-9 strings are the same as those that arise on a single D-brane. The new feature is the 5-9 strings; let us study their massless spectrum. The NS zero-point energy is zero. The moding of the fermions differs from that of the bosons by $\frac{1}{2}$, so there are four periodic world-sheet fermions ψ^m, namely those in the ND directions $m = 6,7,8,9$. The four zero modes then generate $2^{4/2} = 4$ degenerate ground states, which we label by their spins in the (6,7) and (8,9) planes,

$$|s_3, s_4\rangle_{\mathrm{NS}} \,, \tag{13.5.18}$$

with s_3, s_4 taking values $\pm\frac{1}{2}$. Now we need to impose the GSO projection. This was defined in eq. (10.2.22) in terms of s_a, so that with the extra sign from the ghosts it is

$$-\exp[\pi i(s_3 + s_4)] = +1 \;\Rightarrow\; s_3 = s_4 \,. \tag{13.5.19}$$

In terms of the symmetries, the four states (13.5.18) are invariant under $SO(5,1)$ and form spinors $\mathbf{2} + \mathbf{2}'$ of the 'internal' $SO(4)$, and only the $\mathbf{2}$ survives the GSO projection. In the R sector, of the transverse fermions ψ^i only those with $i = 2,3,4,5$ are periodic, so there are again four ground states

$$|s_1, s_2\rangle_{\mathrm{R}} \,. \tag{13.5.20}$$

The GSO projection does not have a extra sign in the R sector so it requires $s_1 = -s_2$. The surviving spinors are invariant under the internal $SO(4)$ and form a $\mathbf{2}'$ of the $SO(4)$ little group of a massless particle.

The system has six-dimensional Lorentz invariance and eight unbroken supersymmetries, so we can classify it by $d = 6$, $N = 1$ supersymmetry (section B.7). The massless content of the 5-9 spectrum amounts to half of a hypermultiplet. The other half comes from strings of opposite orientation, 9-5. The action is fully determined by supersymmetry and the charges; we write the bosonic part:

$$S = -\frac{1}{4g_{\mathrm{D9}}^2}\int d^{10}x\, F_{MN}F^{MN} - \frac{1}{4g_{\mathrm{D5}}^2}\int d^6x\, F'_{MN}F'^{MN} - \int d^6x\left[D_\mu\chi^\dagger D^\mu\chi + \frac{g_{\mathrm{D5}}^2}{2}\sum_{A=1}^{3}(\chi_i^\dagger\sigma^A_{ij}\chi_j)^2\right] . \tag{13.5.21}$$

The integrals run respectively over the 9-brane and the 5-brane, with $M = 0,\ldots,9$, $\mu = 0,\ldots,5$, and $m = 6,\ldots,9$. The covariant derivative is $D_\mu = \partial_\mu + iA_\mu - iA'_\mu$ with A_μ and A'_μ the 9-brane and 5-brane gauge fields. The field χ_i is a doublet describing the hypermultiplet scalars. The 5-9 strings have one endpoint on each D-brane so χ carries charges $+1$ and -1 under the respective symmetries. The gauge couplings $g_{\mathrm{D}p}$ were given in eq. (13.3.25). We are using a condensed notation,

$$A'_M \;\to\; A'_\mu \,,\; X'_m/2\pi\alpha' \,. \tag{13.5.22}$$

The massless 5-5 (and also 9-9) strings separate into $d = 6$, $N = 1$ vector and hypermultiplets. The final potential term is the 5-5 D-term required by the supersymmetry. One might have expected a 9-9 D-term as well by T-duality, but this is inversely proportional to the volume of the D9-brane in the (6,7,8,9)-directions, which we have taken to be infinite.

Under T-dualities in any of the ND directions, one obtains $(p,p') = (8,6)$, $(7,7)$, $(6,8)$, or $(5,9)$, but the intersection of the branes remains $(5+1)$-dimensional and the p-p' strings live on the intersection with action (13.5.21). T-dualities in r NN directions give $(p,p') = (9-r, 5-r)$. The vector components in the dualized directions become collective coordinates as usual,

$$A_i \to X_i/2\pi\alpha' \,, \quad A'_i \to X'_i/2\pi\alpha' \,. \tag{13.5.23}$$

The term $D_i\chi^\dagger D^i\chi$ then becomes

$$\left(\frac{X_i - X_i'}{2\pi\alpha'}\right)^2 \chi^\dagger\chi \,. \tag{13.5.24}$$

This just reflects the fact that when the $(9-r)$-brane and $(5-r)$-brane are separated, the strings stretched between them become massive.

The action for several branes of each type is given by the non-Abelian extension.

13.6 D-brane interactions: bound states

Bound states of D-branes with strings and with each other, and supersymmetric bound states in particular, present a number of interesting dynamical problems. Further, these bound states will play an important role in the next chapter in our attempts to deduce the strongly coupled behavior of string theory.

FD bound states

The first case we consider is a state with p F-strings and q D-strings in the IIB theory, all at rest and aligned along the 1-direction. For a state with these charges, the supersymmetry algebra (13.2.9) becomes

$$\frac{1}{2}\left\{\begin{bmatrix} Q_\alpha \\ \tilde{Q}_\alpha \end{bmatrix}, \begin{bmatrix} Q_\beta^\dagger & \tilde{Q}_\beta^\dagger \end{bmatrix}\right\} = M\delta_{\alpha\beta}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \frac{L_1}{2\pi\alpha'}(\Gamma^0\Gamma^1)_{\alpha\beta}\begin{bmatrix} p & q/g \\ q/g & -p \end{bmatrix}, \tag{13.6.1}$$

where L_1 is the length of the system. The eigenvalues of $\Gamma^0\Gamma^1$ are ± 1, so those of the right-hand side are

$$M \pm L_1\frac{(p^2 + q^2/g^2)^{1/2}}{2\pi\alpha'} \,. \tag{13.6.2}$$

The left-hand side of the algebra is positive — its expectation value in any state is a matrix with positive eigenvalues. This implies a BPS bound on the total energy per unit length,

$$\frac{M}{L_1} \geq \frac{(p^2 + q^2/g^2)^{1/2}}{2\pi\alpha'} \,. \tag{13.6.3}$$

This inequality is saturated by the F-string, which has $(p,q) = (1,0)$, and by the D-string, with $(p,q) = (0,1)$.

For one F-string and one D-string, the total energy per unit length is

$$\tau_{(0,1)} + \tau_{(1,0)} = \frac{g^{-1} + 1}{2\pi\alpha'} \,. \tag{13.6.4}$$

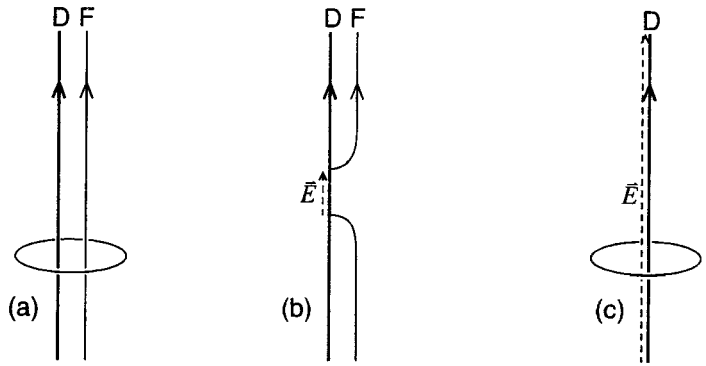


Fig. 13.5. (a) Parallel D-string and F-string. The loop signifies a 7-sphere surrounding the strings. (b) The F-string breaks, its ends attaching to the D-string. (c) Final state: D-string with flux.

This exceeds the BPS bound

$$\tau_{(1,1)} = \frac{(g^{-2} + 1)^{1/2}}{2\pi\alpha'} \,, \tag{13.6.5}$$

and so this configuration is not supersymmetric. One can also see this directly. The F-string is invariant under supersymmetries satisfying

$$\text{left-moving: } \Gamma^0\Gamma^1 Q = Q \,, \quad \text{right-moving: } \Gamma^0\Gamma^1\tilde{Q} = -\tilde{Q} \,, \tag{13.6.6}$$

and no linear combination of these is of the form $Q_\alpha + (\beta^\perp\tilde{Q})_\alpha$ preserved by the D-string (note that $\beta^\perp\Gamma^0\Gamma^1 = \Gamma^0\Gamma^1\beta^\perp$).

However, the system can lower its energy as shown in figure 13.5. The F-string breaks, with its endpoints attached to the D-string. The endpoints can then move off to infinity, leaving only the D-string behind. This cannot be the whole story because the F-string carries the NS–NS 2-form charge, as measured by the integral of $*H$ over the 7-sphere in the figure: this flux must still be nonzero in the final configuration. This comes about because the F-string endpoints are charged under the D-string gauge field, so an electric flux runs between them. This flux remains in the end. Further, from the D-string action

$$S_1 = -T_p\int d^2\xi\, e^{-\Phi}\det{}^{1/2}(G_{ab} + B_{ab} + 2\pi\alpha' F_{ab}) \,, \tag{13.6.7}$$

one sees that $B_{\mu\nu}$ has a source proportional to the invariant electric flux $\mathscr{F}_{ab} = F_{ab} + B_{ab}/2\pi\alpha'$ on the D-string.

The simplest way to see that the resulting state is supersymmetric is

via T-duality along the 1-direction. The D1-brane becomes a D0-brane. The electric field is T-dual to a velocity, $\dot{A}_1 \to \dot{X}^1/2\pi\alpha'$, so the T-dual state is a D0-brane moving with constant velocity. This is invariant under the same number of supersymmetries as a D-brane at rest, namely the Lorentz boost of those supersymmetries. The boosted supersymmetries are linear combinations of the unbroken and broken supersymmetries of the D0-brane at rest. All of this carries over by T-duality to the D1–F1 system. We leave it as an exercise to verify that the tension takes the BPS value.

The F-string 'dissolves' in the D-string, leaving flux behind. For separated D- and F-strings there is an attractive force at long distance, a consequence of the lack of supersymmetry. One might have expected a more standard description of the bound state in terms of the F-string moving in this potential well. However, this description breaks down at short distance; happily, the D-brane effective theory gives a simple alternative description. Note that the bound state is quite deep: the binding tension

$$\tau_{(1,0)} + \tau_{(0,1)} - \tau_{(1,1)} = \frac{1 - O(g)}{2\pi\alpha'} \tag{13.6.8}$$

is almost the total tension of the F-string.

String theory with a constant open string field strength has a simple world-sheet description. The variation of the world-sheet action includes a surface term

$$\oint_{\partial M} ds\, \delta X^\mu \left(\frac{1}{2\pi\alpha'} \partial_n X_\mu + i F_{\mu\nu} \partial_t X^\nu \right) , \tag{13.6.9}$$

implying the linear boundary condition

$$\partial_n X_\mu + 2\pi\alpha' i F_{\mu\nu} \partial_t X^\nu = 0 . \tag{13.6.10}$$

This can also be seen from the T-dual relation to the moving D-brane.

All of the above extends immediately to p F-strings and one D-string forming a supersymmetric $(p,1)$ bound state. The general case of p F-strings and q D-strings is more complicated because the gauge dynamics on the D-strings is non-Abelian. A two-dimensional gauge coupling has units of inverse length-squared; we found the precise value $g_{D1}^2 = g/2\pi\alpha'$ in eq. (13.3.25). For dynamics on length scale l the effective dimensionless coupling is $gl^2/2\pi\alpha'$. No matter how weak the underlying string coupling g, the D-string dynamics at long distances is strongly coupled — this is a *relevant* coupling. The theory cannot then be solved directly, but it has been shown by indirect means that there is a bound string saturating the BPS bound for all (p,q) such that p and q are relatively prime. We will sketch the argument and leave the details to the references.

Focus for example on two D-strings and one F-string. There is a state with a separated $(1,1)$ bound state and $(0,1)$ D-string. The tension

$$\tau_{(1,1)} + \tau_{(0,1)} = \frac{(g^{-2}+1)^{1/2} + g^{-1}}{2\pi\alpha'} = \frac{2g^{-1} + g/2 + O(g^3)}{2\pi\alpha'} \tag{13.6.11}$$

exceeds the BPS bound

$$\tau_{(1,2)} = \frac{(4g^{-2}+1)^{1/2}}{2\pi\alpha'} = \frac{2g^{-1} + g/4 + O(g^3)}{2\pi\alpha'} . \tag{13.6.12}$$

The electric flux is on the first D-brane, so as a $U(2)$ matrix this is proportional to

$$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \frac{1}{2}\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} . \tag{13.6.13}$$

We have separated this into $U(1)$ and $SU(2)$ pieces. When we bring the two D-strings together, the $SU(2)$ field becomes strongly coupled as we have explained but the $U(1)$ part remains free. The $U(1)$ flux is then unaffected by the dynamics, and in particular there are no charged fields that might screen it. However, if the $SU(2)$ part is screened by the massless fields on the D-strings, then the total energy in the flux (which is proportional to the trace of the square of the matrix) is reduced by a factor of 2, from (13.6.11) to the BPS value (13.6.12).

That this does happen has been shown as follows. Focus on four of the 16 supersymmetries, forming the equivalent of $d = 4$, $N = 1$ supersymmetry. The six scalars $X^{4,\dots,9}$ can be written as three chiral superfields Φ_i, with the potential coming from a superpotential $\mathrm{Tr}(\Phi_1, [\Phi_2, \Phi_3])$. Now change the problem, adding to the superpotential a mass term,

$$W(\Phi) = \mathrm{Tr}(\Phi_1, [\Phi_2, \Phi_3]) + m\, \mathrm{Tr}(\Phi_i \Phi_i) . \tag{13.6.14}$$

This is an example of a general strategy for finding supersymmetric bound states: the D-string is a BPS state even under the reduced supersymmetry algebra. Its mass is then determined by the algebra and cannot depend on the parameter m. By now increasing m we can reduce the effective dimensionless coupling $g/2\pi\alpha' m^2$ to a value where the system becomes weakly coupled. It can then be shown that the $SU(2)$ system has a supersymmetric ground state.

The same argument goes through for all relatively prime p and q. When these are not relatively prime, $(p,q) = (kp, kq)$ and the system is only marginally unstable against falling apart into k subsystems. The dynamics is then quite different, and there is believed to be no bound string in this case. The bound string formed from p F-strings and q D-strings is called a (p,q)-string (as opposed to a p-p' string, which is an open string whose endpoints move on Dp- and Dp'-branes).

D0–Dp BPS bound

For a system with the charges of a D0-brane and a Dp-brane extended in the (1,..., p)-directions, the supersymmetry algebra becomes

$$\frac{1}{2}\left\{\begin{bmatrix} Q_\alpha \\ \tilde{Q}_\alpha \end{bmatrix}, \begin{bmatrix} Q_\beta^\dagger & \tilde{Q}_\beta^\dagger \end{bmatrix}\right\} = M\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\delta_{\alpha\beta} + \begin{bmatrix} 0 & Z \\ -Z^\dagger & 0 \end{bmatrix}\Gamma^0_{\alpha\beta}\,, \quad (13.6.15)$$

where

$$Z = \tau_0 + \tau_p V_p \beta\,, \quad \beta = \beta^1 \cdots \beta^p\,. \quad (13.6.16)$$

We have wrapped the Dp-brane on a torus of volume V_p so that its mass will be finite. The positivity of the left-hand side implies that

$$M^2\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \geq \begin{bmatrix} 0 & Z \\ -Z^\dagger & 0 \end{bmatrix}\Gamma^0\begin{bmatrix} 0 & Z \\ -Z^\dagger & 0 \end{bmatrix}\Gamma^0 = \begin{bmatrix} ZZ^\dagger & 0 \\ 0 & Z^\dagger Z \end{bmatrix}, \quad (13.6.17a)$$

$$ZZ^\dagger = \tau_0^2 + \tau_p V_p(\beta + \beta^\dagger) + \tau_p^2 V_p^2 \beta\beta^\dagger\,. \quad (13.6.17b)$$

For p a multiple of 4, β is Hermitean and $\beta^2 = 1$ by the same argument as in eq. (13.4.3). The BPS bound is then

$$M \geq \tau_0 + \tau_p V_p\,. \quad (13.6.18)$$

For $p = 4k+2$, β is anti-Hermitean, $\beta^2 = -1$, and the BPS bound is

$$M \geq (\tau_0^2 + \tau_p^2 V_p^2)^{1/2}\,. \quad (13.6.19)$$

These bounds are consistent with our earlier results on supersymmetry breaking, noting that $\#_{\mathrm{ND}} = p$. For $p = 4k$, a separated 0-brane and p-brane saturate the BPS bound (13.6.18), agreeing with the earlier conclusion that they leave some supersymmetry unbroken. For $p = 4k+2$ they do not saturate the bound and so cannot be in a BPS state, as found before. The reader can extend the analysis of the BPS bound to general values of p and p'.

D0–D0 bound states

The BPS bound for the quantum numbers of two 0-branes is $2\tau_0$, so any bound state will be at the lower edge of the continuous spectrum of two-body states. Nevertheless there is a well-defined, and as it turns out very important, question as to whether a *normalizable* state of energy $2\tau_0$ exists.

Let us first look at an easier problem. Compactify the 9-direction and add one unit of compact momentum, $p_9 = 1/R$. In a two-body state this momentum must be carried by one 0-brane or the other for minimum

total energy

$$\tau_0 + \left(\tau_0 + \frac{p_9^2}{2\tau_0}\right)\,. \quad (13.6.20)$$

For a bound state of mass $2\tau_0$, on the other hand, the minimum energy is

$$2\tau_0 + \frac{p_9^2}{4\tau_0}\,, \quad (13.6.21)$$

a finite distance below the continuum states. The reader may note some resemblance between these energies and the earlier (13.6.11) and (13.6.12). In fact the two systems are T-dual to one another. Taking the T-dual along the 9-direction, the D0-branes become D1-branes and the unit of momentum becomes a unit of fundamental string winding to give the (1,2) system, now at finite radius $R' = \alpha'/R$. Quantizing the (1,2) string wrapped on a circle gives the 2^8 states of an ultrashort BPS multiplet. In terms of the previous analysis, the $SU(2)$ part has a unique ground state in finite volume while the zero modes of the 16 components of the $U(1)$ gaugino generate 2^8 states. The earlier analysis is valid for the T-dual radius R' large, but having found an ultrashort multiplet we know that it must saturate the BPS bound exactly — its mass is determined by its charges and cannot depend on R. Similarly for n D-branes with m units of compact momentum, when m and n are relatively prime there is an ultrashort multiplet of bound states.

Now let us try to take $R \to \infty$ in order to return to the earlier problem. Having found that a bound state exists at any finite radius, it is natural to suppose that it persists in the limit. Since for any n we can choose a relatively prime m, it appears that there is one ultrashort bound state multiplet for any number of D0-branes. However, it is a logical possibility that the size of these states grows with R such that the states becomes nonnormalizable in the limit. To show that the bound states actually exist requires a difficult analysis, which has been carried out fully only for $n = 2$.

D0–D2 bound states

Here the BPS bound (13.6.19) puts any bound state discretely below the continuum. One can see hints of a bound state: the long-distance force is attractive, and for a coincident 0-brane and 2-brane the NS 0-2 string has a negative zero-point energy (13.4.8) and so a tachyon (which survives the GSO projection), indicating instability towards something. We cannot follow the tachyonic instability directly, but there is a simple alternative description of where it must end up.

Let us compactify the 1- and 2-directions and take the T-dual only

in the first, so that the 0-brane becomes a D-string wrapped in the 1-direction and the 2-brane becomes a D-string wrapped in the 2-direction. Now there is an obvious state with the same charges and lower energy, a single D-string running at an angle to wrap once in each direction. A single wrapped D-string is a BPS state (an ultrashort multiplet to be precise). Now use T-duality to return to the original description. As in figure 13.2, this will be a D2-brane with a nonzero magnetic field, such that

$$\int_{D2} F_2 = 2\pi \,. \tag{13.6.22}$$

We can also check that this state has the correct R–R charges. Expanding out the Chern–Simons action (13.3.18) gives

$$i\mu_3 \int (C_3 + 2\pi\alpha' F_2 \wedge C_1) \,. \tag{13.6.23}$$

Thus the magnetic field induces a D0-brane charge on the D2-branes, and the normalizations are consistent with $\mu_1 = 4\pi^2\alpha'\mu_3$.

The D0-brane dissolves in the D2-brane, turning into flux. The reader may note several parallels with the discussion of a D-string and an F-string, and wonder whether the systems are equivalent. In fact, they are not related to one other by T-duality or any other symmetry visible in string perturbation theory, but we will see in the next chapter that they are related by nonperturbative dualities.

The analysis extends directly to n D2-branes and m D0-branes: there is a single ultrashort multiplet of bound states.

D0–D4 bound states

As with the D0–D0 case, the BPS bound (13.6.18) implies that any bound state is marginally stable. We can proceed as before, first compactifying another dimension and adding a unit of momentum so that the bound state lies below the continuum. The low energy D0–D4 action is as discussed at the end of the previous section. Again it is an interacting theory, with a coupling that becomes large at low energy, but again the existence of supersymmetric bound states can be established by deforming the Hamiltonian; the details are left to the references. A difference from the D0–D0 case is that these bound states are invariant only under one-quarter of the original supersymmetries, the intersection of the supersymmetries of the 0-brane and of the 4-brane. The bound states then lie in a short (but not ultrashort) multiplet of 2^{12} states. It is useful to imagine that the D4-brane is wound on a finite but large torus. In this limit the massless 4-4 strings are essentially decoupled from the 0-4 and 0-0 strings. The 16 zero modes of the massless 4-4 fermion then generate 2^8 ground states

delocalized on the D4-brane. The fermion in the 0-4 hypermultiplet has eight real components (the smallest spinor in six dimensions) and their zero modes generate 2^4 ground states localized on the D0-brane. The tensor product gives the 2^{12} states.

For two D0-branes and one D4-brane, one gets the correct count as follows. We can have the two D0-branes bound to the D4-brane independently of one another; for a large D4-brane their interactions can be neglected. Each D0-brane has 2^4 states as noted above, eight bosonic and eight fermionic. Now count the number of ways two D0-branes can be put into these states: there are eight states with both D0-branes in the same (bosonic) state and $\frac{1}{2} \times 8 \times 7$ states with the D-branes in different bosonic states, for a total of $\frac{1}{2} \times 8 \times 9$ states. There are also $\frac{1}{2} \times 8 \times 7$ states with the D0-branes in different fermionic states and 8×8 with one in a bosonic state and one a fermionic state. Summing and tensoring with the 2^8 D4-brane states gives 2^{15} states. However, we could also imagine the two D0-branes first forming a D0–D0 bound state. The $SU(2)$ dynamics decouples and the resulting $U(1)$ dynamics is essentially the same as that of a single D0-brane. This bound state can then bind to the D4-brane, giving 2^{4+8} states as for a single D0-brane. The total number is 9×2^{12}.

This counting extends to n D0-branes and one D4-brane. The degeneracy D_n is given by the generating function

$$\sum_{n=0}^{\infty} q^n D_n = 2^8 \prod_{k=1}^{\infty} \left(\frac{1+q^k}{1-q^k} \right)^8 . \tag{13.6.24}$$

The term k in the product comes from bound states of k D0-branes which are then bound to the D4-brane. For each k there are eight bosonic states and eight fermionic states, and the expression (13.6.24) is then the product of the partition functions for all species. The coefficient of q^2 in its expansion is indeed 9×2^{12}. This proliferation of bound states is in contrast to the single ultrashort multiplet for n D0-branes and one D2-brane. The difference is that all the latter states are spread over the D2-brane, whereas the D0–D4 bound states are localized.

By T-duality the above system is converted into one D0-brane and n D4-branes, so the number of bound states of the latter is the same D_n. For m D0-branes and n D4-branes one gets the correct answer by the following argument. The equality of the degeneracy for one D0-brane and n D4-branes with that for n D0-branes and one D4-brane suggests that the systems are really the same — that in the former case we can somehow picture the D0-brane bound to n D4-branes as separating into n 'fractional branes,' each of which can then bind to each other in all combinations as in the earlier case. Then m D0-branes separate into mn fractional branes.

The degeneracy would then be D_{mn}, defined as in eq. (13.6.24). This is apparently correct, but the justification is not simple.

D-branes as instantons

The D0–D4 system is interesting in other ways. Consider its scalar potential

$$\frac{g_{D0}^2}{4}\sum_{A=1}^{3}(\chi_i^\dagger\sigma_{ij}^A\chi_j)^2+\sum_{i=1}^{5}\frac{(X_i-X_i')^2}{2\pi\alpha'}\chi^\dagger\chi\,, \tag{13.6.25}$$

as at the end of the previous section. The second term by itself has two branches of zeros,

$$X_i-X_i'=0\,,\quad \chi\neq 0 \tag{13.6.26}$$

and

$$X_i-X_i'\neq 0\,,\quad \chi=0\,. \tag{13.6.27}$$

The first of these, where the hypermultiplet scalars are nonzero, is known as a Higgs branch. The second, where the vector multiplet scalars are nonzero, is known as a Coulomb branch. In the present case the first term in the potential, the D-term, eliminates the Higgs branch. The condition

$$D^A\equiv\chi_i^\dagger\sigma_{ij}^A\chi_j=0 \tag{13.6.28}$$

implies that $\chi=0$. For example, if there were a nonzero solution we could by an $SU(2)$ rotation make only the upper component nonzero, and then D^3 is nonzero. However, for *two* D4-branes χ acquires a D4-brane index $a=1,2$ and the D-term condition is

$$\chi_{ia}^\dagger\sigma_{ij}^A\chi_{ja}=0\,. \tag{13.6.29}$$

This now is solved by

$$\chi_{ia}=v\delta_{ia} \tag{13.6.30}$$

for any v, or more generally

$$\chi_{ia}=vU_{ia} \tag{13.6.31}$$

for any constant v and unitary U. Further, U can be taken to lie in $SU(2)$ by absorbing its phase into v, and the latter can then be made real by a 4–4 $U(1)$ gauge rotation.

The Coulomb branch has an obvious physical interpretation, corresponding to the separation of the D0- and D4-brane in the directions transverse to the latter. But what of the Higgs branch?

Recall that non-Abelian gauge theories in four Euclidean dimensions have classical solutions, instantons, that are localized in all four dimensions. Their distinguishing property is that the field strength is self-dual or anti-self-dual,

$$*F_2=\pm F_2\,, \tag{13.6.32}$$

so that the Bianchi identity implies the field equations. Because the classical theory is scale-invariant, the characteristic size of the configuration is undetermined — there is a family of solutions parameterized by scale size ρ. The $U(n)$ gauge theory on coincident D4-branes is five-dimensional, so a configuration that looks like an instanton in the four spatial dimensions and is independent of time is a static classical solution, a soliton.

This soliton has many properties in common with the D0-brane bound to the D4-branes. First, it is a BPS state, breaking half of the supersymmetries of the D4-branes. The supersymmetry variation of the gaugino is

$$\delta\lambda\propto F_{MN}\Gamma^{MN}\zeta\,. \tag{13.6.33}$$

Here the nonzero terms involve the components of Γ^{MN} in the spatial directions of the D4-brane. These are then generators of the $SO(4)=SU(2)\times SU(2)$ rotation group. The self-duality relation (13.6.32) amounts to the statement that only the generators of the first or second $SU(2)$ appear in the variation. The ten-dimensional spinor ζ decomposes into

$$(\mathbf{4},\mathbf{2},\mathbf{1})+(\bar{\mathbf{4}},\mathbf{1},\mathbf{2}) \tag{13.6.34}$$

under $SO(5,1)\times SU(2)\times SU(2)$, so half the components are invariant under each $SU(2)$ and half the supersymmetry variations (13.6.33) are zero. Second, it carries the same R–R charge as the D0-brane. Expanding the Chern–Simons action (13.3.18) gives the term

$$\frac{1}{2}(2\pi\alpha')^2\mu_4\int C_1\wedge\mathrm{Tr}(F_2\wedge F_2)\,. \tag{13.6.35}$$

The topological charge of the instanton is

$$\int_{D4}\mathrm{Tr}(F_2\wedge F_2)=8\pi^2\,, \tag{13.6.36}$$

so the total coupling to a constant C_1 is $(4\pi^2\alpha')^2\mu_4=\mu_0$, exactly the charge of the D0-brane. Finally, the moduli (13.6.31) for the $SU(2)$ Higgs branch just match those of the $SU(2)$ instanton, v to the scale size ρ and U to the orientation of the instanton in the gauge group.[6] Let us check the counting of the moduli, as follows. There are eight real hypermultiplet scalars in χ. The three D-term conditions and the gauge rotation each remove one

[6] For a single instanton the latter are not regarded as moduli because they can be changed by a global gauge transformation, but with more than one instanton there are moduli for the relative orientation. The same is true of the D0-branes.

to leave four moduli. There are also four additional 0-0 moduli for the position of the particle within the D4-branes.

The precise connection between the D0-brane and the instanton is this. When the scale size ρ is large compared to the string scale, the low energy effective field theory on the D4-branes should give a good description of the instanton. However, as ρ is reduced below the string length, this description is no longer accurate. Happily, the D0-brane picture provides a description that is accurate in the opposite limit: the point $v = 0$ where the Higgs and Coulomb branches meet is the zero-size instanton, and turning on the Higgs moduli expands the instanton: as in the D0–D2 case, the D0-brane is dissolving into flux. This picture also accounts for the absence of a Higgs branch for a single D4-brane because there are no instantons for $U(1)$.

The gauge field of the small instanton can be measured directly. Recall that a slow D0-brane probe is sensitive to distances below the string scale. One can consider the D0–D4 bound system with an additional probe D0-brane. This has been studied in a slightly different form, taking first the T-dual to the D5–D9 system and using a D1-brane probe. As discussed earlier, only the effective field theory of the light open string states enters, though this is still rather involved because each open string endpoint can lie on a D1-, D5-, or D9-brane. However, after integrating out the massive fields (which get mass because they stretch from the probe to the other D-branes), the effective theory on moduli space displays the instanton gauge field. This provides a physical realization of the so-called Atiyah–Drinfeld–Hitchin–Manin (ADHM) construction of the general instanton solution.

Note the following curious phenomenon. Start with a large instanton, an object made out of the gauge fields that live on the D4-branes. Contract it to zero size, where the branches meet, and now pull it off the D4-branes along the Coulomb branch. The 'instanton' can no longer be interpreted as being made of the gauge fields, because these exist only on the D4-branes.

It should be noted that because the Higgs moduli are 0-4 fields their vertex operators are rather complicated: the different boundary conditions on the two endpoints mean that the world-sheet boundary conditions on the two sides of the vertex operator are different. They are similar to orbifold twisted state vertex operators — in fact, using the doubling trick, they are essentially half of the latter. It is therefore difficult to discuss in string theory a background with nonzero values for these fields, so the D0-brane picture is really an expansion in ρ, whereas the low energy field theory is an expansion in $1/\rho$.

Returning to the bound state problem, the system with m D0-branes bound to n D4-branes is equivalent to quantum mechanics on the moduli space of m $SU(n)$ instantons. The number of supersymmetric states is related to the topology of this space, and the answer has been argued to be D_{mn} as asserted before. (The connection with the fractional-brane picture is complicated and the latter is perhaps unphysical.)

D0–D6 bound states

The relevant bound is (13.6.19) and again any bound state would be below the continuum. This is as in the D0–D2 case, but the situation is different. The long-distance force is repulsive and the zero-point energy of coincident 0-6 NS strings is positive, so there is no sign of instability toward a supersymmetric state. One can give 0-brane charge to the 6-brane by turning on flux, but there is no configuration that has only these two charges and saturates the BPS bound. So it appears that there are *no* supersymmetric bound states.

D0–D8 bound states

This system is complicated in a number of ways and we will not pursue it. As one example of the complication, the R–R fields of the D8-brane do not fall off with distance (it has codimension 1, like a planar source in $3+1$ dimensions). The total energy is then infinite, and when the couplings to the dilaton and metric are taken into account the dilaton diverges a finite distance from the D8-brane. Thus the D8-brane cannot exist as an independent object, but only in connection with orientifold planes such as arise in the T-dual of the type I theory.

Exercises

13.1 (a) For the various massless fields of each of the type II string theories, write out the relation between the field at (x^μ, x^m) and at the orientifold image point $(x^\mu, -x^m)$. The analogous relation for the bosonic string was given as eq. (8.8.3).
(b) At the eight-dimensional orientifold plane (obtained from type I by T-duality on a single axis), which massless type IIA fields satisfy Dirichlet boundary conditions and which Neumann ones?

13.2 Find the scattering amplitude involving four bosonic open string states attached to a Dp-brane. [Hint: this should be very little work.]

13.3 (a) Consider three D4-branes that are extended along the (6,7,8,9)-, (4,5,8,9)-, and (4,5,6,7)-directions respectively. What are the unbroken supersymmetries?
(b) Add a D0-brane to the previous configuration. Now what are the unbroken supersymmetries?
(c) Call this configuration $(p_1, p_2, p_3, p_4) = (4, 4, 4, 0)$. By T-dualities, what other configurations of D-branes can be reached?

13.4 (a) Calculate the static potential between a D2-brane and a D0-brane from the cylinder amplitude by applying T-duality to the result (13.4.23).
(b) Do the same calculation in low energy field theory and compare the result with part (a) at distances long compared to the string scale.
(c) Extend parts (a) and (b) to a Dp-brane and D$(p+2)$-brane oriented such that $\#\mathrm{ND} = 2$.

13.5 Repeat parts (a) and (b) of the previous exercise for a D0-brane and D6-brane.

13.6 (a) Find the velocity-dependent interaction between a D4-brane and D0-brane due to the cylinder. You can do this by analytic continuation of the potential (13.4.23), with appropriate choice of angles.
(b) Expand the interaction in powers of v and find the explicit r-dependence at $O(v^2)$.
(c) Compare the interaction at distances long compared to the string scale with that obtained from the low energy field theory. One way to do this is to determine the long-range fields of the D4-brane by solving the linearized field equations with a D4-brane source, insert these into the D0-brane action, and expand in the velocity.

13.7 For the D4-brane and D0-brane, determine the interaction at distances short compared to the string scale as follows. Truncate the low energy action given at the end of section 13.5 to the massless 0-0 strings and the lightest 0-4 strings. The D0–D4 interaction arises as a loop correction to the effective action of the 0-0 collective coordinate, essentially a propagator correction for the field we called X_i'. Calculate this Feynman graph and compare with part (b) of the previous exercise at short distance. This is a bit easier than the corresponding D0–D0 calculation because the 0-4 strings do not include gauge fields. You need the Lagrangian for the 0-4 fermions; this is the dimensional reduction of the $(5+1)$-dimensional fermionic Lagrangian density $-i\bar{\psi}\Gamma^\mu D_\mu \psi$.

13.8 (a) Continuing the previous two exercises, obtain the full v-dependence at large r from the cylinder amplitude. Compare the result with the low energy supergravity (graviton–dilaton–R–R) exchange.
(b) Obtain the full v-dependence at small r and compare with the same from the open string loop.

13.9 Find a configuration of an infinite F-string and infinite D3-brane that leaves some supersymmetry unbroken.

13.10 From the D-string action, calculate the tension with q units of electric flux and compare with the BPS bound (13.6.3) for a $(q,1)$ string.

13.11 Carry out in detail the counting that leads to the bound state degeneracy (13.6.24).

13.12 Consider one of the points in figure 13.5(b) at which the F-string attaches to the D-string. At this point a $(1,0)$ and a $(0,1)$ string join to form a $(1,1)$ string; alternatively, if we count positive orientation as being inward, it is a junction of $(1,0)$, $(0,1)$, and $(-1,-1)$ strings. Consider the junction of three semi-infinite straight strings of general (p_i, q_i), with vanishing total p and q. Find the conditions on the angles such that the system is mechanically stable. Show that, with these angles, one-quarter of the original supersymmetries leave all three strings invariant.

14

Strings at strong coupling

Thus far we have understood string interactions only in terms of perturbation theory — small numbers of strings interacting weakly. We know from quantum field theory that there are many important phenomena, such as quark confinement, the Higgs mechanism, and dynamical symmetry breaking, that arise from having many degrees of freedom and/or strong interactions. These phenomena play an essential role in the physics of the Standard Model. If one did not understand them, one would conclude that the Standard Model incorrectly predicts that the weak and strong interactions are both long-ranged like electromagnetism; this is the famous criticism of Yang–Mills theory by Wolfgang Pauli.

Of course string theory contains quantum field theory, so all of these phenomena occur in string theory as well. In addition, it likely has new nonperturbative phenomena of its own, which must be understood before we can connect it with nature. Perhaps even more seriously, the perturbation series does not even define the theory. It is at best asymptotic, not convergent, and so gives the correct qualitative and quantitative behavior at sufficiently small coupling but becomes useless as the coupling grows.

In quantum field theory we have other tools. One can define the theory (at least in the absence of gravity) by means of a nonperturbative lattice cutoff on the path integral. There are a variety of numerical methods and analytic approximations available, as well as exactly solvable models in low dimensions. The situation in string theory was, until recently, much more limited.

In the past few years, new methods based on supersymmetry have revolutionized the understanding both of quantum field theory and of string theory. In the preceding chapters we have assembled the tools needed to study this. We now consider each of the five string theories and deduce the physics of its strongly coupled limit. We will see that all are limits of a single theory, which most surprisingly has a limit in

which spacetime becomes *eleven-dimensional*. We examine one proposal, matrix theory, for a formulation of this unified theory. We conclude with a discussion of related progress on one of the central problems of quantum gravity, the quantum mechanics of black holes.

We will use extensively the properties of D-brane states in mapping out the physics of strongly coupled strings. This allows a natural connection with our previous perturbative discussion. We should note, however, that most of these results were deduced by other methods before the role of D-branes was understood. Many properties of the R–R states were guessed (subject to many consistency checks) before the explicit D-brane picture was known.

14.1 Type IIB string and $SL(2,Z)$ duality

In the IIB theory, consider an infinite D-string stretched in the 1-direction. Let us determine its massless excitations, which come from the attached strings. The gauge field has no dynamics in two dimensions, so the only bosonic excitations are the transverse fluctuations. The Dirac equation for the massless R sector states

$$(\Gamma^0\partial_0 + \Gamma^1\partial_1)u = 0 \tag{14.1.1}$$

implies that $\Gamma^0\Gamma^1 u = \pm u$ for the left- and right-movers respectively, or that the boost eigenvalue $s_0 = \pm\frac{1}{2}$. The open string R sector ground state decomposes as

$$\mathbf{16} \to (\tfrac{1}{2},\mathbf{8}) + (-\tfrac{1}{2},\mathbf{8}') \tag{14.1.2}$$

under $SO(9,1) \to SO(1,1) \times SO(8)$, so the left-moving fermionic open strings on the D-string are in an $\mathbf{8}$ of $SO(8)$ and the right-movers are in an $\mathbf{8}'$.

Now consider an infinite fundamental string in the same theory. The massless bosonic fluctuations are again the transverse fluctuations. The massless fermionic fluctuations are superficially different, being the spacetime vectors ψ^μ and $\tilde{\psi}^\mu$. However, these are not entirely physical — the GSO projection forbids a single excitation of these fields. To identify the physical fermionic fluctuations, recall from the discussion in section 13.2 that these can be thought of as the Goldstone fermions of the supersymmetries broken by the string. The supersymmetry algebra for a state containing a long string was given in eq. (13.6.1), where $(p,q) = (1,0)$ for the fundamental string. The broken supersymmetries are those whose anticommutators do not vanish when acting on the BPS state; for the IIB F-string these are the Q_α with $\Gamma^0\Gamma^1 = +1$ and the $\tilde{Q}_\alpha$ with $\Gamma^0\Gamma^1 = -1$. The decomposition (14.1.2) then shows that the Goldstone fermions on

the IIB F-string have the same quantum numbers as on the IIB D-string; for the IIA F-string, on the other hand, the Goldstone fermions moving in both directions are **8**s. The relation between these excitations and the ψ^μ is just the refermionization used in section 12.6.

The D-string and F-string have the same massless excitations but they are not the same object. Their tensions are different,

$$\frac{\tau_{\rm F1}}{\tau_{\rm D1}} = g = e^{\Phi} . \tag{14.1.3}$$

This relation is a consequence of supersymmetry and so is exact. The field dependence of the central charge is connected by supersymmetry to the field dependence of the moduli space metric, and this receives no corrections for 16 or more supersymmetries. At weak coupling the F-string is much lighter than the D-string, but consider what happens as the coupling is adiabatically increased. Quantum mechanics does not allow the D-string states to simply disappear from the spectrum, and they must continue to saturate the BPS bound because their multiplet is smaller than the non-BPS multiplet. Thus at very strong coupling the D-string is still in the spectrum but it is much *lighter* than the F-string. It is tempting to conclude that the theory with coupling $1/g$ is the same as the theory with coupling g, but with the two strings reversing roles.

Let us amplify this as follows. Consider also a third scale, the gravitational length

$$l_0 = (4\pi^3)^{-1/8}\kappa^{1/4} , \tag{14.1.4}$$

where the important feature is the dependence on κ; the numerical constants are just included to simplify later equations. The relevant length scales are in the ratios

$$\tau_{\rm F1}^{-1/2} : l_0 : \tau_{\rm D1}^{-1/2} = g^{-1/4} : 1 : g^{1/4} . \tag{14.1.5}$$

At $g \ll 1$, if we start at long distance and consider the physics at progressively shorter scales, before reaching the scale where gravity would become strong we encounter the fundamental string scale and all the excited states of the fundamental string. At $g \gg 1$, we again encounter another scale before reaching the scale where gravity is strong, namely the D-string scale. We cannot be certain that the physics is the same as at weak coupling, but we do know that gravity is weak at this scale, and we can reproduce much of the same spectrum — the long straight string is a BPS state, as are states with arbitrary left- *or* right-moving excitations, so we can identify these. States with both left- and right-moving excitations are not BPS states, but at low energy the interactions are weak and we can identify them approximately.

Of course we have no nonperturbative definition of string theory and anything can happen. For example there could be very light non-BPS

states below the D-stri
no analogs in the we
identify many similarit
reversed, the simplest
them. Furthermore we
natural candidate for
can make on the basi
well with observations
and in some cases wit

One might have im
encounter a phase wit
time physics, but what
weak coupling. Of cou
there is no quantitativ
have the $g \approx 1$ theory
weak–strong dualities
theories for some tim
dimensional theories,
now very strong eviden
even in field theory, w
theory, weak–strong d
require new ideas, whi

The D-string has m
no analog in the hete
supersymmetric and d
As g becomes strong t
disappear into multi-p

As a further test, th
exactly from supersym
determined by the valu
the R–R scalar C_0 to
action is invariant und

The Einstein metric, d

and so is invariant.

$SL(2,\mathbf{Z})$ duality

The transformation (14.1.6) is one of the $SL(2,\mathbf{R})$ symmetries (12.1.32) of the low energy theory, with $\tau' = -1/\tau$. Consider the action of a general element on the 2-form coupling of the fundamental string,

$$\int_M B_2' = \int_M (B_2 d + C_2 c) \,. \tag{14.1.8}$$

For general real c and d there is no state with this coupling, but for the integer subgroup $SL(2,\mathbf{Z})$, the condition $ad - bc = 1$ implies that d and c are relatively prime. In this case we know from the previous chapter that there is a supersymmetric (d,c)-string with these quantum numbers. It is described at weak coupling as a bound state of c D-strings and d F-strings, and its existence at strong coupling follows from the continuation argument used above. This is a strong indication that this integer subgroup is an exact symmetry of the theory, with the weak–strong duality as one consequence. The BPS bound can be written in $SL(2,\mathbf{Z})$-invariant form as

$$\tau^2_{(p,q)} = l_0^4 (\mathcal{M}^{-1})^{ij} q_i q_j = l_0^4 \left[e^{\Phi} (p + C_0 q)^2 + e^{-\Phi} q^2 \right] . \tag{14.1.9}$$

Note that a subgroup of $SL(2,\mathbf{R})$, with $a = d = 1$ and $c = 0$, is visible in perturbation theory. This leaves the dilaton invariant and shifts

$$C_0 \to C_0 + b \,. \tag{14.1.10}$$

This shift is a symmetry of perturbation theory because the R–R scalar C_0 appears only through its field strength (gradient). The coupling to D-strings then breaks this down to integer shifts. This is evident from the bound (14.1.9), which is invariant under $C_0 \to C_0 + 1$ with $(p,q) \to (p-q,q)$. The integer shift takes τ to $\tau + 1$, and the full $SL(2,\mathbf{Z})$ is generated by this symmetry plus the weak–strong duality.

The IIB NS5-brane

Let us consider how the weak–strong duality acts on the various extended objects in the theory. We know that it takes the F- and D-strings into one another. It leaves the potential C_4 invariant and so should take the D3-brane into itself. The D5-brane is a magnetic source for the R–R 2-form charge: the integral of F_3 over a 3-sphere surrounding it is nonzero. This must be transformed into a magnetic source for the NS–NS 2-form charge. We have not encountered such an object before — it is neither a string nor a D-brane. Rather, it is a soliton, a localized classical solution to the field equations.

Consider the action for a graviton, dilaton, and q-form field strength in d dimensions, of the general form

$$\int d^{10}x\,(-G)^{1/2} e^{-2\Phi} (R + 4\partial_\mu \Phi \partial^\mu \Phi) - \frac{1}{2} \int e^{2\alpha\Phi} |F_q|^2 \,, \tag{14.1.11}$$

where α is -1 for an NS–NS field and 0 for an R–R field. We can look for a solution which is spherically symmetric in $q+1$ directions and independent of the other $8-q$ spatial dimensions and of time, and which has a fixed 'magnetic' charge

$$\int_{S_q} F_q = Q \,. \tag{14.1.12}$$

Here the q-sphere is centered on the origin in the $q+1$ spherically symmetric dimensions. This would be an $(8-q)$-brane. The field equation

$$d * (e^{2\alpha\Phi} F_q) = 0 \tag{14.1.13}$$

is automatic as a consequence of the spherical symmetry. The dual field strength is $F_{10-q} = *e^{2\alpha\Phi} F_q$, for which eq. (14.1.13) becomes the Bianchi identity. An 'electric' solution with

$$\int_{S_{10-q}} *e^{2\alpha\Phi} F_q = Q' \tag{14.1.14}$$

would be a $(q-2)$-brane.

A generalization of Birkhoff's theorem from general relativity guarantees a unique solution for given mass M and charge Q. For M/Q greater than a critical value $(M/Q)_c$ the solution is a black hole, with a singularity behind a horizon. More precisely, the solution is a *black p-brane*, meaning that it is extended in p spatial dimensions and has a black hole geometry in the other $9-p$. Essentially the source for the field strength is hidden in the singularity. For $M/Q < (M/Q)_c$, there is a naked singularity. The solution with $M/Q = (M/Q)_c$ is called extremal, and in most cases it is a supersymmetric solution, saturating the BPS bound. The naked singularities would then be excluded by the bound.

For the NS5-brane, the extremal solution is supersymmetric and takes the form

$$G_{mn} = e^{2\Phi} \delta_{mn} \,, \quad G_{\mu\nu} = \eta_{\mu\nu} \,, \tag{14.1.15a}$$

$$H_{mnp} = -\epsilon_{mnp}{}^q \partial_q \Phi \,, \tag{14.1.15b}$$

$$e^{2\Phi} = e^{2\Phi(\infty)} + \frac{Q}{2\pi^2 r^2} \,. \tag{14.1.15c}$$

Here the x^m are transverse to the 5-brane, the x^μ are tangent to it, and $r^2 = x^m x^m$. This is the magnetically charged object required by string duality. The product $\tau_{D1}\tau_{D5} = \pi/\kappa^2$ should equal $\tau_{F1}\tau_{NS5}$ by the Dirac

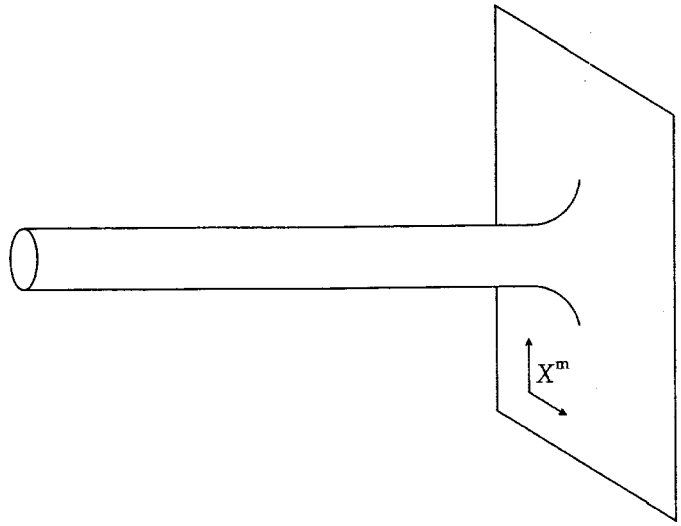


Fig. 14.1. Infinite throat of an NS5-brane, with asymptotically flat spacetime on the right. The x^μ-directions, in which the 5-brane is extended, are not shown.

quantization condition (which determines the product of the charges) combined with the BPS condition (which relates the charges to the tensions). This gives

$$\tau_{\rm NS5} = \frac{2\pi^2\alpha'}{\kappa^2} = \frac{1}{(2\pi)^5 g^2 \alpha'^3} \,. \tag{14.1.16}$$

There must also be bound states of this with the D5-brane, which are presumably described by adding R–R flux to the above solution.

The geometry of the metric (14.1.15), shown in figure 14.1, is interesting. There is an infinite throat. The point $x^m = 0$ is at infinite distance, and as one approaches it the radius of the angular 3-spheres does not shrink to zero but approaches an asymptotic value $(Q/2\pi^2)^{1/2}$. The dilaton grows in the throat of the 5-brane, diverging at infinite distance. String perturbation theory thus breaks down some distance down the throat, and the effective length is probably finite. Because of the strong coupling one cannot describe this object quite as explicitly as the fundamental strings and D-brane, but one can look at fluctuations of the fields around the classical solution. There are normalizable massless fluctuations corresponding to translations and also ones which transform as a vector on the 5-brane, and

the fermionic partners of these. These are the same as for the D5-brane, as should be true by duality.

It should be noted that the description above is in terms of the string metric, which is what appears in the string world-sheet action and is relevant for the dynamics of the string. The geometry is rather different in the Einstein metric $G_{{\rm E}\mu\nu} = e^{-\Phi/2} G_{\mu\nu}$. In the string metric the radial distance is $ds \propto x^{-1} dx$, while in the Einstein metric it is $ds \propto x^{-3/4} dx$. The latter is singular but integrable, so the singularity is at finite distance. Thus, different probes can see a very different geometry.

Let us make a few more comments on this solution. For an NS field strength, a shift of the dilaton just multiplies the classical action by a constant. The solution is then independent of the dilaton, and its size can depend only on α' and the charge Q. The charge is quantized, $Q = nQ_0$, by the Dirac condition. The radius is then of order $\alpha'^{1/2}$ times a function of n, which in fact is $n^{1/2}$. For small n the characteristic scale of the solution is the string scale, so the low energy theory used to find the solution is not really valid. However, there are nonrenormalization theorems, which have been argued to show that the solution does not receive corrections. There is also a description of the throat region that is exact at string tree level — it does not use sigma model perturbation theory but is an exact CFT. The geometry of the throat is $S_3 \times R_1 \times$ six-dimensional Minkowski space. The CFT similarly factorizes. The six dimensions parallel to the brane world-volume are the usual free fields. The CFT of the radial coordinate is the linear dilaton theory that we have met before, with the dilaton diverging at infinite distance. The CFT of the angular directions is an $SU(2) \times SU(2)$ current algebra at level n, in a form that we will discuss in the next chapter.

This construction might seem to leave us with an embarrassment of riches, for we can similarly construct NS–NS electrically charged solutions and R–R charged solutions, for which we already have the F-string and D-branes as sources. In fact, the NS–NS electrically charged solution has a pointlike singularity and the fields satisfy the field equations with a δ-function source at the singularity. Thus this solution just gives the external fields produced by the F-string. The R–R charged solutions are black p-branes. Their relation to the D-branes will be considered at the end of this chapter.

A fundamental string can end on a D5-brane. It follows by weak–strong duality that a D-string should be able to end on an NS5-brane. A plausible picture is that it extends down the infinite throat. Its energy is finite in spite of the infinite length because of the position-dependence of the dilaton. Similarly a D-string should be able to end on a D3-brane. There is a nontrivial aspect to the termination of one object 'A' on a second

'B', which we saw in figure 13.5. Since A carries a conserved charge, the coupling between spacetime forms and world-brane fields of B must be such as to allow it to carry the charge of A.

The solution for any number of parallel NS5-branes is simply given by substituting

$$e^{2\Phi} = e^{2\Phi(\infty)} + \frac{1}{2\pi^2}\sum_{i=1}^{N}\frac{Q_i}{(x-x_i)^2} \tag{14.1.17}$$

into the earlier solution (14.1.15). A D-string can run from one 5-brane to another, going down the throat of each. The ground state of this D-string is a BPS state. It is related by string duality to a ground state F-string stretched between two D5-branes, which is related by T-duality to a massless open string in the original type I theory. The mass of the D-string is given by the classical D-string action in the background (14.1.17), and agrees with string duality. In particular it vanishes as the NS5-branes become coincident, so like D-branes these have a non-Abelian symmetry in this limit. The limiting geometry is a single throat with twice the charge; in the limit, the non-Abelian degrees of freedom are in the strong coupling region down the throat and cannot be seen explicitly.

D3-branes and Montonen–Olive duality

Consider a system of n D3-branes. The dynamics on the D-branes is a $d = 4$, $N = 4$ $U(n)$ gauge theory, with the gauge coupling (13.3.25) equal to

$$g_{\mathrm{D3}}^2 = 2\pi g\ . \tag{14.1.18}$$

In particular this is dimensionless, as it should be for a gauge theory in four dimensions. At energies far below the Planck scale, the couplings of the closed strings to the D-brane excitations become weak and we can consider the D-brane gauge theory separately.

The $SL(2,\mathbf{Z})$ duality of the IIB string takes this system into itself, at a different coupling. In particular the weak–strong duality $g \to 1/g$ takes

$$g_{\mathrm{D3}}^2 \to \frac{4\pi^2}{g_{\mathrm{D3}}^2}\ . \tag{14.1.19}$$

This is a weak–strong duality transformation within the gauge theory itself. Thus, the self-duality of the IIB string implies a similar duality within $d = 4$, $N = 4$ gauge theory. Such a duality was conjectured by Montonen and Olive in 1979. The evidence for it is of the same type as for string duality: duality of BPS masses and degeneracies and of the low energy effective action. Nevertheless the reaction to this conjecture was for a long time skeptical, until the development of supersymmetric gauge theory in the past few years placed it in a broader and more systematic context.

To understand the full $SL(2,\mathbf{Z})$ symmetry we need also to include the coupling (13.3.18) to the R–R scalar,

$$\frac{1}{4\pi}\int C_0\,\mathrm{Tr}(F_2\wedge F_2)\ . \tag{14.1.20}$$

This is the Pontrjagin (instanton winding number) term, with $C_0 = \theta/2\pi$. The full gauge theory action, in a constant C_0 background, is

$$-\frac{1}{2g_{\mathrm{D3}}^2}\int d^4x\,\mathrm{Tr}(|F_2|^2) + \frac{\theta}{8\pi^2}\int \mathrm{Tr}(F_2\wedge F_2)\ . \tag{14.1.21}$$

The duality $C_0 \to C_0 + 1$ is then the shift $\theta \to \theta + 2\pi$, corresponding to quantization of instanton charge. This and the weak–strong duality generate the full $SL(2,\mathbf{Z})$.

Let the D3-branes be parallel but slightly separated, corresponding to spontaneous breaking of $U(n)$ to $U(1)^n$. The ground state of an F-string stretched between D3-branes is BPS, and corresponds to a vector multiplet that has gotten mass from spontaneous breaking. The weak–strong dual is a D1-string stretched between D3-branes. To be precise, this is what it looks like when the separation of the D3-branes is large compared to the string scale. When the separation is small there is an alternative picture of this state as an 't Hooft-Polyakov magnetic monopole in the gauge theory. The size of the monopole varies inversely with the energy scale of gauge symmetry breaking and so inversely with the separation. This is similar to the story of the instanton in section 13.6, which has a D-brane description when small and a gauge theory description when large.

The relation between the IIB and Montonen–Olive dualities is one example of the interplay between the spacetime dynamics of various branes and the nonperturbative dynamics of the gauge theories that live on them. This is a very rich subject, and one which at this time is developing rapidly.

14.2 U-duality

The effect of toroidal compactification is interesting. The symmetry group of the low energy supergravity theory grows with the number k of compactified dimensions, listed as G in table B.3. We are familiar with two subgroups of each of these groups. The first is the $SL(2,\mathbf{R})$ symmetry of the uncompactified IIB theory. The second is the perturbative $O(k,k,\mathbf{R})$ symmetry of compactification of strings on T^k, which we encountered in the discussion of Narain compactification in chapter 8. In each case the actual symmetry of the full theory is the integer subgroup, the $O(k,k,\mathbf{Z})$

T-duality group and the $SL(2,\mathbf{Z})$ of the ten-dimensional IIB theory. The continuous $O(k,k,\mathbf{R})$ is reduced to the discrete $O(k,k,\mathbf{Z})$ by the discrete spectrum of (p_L,p_R) charges, and the continuous $SL(2,\mathbf{R})$ to the discrete $SL(2,\mathbf{Z})$ by the discrete spectrum of (p,q)-strings. In the massless limit the charged states do not appear and the symmetry appears to be continuous.

The natural conjecture is that in each case the maximal integer subgroup of the low energy symmetry is actually a symmetry of the full theory. This subgroup has been given the name *U-duality*. In perturbation theory we only see symmetries that act linearly on g and so are symmetries of each term in the perturbation series — these are the T-dualities plus shifts of the R–R fields. The other symmetries take small g to large and so require some understanding of the exact theory. The principal tools here are the constraints of supersymmetry on the low energy theory, already used in writing table B.3, and the spectrum of BPS states, which can be determined at weak coupling and continued to strong.

Let us look at the example of the IIB string on T^5, which by T-duality is the same as the IIA string on T^5. This is chosen because it is the setting for the simplest black hole state counting, and also because the necessary group theory is somewhat familiar from grand unification.

Let us first count the gauge fields. From the NS–NS sector there are five Kaluza–Klein gauge bosons and five gauge bosons from the antisymmetric tensor. There are also 16 gauge bosons from the dimensional reduction of the various R–R forms: five from $C_{\mu n}$, ten from $C_{\mu npq}$ and one from $C_{\mu npqrs}$. The index μ is in the noncompact dimensions, and in each case one sums over all antisymmetric ways of assigning the compact dimensions to the roman indices. Finally, in five noncompact dimensions the 3-form $B_{\mu\nu\rho}$ is equivalent by Poincaré duality to a vector field, giving 27 gauge bosons in all.

Let us see how T-duality acts on these. This group is $O(5,5,\mathbf{Z})$, generated by T-dualities on the various axes, linear redefinitions of the axes, and discrete shifts of the antisymmetric tensor. This mixes the first ten NS–NS gauge fields among themselves, and the 16 R–R gauge fields among themselves, and leaves the final NS–NS field invariant. Now, a representation of $O(10,\mathbf{R})$ automatically gives a representation of $O(5,5,\mathbf{R})$ by analytic continuation, and so in turn a representation of the subgroup $O(5,5,\mathbf{Z})$. The group $O(10,\mathbf{R})$ has a vector representation **10**, spinor representations **16** and **16′**, and of course a singlet **1**. The gauge fields evidently transform in these representations; which spinor occurs depends on whether we start with the IIA or IIB theory, which differ by a parity transformation on $O(5,5,\mathbf{Z})$.

According to table B.3, the low energy supergravity theory for this compactification has a continuous symmetry $E_{6(6)}$, which is a noncompact version of E_6. The maximal discrete subgroup is denoted $E_{6(6)}(\mathbf{Z})$. The

group E_6 has a representation **27**, and a subgroup $SO(10)$ under which

$$\mathbf{27} \to \mathbf{10} + \mathbf{16} + \mathbf{1} \,. \tag{14.2.1}$$

This may be familiar to readers who have studied grand unification; some of the relevant group theory was summarized in section 11.4 and exercise 11.5. Evidently the gauge bosons transform as this **27**.

Now let us identify the states carrying the various charges. The charges **10** are carried by the Kaluza–Klein and winding strings. Then U-duality also requires states in the **16**. These are just the various wrapped D-branes. Finally, the state carrying the **1** charge is the NS5-brane, fully wrapped around the T^5 so that it is localized in the noncompact dimensions.

U-duality and bound states

It is interesting to see how some of the bound state results from the previous chapter fit the predictions of U-duality in detail. We will generate U transformations as a combination of $T_{mn\cdots p}$, which is a T-duality in the indicated directions, and S, the IIB weak–strong transformation. The former switches between Neumann and Dirichlet boundary conditions and between momentum and winding number in the indicated directions. The latter interchanges the NS–NS and R–R 2-forms but leaves the R–R 4-form invariant, and acts correspondingly on the solitons carrying these charges. We denote by $\mathrm{D}_{mn\cdots p}$ a D-brane extended in the indicated directions, and similarly for F_m a fundamental string extended in the given direction and p_m a momentum-carrying BPS state.

The first duality chain is

$$(\mathrm{D}_9,\mathrm{F}_9) \overset{T_{78}}{\to} (\mathrm{D}_{789},\mathrm{F}_9) \overset{S}{\to} (\mathrm{D}_{789},\mathrm{D}_9) \overset{T_9}{\to} (\mathrm{D}_{78},\mathrm{D}_\emptyset) \,. \tag{14.2.2}$$

Thus the D-string/F-string bound state is U-dual to the D0–D2 bound state. The constructions of these bound states were similar, but the precise relation goes through the nonperturbative step S. In each case there is one short multiplet of BPS states.

The second chain is

$$(\mathrm{D}_{6789},\mathrm{D}_\emptyset) \overset{T_6}{\to} (\mathrm{D}_{789},\mathrm{D}_6) \overset{S}{\to} (\mathrm{D}_{789},\mathrm{F}_6) \overset{T_{6789}}{\to} (\mathrm{D}_6,p_6) \overset{S}{\to} (\mathrm{F}_6,p_6) \,. \tag{14.2.3}$$

The bound states of n D0-branes and m D4-branes are thus U-dual to fundamental string states with momentum n and winding number m in one direction. Let us compare the degeneracy of BPS states in the two cases. For the winding string, the same argument as led to eq. (11.6.28) for the heterotic string shows that the BPS strings satisfy

$$(N,\tilde{N}) = \begin{cases} (nm,0) \,, & nm > 0 \,, \\ (0,-nm) \,, & nm < 0 \,. \end{cases} \tag{14.2.4}$$

Here N and $\tilde{N}$ are the number of excitations above the massless ground state. We see that BPS states have only left-moving or only right-moving excitations. The generating function for the number of BPS states is the usual string partition function,

$$\mathrm{Tr}\, q^N = 2^8 \prod_{k=1}^{\infty} \left(\frac{1+q^k}{1-q^k}\right)^8 , \tag{14.2.5}$$

or the same with $\tilde{N}$. Note that we are counting the states of one string with winding number m, not of a bound state of m strings of winding number 1. The latter does not exist at small g — except insofar as one can think of the multiply wound string in this way. The counting (14.2.5) is most easily done with the refermionized θ_α. In terms of the ψ^μ the GSO projection gives several terms, which simplify using the abstruse identity. The string degeneracy (14.2.5) precisely matches the degeneracy D_{nm} of D0–D4 bound states in section 13.6.

14.3 $SO(32)$ type I–heterotic duality

In the type I theory, the only R–R fields surviving the Ω projection are the 2-form, which couples electrically to the D1-brane and magnetically to the D5-brane, and the nondynamical 10-form which couples to the D9-brane. This is consistent with the requirement for unbroken supersymmetry — the D1- and D5-branes both have $\#_{\mathrm{ND}} = 4k$ relative to the D9-brane.[1]

Consider again an infinite D-string stretched in the 1-direction. The type I D-string differs from that of the IIB theory in two ways. The first is the projection onto oriented states. The $U(1)$ gauge field, with vertex operator $\partial_t X^\mu$, is removed. The collective coordinates, with vertex operators $\partial_n X^\mu$, remain in the spectrum because the normal derivative is even under reversal of the orientation of the boundary. That is, in terms of its action on the X oscillators Ω has an additional -1 for the $m = 2, \ldots, 9$ directions, as compared to the action on the usual 9-9 strings. By superconformal symmetry this must extend to the ψ^μ, so that in particular on the ground states Ω is no longer -1 but acts as

$$-\beta = -\exp[\pi i(s_1 + s_2 + s_3 + s_4)] , \tag{14.3.1}$$

with an additional rotation by π in the four planes transverse to the string. From the fermionic 1-1 strings of the IIB D-string, this removes the left-moving **8** and leaves the right-moving **8′**.

[1] It is conceivable that the D3- and D7-branes exist as non-BPS states. However, they would be expected to decay rapidly; also, there is some difficulty at the world-sheet level in defining them, as explained later.

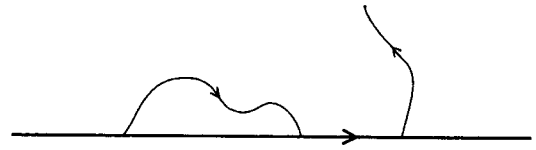


Fig. 14.2. D-string in type I theory with attached 1-1 and 1-9 strings.

The second modification is the inclusion of 1-9 strings, strings with one end on the D1-brane and one on a D9-brane. The end on the D9-brane carries the type I Chan–Paton index, so these are vectors of $SO(32)$. These strings have $\#_{\mathrm{ND}} = 8$ so that the NS zero-point energy (13.4.8) is positive, and there are no massless states in the NS sector. The R ground states are massless as always. Only ψ^0 and ψ^1 are periodic in the R sector, so their zero modes generate two states

$$|s_0; i\rangle , \tag{14.3.2}$$

where $s_0 = \pm\frac{1}{2}$ and i is the Chan–Paton index for the 9-brane end. One of these two states is removed by the GSO projection; our convention has been

$$\exp(\pi i F) = -i \exp[2\pi i(s_0 + \ldots + s_4)] , \tag{14.3.3}$$

so that the state with $s_0 = +\frac{1}{2}$ would survive. We now impose the G_0 condition, which as usual (e.g. eq. (14.1.1)) reduces to a Dirac equation and then to the condition $s_0 = +\frac{1}{2}$ for the left-movers and $s_0 = -\frac{1}{2}$ for the right-movers. The right-moving 1-9 strings are thus removed from the spectrum by the combination of the GSO projection and G_0 condition. Finally we must impose the Ω projection; this determines the 9-1 state in terms of the 1-9 state, but otherwise makes no constraint.

To summarize, the massless bosonic excitations are the usual collective coordinates. The massless fermionic excitations are right-movers in the **8′** of the transverse $SO(8)$ and left-movers that are invariant under $SO(8)$ and are vectors under the $SO(32)$ gauge group. This is the same as the excitation spectrum of a long $SO(32)$ heterotic string. Incidentally, this explains how it can be consistent with supersymmetry that the 1-9 strings have massless R states and no massless NS states: the supersymmetry acts only on the right-movers. This is also a check that our conventions above were consistent — supersymmetry requires the 1-9 fermions to move in the opposite direction to the 1-1 fermions. From a world-sheet point of view, this is necessary in order that the gravitino OPE be consistent.

The D-string tension $\tau_{D1} = 1/2\pi\alpha' g$ is again exact, and at strong coupling this is the lowest energy scale in the theory, below the gravitational scale and the fundamental string tension. By the same arguments as in the IIB case, the simplest conclusion is that the strongly coupled type I theory is actually a weakly coupled $SO(32)$ heterotic string theory. As a check, this must be consistent with the low energy supergravity theories. We have already noted that these must be the same up to field redefinition, because the supersymmetry algebras are the same. It is important, though, that the redefinition (12.1.41),

$$G_{I\mu\nu} = e^{-\Phi_h} G_{h\mu\nu} , \quad \Phi_I = -\Phi_h , \tag{14.3.4a}$$

$$\tilde{F}_{I3} = \tilde{H}_{h3} , \quad A_{I1} = A_{h1} , \tag{14.3.4b}$$

includes a reversal of the sign of the dilaton.

The conclusion is that there is a single theory, which looks like a weakly coupled type I theory when $e^{\Phi_I} \ll 1$ and like a weakly coupled $SO(32)$ heterotic theory when $e^{\Phi_I} \gg 1$. The type I supergravity theory is a good description of the *low energy* physics throughout. Even if the dimensionless string coupling is of order 1, the couplings in the low energy theory are all irrelevant in ten dimensions (and remain irrelevant as long as there are at least five noncompact dimensions) and so are weak at low energy.

As a bonus we have determined the strong-coupling physics of the $SO(32)$ heterotic string, namely the type I string. It would have been harder to do this directly. The strategy we have used so far, which would require finding the type I string as an excitation of the heterotic theory, would not work because a long type I string is not a BPS state. The NS–NS 2-form, whose charge is carried by most fundamental strings, is not present in the type I theory. The R–R 2-form remains, but its charge is carried by the type I D-string, not the F-string. That the long type I F-string is not a BPS state is also evident from the fact that it can break and decay. As the type I coupling increases, this becomes rapid and the type I string disappears as a recognizable excitation.

The strings of the type I theory carry only symmetric and antisymmetric tensor representations of the gauge group, while the strings of the heterotic theory can appear in many representations. We see that the corresponding states appear in the type I theory as D-strings, where one gets large representations of the gauge group by exciting many 1-9 strings. Note in particular that type I D-strings can carry the spinor representation of $SO(32)$; this representation is carried by fundamental heterotic strings but cannot be obtained in the product of tensor representations. Consider a long D-string wrapped around a periodic dimension of length L. The massless 1-9 strings are associated with fermionic fields Λ^i living on the D-string, with i the $SO(32)$ vector index. The zero modes of these,

$$\Lambda_0^i = L^{-1/2} \int_0^L dx^1 \, \Lambda^i(x^1) , \tag{14.3.5}$$

satisfy a Clifford algebra

$$\{\Lambda_0^i, \Lambda_0^j\} = \delta^{ij} . \tag{14.3.6}$$

The quantization now proceeds just as for the *fundamental* heterotic string, giving spinors $\mathbf{2^{15}} + \mathbf{2^{15'}}$ of $SO(32)$. Again, the Λ^i are fields that create light *strings*, but they play the same role here as the λ^i that create excitations *on* the heterotic string.

The heterotic string automatically comes out in fermionic form, and so a GSO projection is needed. We can think of this as gauging a discrete symmetry that acts as -1 on every D-string endpoint (the idea of gauging a discrete group was explained in section 8.5). This adds in the NS sectors for the fields Λ^i and removes one of the two spinor representations. Recall that in the IIB D-string there is a continuous $U(1)$ gauge symmetry acting on the F-string endpoints. The part of this that commutes with the Ω projection and so remains on the type I D-string is just the discrete gauge symmetry that we need to give the current algebra GSO projection.

Quantitative tests

Consider the tension of the heterotic D-string,

$$\tau_{D1} \text{ (type I)} = \frac{\pi^{1/2}}{2^{1/2}\kappa}(4\pi^2\alpha') = \frac{g_{YM}^2}{8\pi\kappa^2} . \tag{14.3.7}$$

We have used the type I relation (13.3.31) to express the result in terms of the low energy gauge and gravitational couplings, which are directly measurable in scattering experiments. It should be noted that the type I cylinder amplitude for the D-brane interaction has an extra $\frac{1}{2}$ from the orientation projection as compared to the type II amplitude, so the D-brane tension is multiplied by $2^{-1/2}$. The result (14.3.7), obtained at weak type I coupling, is exact as a consequence of the BPS property. Hence it should continue to hold at strong type I coupling, and therefore agree with the relation between the *heterotic* string tension and the low energy couplings at weak heterotic coupling. Indeed, this is precisely eq. (12.3.37).

As another example, consider the $F_{\mu\nu}^4$ interaction (12.4.28) found in type I theory from the disk amplitude,

$$\frac{\pi^2\alpha'^2}{2\times 4!\, g_{YM}^2}(tF^4) = \frac{g_{YM}^2}{2^{10}\pi^5 4!\,\kappa^2}(tF^4) , \tag{14.3.8}$$

and the same interaction found in the $SO(32)$ heterotic theory from the torus,

$$\frac{1}{2^8\pi^5 4!\,\alpha'}(tF^4) = \frac{g_{\rm YM}^2}{2^{10}\pi^5 4!\,\kappa^2}(tF^4)\,. \tag{14.3.9}$$

Here (tF^4) is an abbreviation for the common Lorentz and gauge structure of the two amplitudes. In each theory we have expressed α' appropriately in terms of the low energy couplings. The agreement between the numerical coefficients of the respective interactions is not an accident but is required by type I–heterotic duality. To explain this, first we must assert without proof the fact that supersymmetry completely determines the dilaton dependence of the $F_{\mu\nu}^4$ interaction in a theory with 16 supersymmetries.[2] Hence we can calculate the coefficient when $\Phi_{\rm I}$ is large and negative and the type I calculation is valid, and it must agree with the result at large positive $\Phi_{\rm I}$ where the heterotic calculation is valid.

Actually, this particular agreement is not an independent test of duality, but is a consequence of the consistency of each string theory separately. The (tF^4) interaction is related by supersymmetry to the $B_2F_2^4$ interaction, and the coefficient of the latter is fixed in terms of the low energy spectrum by anomaly cancellation. However, this example illustrates the fact that weak–strong dualities in general can relate calculable amplitudes in the dual theories, and not only incalculable strong-coupling effects. In more complicated examples, such as compactified theories, there are many such successful relations that are not preordained by anomaly cancellation. As in this example, a tree-level amplitude on one side can be related to a loop amplitude on the other, or to an instanton calculation.

Type I D5-branes

The type I D5-brane has some interesting features. The D5–D9 system is related by T-duality to the D0–D4 system. We argued that in the latter case the D0-brane was in fact the zero-size limit of an instanton constructed from the D4-brane gauge fields. The same is true here. The type I theory has gauge field solutions in which the fields are independent of five spatial dimensions and are a localized Yang–Mills instanton configuration in the other four: this is a 5-brane. It has collective coordinates for its shape, and also for the size and gauge orientation of the instanton. In the zero-size limit, the D5-brane description is accurate. As in the discussion of D0–D4 bound states, there are flat directions for the 5-9 fields. Again these have the interpretation of blowing the D5-brane up into a 9-9 gauge field configuration whose cross-section is the $SO(32)$ instanton and which is independent of the other six dimensions.

The heterotic dual of the type I D5-brane is simple to deduce. The blown-up instanton is an ordinary field configuration. The transformation (14.3.4) between the type I and heterotic fields leaves the gauge field invariant, so this just becomes an instanton in the heterotic theory. The transformation of the metric has an interesting effect. What looks in the type I theory like a small instanton becomes in the heterotic theory an instanton at the end of a long but finite throat; in the zero-size limit the throat becomes infinite as in figure 14.1.

There is one difference from the earlier discussion of D-branes. It turns out to be necessary to assume that the type I D5-brane carries an $SU(2)$ symmetry — that is, a two-valued Chan–Paton index. More specifically, it is necessary on the D5-branes to take a symplectic rather than orthogonal projection. We will first work out the consequences of this projection, and then discuss why it must be so.

The bosonic excitation spectrum consists of

$$\psi_{-1/2}^{\mu}|0,k;ij\rangle\lambda_{ij}\,, \quad \psi_{-1/2}^{m}|0,k;ij\rangle\lambda'_{ij}\,, \tag{14.3.10}$$

which are the D5-brane gauge field and collective coordinate respectively; i and j are assumed to be two-valued. The symplectic Ω projection gives

$$M\lambda M^{-1} = -\lambda^T\,, \quad M\lambda' M^{-1} = \lambda'^T\,, \tag{14.3.11}$$

with M the antisymmetric 2×2 matrix. The general solutions are

$$\lambda = \sigma^a\,, \quad \lambda' = I\,. \tag{14.3.12}$$

In particular the Chan–Paton wavefunction for the collective coordinate is the identity, so 'both' D5-branes move together. We should really then refer to *one* D5-brane, with a two-valued Chan–Paton index. This is similar to the T-dual of the type I string, where there are 32 Chan–Paton indices but 16 D-branes, each D-brane index being doubled to account for the orientifold image. The world-brane vectors have Chan–Paton wavefunctions σ_{ij}^a so the gauge group is $Sp(1) = SU(2)$, unlike the IIB D5-brane whose gauge group is $U(1)$. For k coincident D5-branes the group is $Sp(k)$.

The need for a two-valued Chan–Paton index can be seen in four independent ways. The first is that it is needed in order to get the correct instanton moduli space, the instanton gauge group now being $SO(n)$ rather than $SU(n)$. We will not work out the details of this, but in fact this is how the $SU(2)$ symmetry was first deduced. Note that

[2] Notice that there are dilaton dependences hidden in the couplings in (14.3.8) and (14.3.9), which moreover are superficially different because of the different dilaton dependences of $g_{\rm YM}$ in the two string theories. However, the dilaton dependences are related by the field redefinition (14.3.4), and are correlated with the fact that the lower order terms in the action also have different dilaton dependences.

starting from a large instanton, it is rather surprising that in the zero-size limit a new internal gauge symmetry appears. The appearance of new gauge symmetries at special points in moduli space is now known to occur in many contexts. The non-Abelian gauge symmetry of coincident IIB NS5-branes, pointed out in section 14.1, was another such surprise. The enhanced gauge symmetry of the toroidally compactified string is a perturbative example.

The second argument for a symplectic projection is based on the fact that in the type I theory the force between D1-branes, and between D5-branes, is half of what was calculated in section 13.3 due to the orientation projection. The tension and charge are then each reduced by a factor $2^{-1/2}$, so the product of the charges of a single D1-brane and single (one-valued) D5-brane would then be only half a Dirac unit. However, since the D5-branes with a symplectic projection always move in pairs, the quantization condition is respected. The third argument is based on the spectrum of 5-9 strings. For each value of the Chan–Paton indices there are two bosonic states, as in eq. (13.5.19). The D5–D9 system has eight supersymmetries, and these two bosons form half of a hypermultiplet (section B.7). In an oriented theory the 9-5 strings are the other half, but in this unoriented theory these are not independent. A half-hypermultiplet is possible only for pseudoreal representations, like the **2** of $SU(2)$ — hence the need for the $SU(2)$ on the D5-brane.

The final argument is perhaps the most systematic, but also the most technical. Return to the discussion of the orientation projection in section 6.5. The general projection was of the form

$$\hat{\Omega}|\psi; ij\rangle = \gamma_{jj'}|\Omega\psi; j'i'\rangle\gamma^{-1}_{i'i} \ . \tag{14.3.13}$$

We can carry over this formalism to the present case, where now the Chan–Paton index in general runs over 1-, 5-, and 9-branes. In order for this to be a symmetry the matrix $\gamma_{jj'}$ must connect D-branes that are of the same dimension and coincident.[3] In chapter 6 we argued that $\hat{\Omega}^2 = 1$ and therefore that γ was either purely symmetric or purely antisymmetric. The first argument still holds, but the second rested on an assumption that is not true in general: that the operator Ω, the part of $\hat{\Omega}$ that acts on the fields, squares to the identity, $\Omega^2 = 1$. More generally, it may in fact be a phase.

Working out the phase of Ω is a bit technical. It is determined by the requirement that the symmetry be conserved by the operator product of the corresponding vertex operators. In the 5-5 sector, the massless

[3] This formalism also applies to the more general orientifold projection, where $\hat{\Omega}$ is combined with a spacetime symmetry. The matrix γ then connects each D-brane with its image under the spacetime symmetry.

vertex operator is $\partial_t X^\mu$ ($\Omega = -1$) for μ parallel to the 5-brane, and $\partial_n X^\mu$ ($\Omega = +1$) for μ perpendicular. On these states, $\Omega^2 = 1$, and the same is true for the rest of the 9-9 and 5-5 Hilbert spaces. To see this, use the fact that Ω multiplies any mode operator ψ_r by $\pm\exp(i\pi r)$. The mode expansions were given in section 13.4. In the NS sector this is $\pm i$, but the GSO projection requires that these mode operators act in pairs (the OPE is single-valued only for GSO-projected vertex operators). So $\Omega = \pm 1$, and this holds in the R sector as well by supersymmetry.

Now consider the NS 5-9 sector. The four X^μ with mixed Neumann–Dirichlet boundary conditions, say $\mu = 6, 7, 8, 9$, have a half-integer-mode expansion. Their superconformal partners ψ^μ then have an integer-mode expansion and the ground state is a representation of the corresponding zero-mode algebra. The vertex operator is thus a spin field: the periodic ψ^μ contribute a factor

$$V = e^{i(H_3+H_4)/2} \ , \tag{14.3.14}$$

where $H_{3,4}$ are from the bosonization of the four periodic $\psi^{6,7,8,9}$. We need only consider this part of the vertex operator, as the rest is the same as in the 9-9 string and so has $\Omega^2 = +1$. Now, the operator product of V with itself (which is in the 5-5 or 9-9 sector) involves $e^{i(H_3+H_4)}$, which is the bosonization of $(\psi^6 + i\psi^7)(\psi^8 + i\psi^9)$. This in turn is the vertex operator for the state

$$(\psi^6 + i\psi^7)_{-1/2}(\psi^8 + i\psi^9)_{-1/2}|0\rangle \ . \tag{14.3.15}$$

Finally we can deduce the Ω eigenvalue. For $|0\rangle$ it is $+1$, because its vertex operator is the identity, while each $\psi_{-1/2}$ contributes either $-i$ (for a 9-9 string) or $+i$ (for a 5-5 string), giving an overall -1. That is, the Ω eigenvalue of VV is -1, and so therefore is the Ω^2 eigenvalue of V.

In the 5-9 sector $\Omega^2 = -1$. Separate γ into a block γ_9 that acts on the D9-branes and a block γ_5 that acts on the D5-branes. Then repeating the argument in section 6.5 gives

$$\gamma_9^T\gamma_9^{-1} = \Omega^2_{5\text{-}9}\gamma_5^T\gamma_5^{-1} \ . \tag{14.3.16}$$

We still have $\gamma_9^T = +\gamma_9$ from tadpole cancellation, so we need $\gamma_5^T = -\gamma_5$, giving symplectic groups on the D5-brane. The minimum dimension for the symplectic projection is 2, so we need a two-valued Chan–Paton state. This argument seems roundabout, but it is faithful to the logic that the actions of Ω in the 5-5 and 9-9 sectors are related because they are both contained in the 5-9 × 9-5 product. Further, there does not appear to be any arbitrariness in the result. It also seems to be impossible to define the D3- or D7-brane consistently, as $\Omega^2 = \pm i$.

14.4 Type IIA string and M-theory

The type IIA string does not have D-strings but does have D0-branes, so let us consider the behavior of these at strong coupling. We focus on the D-brane of smallest dimension for the following reason. The D-brane tension $\tau_p = O(g^{-1}\alpha'^{-(p+1)/2})$ translates into a mass scale

$$(\tau_p)^{1/(p+1)} \approx g^{-1/(p+1)}\alpha'^{-1/2} \tag{14.4.1}$$

so that at strong coupling the smallest p gives the lowest scale. Thus we need to find an effective field theory describing these degrees of freedom.

The D0-brane mass is

$$\tau_0 = \frac{1}{g\alpha'^{1/2}} . \tag{14.4.2}$$

This is heavy at weak coupling but becomes light at strong coupling. We also expect that for any number n of D0-branes there is an ultrashort multiplet of bound states with mass

$$n\tau_0 = \frac{n}{g\alpha'^{1/2}} . \tag{14.4.3}$$

This is exact, so as the coupling becomes large all these states become light and the spectrum approaches a continuum. Such a continuous spectrum of particle states is characteristic of a system that is becoming noncompact. In particular, the evenly spaced spectrum (14.4.3) matches the spectrum of momentum (Kaluza–Klein) states for a periodic dimension of radius

$$R_{10} = g\alpha'^{1/2} . \tag{14.4.4}$$

Thus, as $g \to \infty$ an *eleventh* spacetime dimension appears. This is one of the greatest surprises in this subject, because perturbative superstring theory is so firmly rooted in ten dimensions.

From the point of view of supergravity all this is quite natural. Eleven-dimensional supergravity is the supersymmetric field theory with the largest possible Poincaré invariance. Beyond this, spinors have at least 64 components, and this would lead to massless fields with spins greater than 2; such fields do not have consistent interactions. We have used dimensional reduction of eleven-dimensional supergravity as a crutch to write down ten-dimensional supergravity, but now we see that it was more than a crutch: dimensional reduction keeps only the $p_{10} = 0$ states, but string theory has also states of $p_{10} \neq 0$ in the form of D0-branes and their bound states. Recall that in the reduction of the eleven-dimensional theory to IIA string theory, the Kaluza–Klein gauge boson which couples to p_{10} became the R–R gauge boson which couples to D0-branes. The eleventh dimension is invisible in string perturbation theory because this is an expansion around the zero-radius limit for the extra dimension, as is evident from eq. (14.4.4).

The eleven-dimensional gravitational coupling is given by dimensional reduction as

$$\kappa_{11}^2 = 2\pi R_{10}\kappa^2 = \frac{1}{2}(2\pi)^8 g^3 \alpha'^{9/2} . \tag{14.4.5}$$

The numerical factors here are inconvenient so we will define instead an eleven-dimensional Planck mass

$$M_{11} = g^{-1/3}\alpha'^{-1/2} , \tag{14.4.6}$$

in terms of which $2\kappa_{11}^2 = (2\pi)^8 M_{11}^{-9}$. The two parameters of the IIA theory, g and α', are related to the eleven-dimensional Planck mass and the radius of compactification by eqs. (14.4.4) and (14.4.6). Inverting these,

$$g = (M_{11}R_{10})^{3/2} , \quad \alpha' = M_{11}^{-3}R_{10}^{-1} . \tag{14.4.7}$$

The reader should be be alert to possible differences in convention in the definition of M_{11}, by powers of 2π; the choice here makes the conversion between string and M-theory parameters simple.

We know little about the eleven-dimensional theory. Its low energy physics must be described by $d = 11$ supergravity, but it has no dimensionless parameter in which to make a perturbation expansion. At energies of order M_{11} neither supergravity nor string theory is a useful description. It is hard to name a theory when one does not know what it is; it has been given the tentative and deliberately ambiguous name *M-theory*. Later in the chapter we will discuss a promising idea as to the nature of this theory.

U-duality and F-theory

Since we earlier deduced the strongly coupled behavior of the IIB string, and this is T-dual to the IIA string, we can also understand the strongly coupled IIA string in this way. Periodically identify the 9-direction. The IIB weak–strong duality S interchanges a D-string wound in the 9-direction with an F-string wound in the 9-direction. Under T-duality, the D-string becomes a D0-brane and the wound F-string becomes a string with nonzero p_9. So TST takes D0-brane charge into p_9 and vice versa. Thus we should be able to interpret D0-brane charge as momentum in a dual theory, as indeed we argued above. The existence of states with R–R charge and of the eleventh dimension was inferred in this way — as were the various other dualities — before the role of D-branes was understood.

It is notable that while the IIA and IIB strings are quite similar in perturbation theory, their strongly coupled behaviors are very different. The strongly coupled dual of the IIB theory is itself, while that of the

IIA theory is a new theory with an additional spacetime dimension. Nevertheless, we see that these results are consistent with the equivalence of the IIA and IIB theories under T-duality. The full set of dualities forms a rich interlocking web.

For the type II theory on a circle, the noncompact symmetry of the low energy theory is $SL(2,\mathbf{R}) \times SO(1,1,\mathbf{R})$ (table B.3) and the discrete U-duality subgroup is

$$d = 9: \quad U = SL(2,\mathbf{Z}) . \tag{14.4.8}$$

Regarded as a compactification of the IIB string, this is just the $SL(2,\mathbf{Z})$ symmetry of the ten-dimensional theory. Regarded as a compactification of the IIA string on a circle and therefore of M-theory on T^2, it is a geometric symmetry, the modular transformations of the spacetime T^2.

For the type II theory on T^2, the noncompact symmetry of the low energy theory is $SL(3,\mathbf{R}) \times SL(2,\mathbf{R})$ and the discrete U-duality subgroup is

$$d = 8: \quad U = SL(3,\mathbf{Z}) \times SL(2,\mathbf{Z}) . \tag{14.4.9}$$

In section 8.4 we studied compactification of strings on T^2 and found that the T-duality group was $SL(2,\mathbf{Z}) \times SL(2,\mathbf{Z})$, one factor being the geometric symmetry of the 2-torus and one factor being stringy. In the U-duality group the geometric factor is enlarged to the $SL(3,\mathbf{Z})$ of the M-theory T^3.

Under compactification of more dimensions, it is harder to find a geometric interpretation of the U-duality group. The type II string on T^4, which is M-theory on T^5, has the U-duality symmetry

$$d = 6: \quad U = SO(5,5,\mathbf{Z}) . \tag{14.4.10}$$

This is the same as the T-duality of *string theory* on T^5. This is suggestive, but this identity holds only for T^5 so the connection if any will be intricate. For compactification of M-theory on T^k for $k \geq 6$, the U-duality symmetry is a discrete exceptional group, which has no simple geometric interpretation. A good interpretation of these symmetries would likely be an important step in understanding the nature of M-theory.

Returning to the IIB string in ten dimensions, it has been suggested that the $SL(2,\mathbf{Z})$ duality has a geometric interpretation in terms of two additional toroidal dimensions. This construction was christened *F-theory*. It is clear that these dimensions are not on the same footing as the eleventh dimension of M-theory, in that there is no limit of the parameters in which the spectrum becomes that of twelve noncompact dimensions. However, there may be some sense in which it is useful to begin with twelve dimensions and 'gauge away' one or two of them. Independent of this, F-theory has been a useful technique for finding solutions to

the field equations with nontrivial behavior of the dilaton and R–R scalar. As in eq. (12.1.30), these fields are joined in a complex parameter $\tau = C_0 + ie^{-\Phi}$ characterizing the complex structure of the additional 2-torus. Ten-dimensional solutions are then usefully written in terms of *twelve-dimensional* geometries.

IIA branes from eleven dimensions

The IIA theory has a rich spectrum of extended objects. It is interesting to see how each of these originates from compactification of M-theory on a circle. Let us first consider the extended objects of the eleven-dimensional theory. There is one tensor gauge field, the 3-form $B_{\mu\nu\rho}$. The corresponding electrically charged object is a 2-brane; in the literature the term *membrane* is used specifically for 2-branes. The magnetically charged object is a 5-brane. Of course the designations *electric* and *magnetic* interchange if we use instead a 6-form potential. However, $d = 11$ supergravity is one case in which one of the two Poincaré dual forms seems to be preferred (the 3-form) because the Chern–Simons term in the action cannot be written with a 6-form.

As in the discussion of the IIB NS5-brane, but with the dilaton omitted, we can always find a supersymmetric solution to the field equations having the appropriate charges. The M2- and M5-brane solutions are black p-branes, as described below eq. (14.1.14).

0-branes: The D0-branes of the IIA string are the BPS states of nonzero p_{10}. In M-theory these are the states of the massless graviton multiplet, an ultrashort multiplet of 2^8 states for each value of p_{10}.

1-branes: The 1-brane of the IIA theory is the fundamental IIA string. Its natural origin is as an M-theory supermembrane wrapped on the hidden dimension. As a check, such a membrane would couple to $B_{\mu\nu 10}$; this reduces to the NS–NS $B_{\mu\nu}$ field which couples to the IIA string. It was noted some time ago that the classical action of a wrapped M2-brane reduces to that of the IIA string.

2-branes: The obvious origin of the IIA D2-brane is as a transverse (rather than wrapped) M2-brane. The former couples to the R–R $B_{\mu\nu\rho}$, which is the reduction of the $d = 11$ $B_{\mu\nu\rho}$ to which the latter couples. Note that when written in terms of M-theory parameters, the D2-brane tension

$$\tau_{D2} = \frac{1}{(2\pi)^2 g \alpha'^{3/2}} = \frac{M_{11}^3}{(2\pi)^2} \tag{14.4.11}$$

depends only on the fundamental scale M_{11} and not on R_{10}, as necessary for an object that exists in the eleven-dimensional limit. On the other hand, the F-string tension

$$\tau_{F1} = \frac{1}{2\pi\alpha'} = 2\pi R_{10}\tau_{D2} \tag{14.4.12}$$

is linear in R_{10}, as should be the case for a wrapped object.

The D2-brane is perpendicular to the newly discovered 10-direction, and so should have a collective coordinate for fluctuations in that direction. This is puzzling, because D-branes in general have collective coordinates only for their motion in the ten-dimensional spacetime of perturbative string theory. However, the D2-brane is special, because in 2+1 dimensions a vector describes the same physics, by Poincaré duality, as a scalar. It is interesting to see this in detail. The bosonic action for a D2-brane in flat spacetime is

$$S[F,\lambda,X] = -\tau_2 \int d^3x \left\{ [-\det(\eta_{\mu\nu} + \partial_\mu X^m \partial_\nu X^m + 2\pi\alpha' F_{\mu\nu})]^{1/2} + \frac{\epsilon^{\mu\nu\rho}}{2}\lambda\partial_\mu F_{\nu\rho} \right\} . \tag{14.4.13}$$

We are treating $F_{\mu\nu}$ as the independent field and so include a Lagrange multiplier λ to enforce the Bianchi identity. In this form $F_{\mu\nu}$ is an auxiliary field (its equation of motion determines it completely as a local function of the other fields) and it can be eliminated with the result

$$S[\lambda,X] = -\tau_2 \int d^3x \left\{ -\det[\eta_{\mu\nu} + \partial_\mu X^m \partial_\nu X^m + (2\pi\alpha')^{-2}\partial_\mu\lambda\partial_\nu\lambda] \right\}^{-1/2} . \tag{14.4.14}$$

The algebra is left as an exercise. Defining $\lambda = 2\pi\alpha' X^{10}$, this is the action for a membrane in eleven dimensions. Somewhat surprisingly, it displays the full eleven-dimensional Lorentz invariance, even though this is broken by the compactification of X^{10}. This can be extended to the fermionic terms, and to membranes moving in background fields.

4-branes: These are wrapped M5-branes.

5-branes: The IIA theory, like the IIB theory, has a 5-brane solution carrying the magnetic NS–NS $B_{\mu\nu}$ charge. The solution is the same as in the IIA theory, because the actions for the NS–NS fields are the same. However, there is an interesting difference. Recall that a D1-brane can end on the IIB NS 5-brane. Under T-duality in a direction parallel to the 5-brane, we obtain a D2-brane ending on a IIA NS5-brane. From the point of view of the $(5+1)$-dimensional theory on the 5-brane, the end of a D1-brane in the IIB theory is a point, and is a source for the $U(1)$ gauge field living on the 5-brane. This is necessary so that the 5-brane can, through a Chern–Simons interaction, carry the R–R charge of the D1-brane. Similarly the end of the D2-brane in the IIA theory is a string in the 5-brane, and so should couple to a *2-form* field living on the IIA NS5-brane.

We were not surprised to find a $U(1)$ gauge field living on the IIB NS 5-brane, because it is related by S-duality to the IIB D5-brane which we know to have such a field. We cannot use this argument for the IIA NS 5-brane. However, in both cases the fields living on the world-sheet can be seen directly by looking at small fluctuations around the soliton solution. We do not have space here to develop in detail the soliton solutions and their properties, but we summarize the results. Modes that are normalizable in the directions transverse to the 5-brane correspond to degrees of freedom living on the 5-brane. These include the collective coordinates for its motion and in each case some R–R modes, which do indeed form a vector in the IIB case and a 2-form in the IIA case. The field strength of the 2-form is self-dual.

It is also interesting to look at this in terms of the unbroken supersymmetry algebras in the 5-brane world-volumes. Again we have space only to give a sketch. The supersymmetry variations of the gravitinos in a general background are

$$\delta\psi_M = D_M^- \zeta , \quad \delta\tilde{\psi}_M = D_M^+ \zeta . \tag{14.4.15}$$

Here $D_M^\pm$ is a covariant derivative where the spin connection ω is replaced with $\omega^\pm = \omega \pm \frac{1}{2}H$ with H the NS–NS 3-form field strength. We have already encountered $\omega^\pm$ in the world-sheet action (12.3.28). The difference of sign on the two sides occurs because H is odd under world-sheet parity. Under

$$SO(9,1) \to SO(5,1) \times SO(4) , \tag{14.4.16}$$

the ten-dimensional spinors decompose

$$\mathbf{16} \to (\mathbf{4},\mathbf{2}) + (\mathbf{4}',\mathbf{2}') , \tag{14.4.17a}$$

$$\mathbf{16}' \to (\mathbf{4},\mathbf{2}') + (\mathbf{4}',\mathbf{2}) . \tag{14.4.17b}$$

The nonzero components of the connection for the 5-brane solution lie in the transverse $SO(4) = SU(2) \times SU(2)$, and for the NS5-branes ω^+ and ω^- have the property that they lie entirely in the first or second $SU(2)$ respectively. A constant spinor carrying the second $SU(2)$ (that is, a $\mathbf{2}'$ of $SO(4)$) is then annihilated by D_M^+, and one carrying the first (a $\mathbf{2}$) by D_M^-; these correspond to unbroken supersymmetries. The left-moving supersymmetries transforming as a $\mathbf{2}$ of $SO(4)$ are thus unbroken — these are a $\mathbf{4}$ in both the IIA and IIB theories. Also unbroken are the right-moving supersymmetries transforming as a $\mathbf{2}'$ of $SO(4)$, which for the IIA

theory are a **4** and for the IIB theory a **4′**. In other words, the unbroken supersymmetry of the IIA NS5-brane is $d = 6$ (2,0) supersymmetry, and the unbroken supersymmetry of the IIB NS5-brane is $d = 6$ (1,1) supersymmetry. These supersymmetries are reviewed in section B.6. Curiously the nonchiral IIA theory has a chiral 5-brane, and the chiral IIB theory a nonchiral 5-brane.

These results fit with the fluctuation spectra. For the IIB NS5-brane the collective coordinates plus vector add up to a vector multiplet of $d = 6$ (1,1) supersymmetry. For the IIA NS5-brane, the only low-spin multiplet is the tensor, which contains the self-dual tensor argued for above and five scalars.

The obvious interpretation of the IIA NS5-brane is as an M-theory 5-brane that is transverse to the eleventh dimension. As in the discussion of the 2-brane, it should then have a collective coordinate for motion in this direction. Four of the scalars in the tensor multiplet are from the NS–NS sector and are collective coordinates for the directions perpendicular to the 5-brane that are visible in string perturbation theory. The fifth scalar, from the R–R sector, must be the collective coordinate for the eleventh dimension. It is remarkable that the 2-brane and the 5-brane of the IIA theory know that they secretly live in eleven dimensions.

The tension of the IIA NS5-brane is the same as that of the IIB NS5-brane,

$$\tau_{\mathrm{NS5}} = \frac{1}{(2\pi)^5 g^2 \alpha'^3} = \frac{\tau_{\mathrm{D2}}^2}{2\pi} = \frac{M_{11}^6}{(2\pi)^5} . \tag{14.4.18}$$

Like the tension of the D2-brane this is independent of R_{10}, as it must be for the eleven-dimensional interpretation,

$$\tau_{\mathrm{D2}} = \tau_{\mathrm{M2}} , \quad \tau_{\mathrm{NS5}} = \tau_{\mathrm{M5}} . \tag{14.4.19}$$

This also fits with the interpretation of the D4-brane,

$$\tau_{\mathrm{D4}} = 2\pi R_{10} \tau_{\mathrm{M5}} . \tag{14.4.20}$$

Since the IIA NS5-brane and D2-brane are both localized in the eleventh dimension, the configuration of a D2-brane ending on an NS5-brane lifts to an eleven-dimensional configuration of an M-theory 2-brane ending on an M-theory 5-brane. It is interesting to consider two nearby 5-branes with a 2-brane stretched between them, either in the IIA or M-theory context. The 2-brane is still extended in one direction and so behaves as a string. The tension is proportional to the distance r between the two 5-branes,

$$\tau_1 = r \tau_{\mathrm{M2}} . \tag{14.4.21}$$

In the IIB theory, the $r \to 0$ limit was a point of non-Abelian gauge

symmetry. Here it is something new, a *tensionless string theory*. For small r the lightest scale in the theory is set by the tension of these strings. They are entirely different from the strings we have studied: they live in six dimensions, they are not associated with gravity, and they have no adjustable coupling constant — their interactions in fact are of order 1. Of all the new phases of gauge and string theories that have been discovered this is perhaps the most mysterious, and may be a key to understanding many other things.

6-branes: The D6-brane field strength is dual to that of the D0-brane. Since the D0-brane carries Kaluza–Klein electric charge, the D6-brane must be a Kaluza–Klein magnetic monopole. Such an object exists as a soliton, where the Kaluza–Klein direction is not independent of the noncompact directions but is combined with them in a smooth and topologically nontrivial way. This is a local object in three noncompact spatial dimensions and so becomes a 6-brane in nine noncompact spatial dimensions.

8-branes: The eleven-dimensional origin of the D8-brane will be seen in the next section.

14.5 The $E_8 \times E_8$ heterotic string

The final ten-dimensional string theory is the $E_8 \times E_8$ heterotic string. We should be able to figure out its strongly coupled behavior, since it is T-dual to the $SO(32)$ heterotic string whose strongly coupled limit is known. We will need to trace through a series of T- and S-dualities before we come to a weakly coupled description. In order to do this we will keep track of how the moduli — the dilaton and the various components of the metric — are related at each step.

Recall that in each string theory the natural metric to use is the one that appears in the F-string world-sheet action. The various dualities interchange F-strings with other kinds of string, and the 'string metrics' in the different descriptions differ, as one sees explicitly in the IIB transformation (14.1.6) and the type I–heterotic transformation (14.3.4). After composing a series of dualities, one is interested in how the final dilaton and metric vary as the original dilaton becomes large. We seek to reach a description in which the final dilaton becomes small (or at least stays fixed), and in which the final radii grow (or at least stay fixed). A description in which the dilaton becomes small and also the radii become small is not useful, because the effective coupling in a small-radius theory

is increased by the contributions of light winding states. To get an accurate estimate of the coupling one must take the T-dual to a large-radius description.

T: Heterotic $E_8 \times E_8$ on S_1 to heterotic $SO(32)$ on S_1. Compactify the heterotic $E_8 \times E_8$ theory on a circle of large radius R_9 and turn on the Wilson line that breaks $E_8 \times E_8$ to $SO(16) \times SO(16)$. We will eventually take $R_9 \to \infty$ to get back to the ten-dimensional theory of interest, and then the Wilson line will be irrelevant. As discussed in section 11.6 this theory is T-dual to the $SO(32)$ heterotic string, again with a Wilson line breaking the group to $SO(16) \times SO(16)$. The couplings and radii are related

$$R_9' \propto R_9^{-1} , \quad g' \propto g R_9^{-1} . \tag{14.5.1}$$

Here primed quantities are for the $SO(32)$ theory and unprimed for the $E_8 \times E_8$ theory. We are only keeping track of the field dependence on each side, $R_9 \propto G_{99}^{1/2}$ and $g \propto e^{\Phi}$. The transformation of g follows by requiring that the two theories give the same answer for scattering of low energy gravitons. The low energy actions are proportional to

$$\frac{1}{g^2} \int d^{10}x = \frac{2\pi R_9}{g^2} \int d^9 x \tag{14.5.2}$$

and so $R_9/g^2 = R_9'/g'^2$.

S: Heterotic $SO(32)$ on S_1 to type I on S_1. Now use type I–heterotic duality to write this as a type I theory with

$$g_{\rm I} \propto g'^{-1} \propto g^{-1} R_9 , \quad R_{9\rm I} \propto g'^{-1/2} R_9' \propto g^{-1/2} R_9^{-1/2} . \tag{14.5.3}$$

The transformation of G_{99} follows from the field redefinition (14.3.4). We are interested in the limit in which g and R_9 are both large. It appears that we can make $g_{\rm I}$ small by an appropriate order of limits. However, the radius of the type I theory is becoming very small and so we must go to the T-dual description as warned above.

T: Type I on S_1 to type IIA on $S_1/\mathbf{Z}_2$. Consider a T-duality in the 9-direction of the type I theory. The compact dimension becomes a segment of length $\pi\alpha'/R_{9\rm I}$ with eight D8-branes at each end, and

$$g_{\rm I'} \propto g_{\rm I} R_{9\rm I}^{-1} \propto g^{-1/2} R_9^{3/2} , \quad R_{9\rm I'} \propto R_{9\rm I}^{-1} \propto g^{1/2} R_9^{1/2} . \tag{14.5.4}$$

If we are taking $g \to \infty$ at fixed R_9 then we have reached a good description. However, our real interest is the ten-dimensional theory at fixed large coupling. The coupling $g_{\rm I'}$ then becomes large, but one final duality brings us to a good description. The theory that we have reached is often called the type I′ theory. In the bulk, between the orientifold planes, it is the IIA theory, so we can also think of it as the IIA theory on the segment $S_1/\mathbf{Z}_2$. The coset must be an orientifold because the only

spacetime parity symmetry of the IIA theory also includes a world-sheet parity transformation.

S: Type IIA on $S_1/\mathbf{Z}_2$ to M-theory on $S_1/\mathbf{Z}_2$. The IIA theory is becoming strongly coupled, so the physics between the orientifold planes is described in terms of a new periodic dimension. The necessary transformations (12.1.9) were obtained from the dimensional reduction of $d = 11$ supergravity, giving

$$R_{10\rm M} \propto g_{\rm I'}^{2/3} \propto g^{-1/3} R_9 , \quad R_{9\rm M} \propto g_{\rm I'}^{-1/3} R_{9\rm I'} \propto g^{2/3} . \tag{14.5.5}$$

As the original R_9 is taken to infinity, the new R_{10} diverges linearly. Evidently we should identify the original 9-direction with the final 10-direction. Hence at the last step we also rename $(9, 10) \to (10', 9')$. The final dual for the strongly coupled $E_8 \times E_8$ theory in ten dimensions is M-theory, with ten noncompact dimensions and the 10′-direction, which is the transform of the original 9-direction, compactified. This is the same as the strongly coupled IIA theory. The difference is that here the 10′-direction is not a circle but a segment, with boundaries at the orientifold planes. M-theory on S_1 is the strongly coupled IIA theory. M-theory on $S_1/\mathbf{Z}_2$ is the strongly coupled $E_8 \times E_8$ heterotic theory. At each end are the orientifold plane and eight D8-branes, but now both are nine-dimensional as they bound a ten-dimensional space. The gauge degrees of freedom thus live in these walls, one E_8 in each wall.

The full sequence of dualities is

$$\text{heterotic } E_8 \times E_8 \xrightarrow{T_9} \text{heterotic } SO(32) \xrightarrow{S} \text{type I} \xrightarrow{T_9} \text{type I'} \xrightarrow{S} \text{M-theory} . \tag{14.5.6}$$

A heterotic string running in the 8-direction becomes

$$\mathrm{F}_8 \xrightarrow{T_9} \mathrm{F}_8 \xrightarrow{S} \mathrm{D}_8 \xrightarrow{T_9} \mathrm{D}_{89} \xrightarrow{S} \mathrm{M}_{8,10'} . \tag{14.5.7}$$

That is, it is a membrane running between the boundaries, as in figure 14.3. This whole picture is highly constrained by anomalies, and this in fact is how it was originally discovered. The $d = 11$ supergravity theory in a space with boundaries has anomalies unless the boundaries carry precisely E_8 degrees of freedom. Note also that

$$\mathrm{p}_9 \xrightarrow{T_9} \mathrm{F}_9 \xrightarrow{S} \mathrm{D}_9 \xrightarrow{T_9} \mathrm{D}_\phi \xrightarrow{S} \mathrm{p}_{10} = \mathrm{p}_{9'} . \tag{14.5.8}$$

This confirms the identification of the original 9-direction and final 10-direction.

Let us comment on the D8-branes. In string theory the D8-brane is a source for the dilaton. To first order the result is a constant gradient for the dilaton (since the D8-brane has codimension one), but the full nonlinear supergravity equations for the dilaton, metric, and R–R 9-form imply that

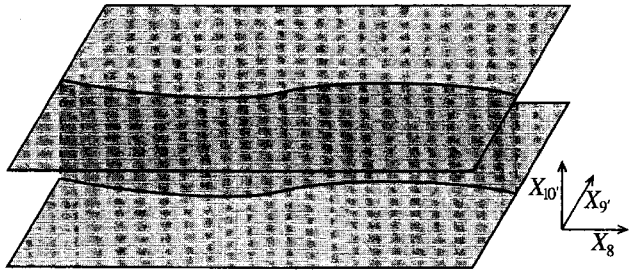


Fig. 14.3. Strongly coupled limit of $E_8 \times E_8$ heterotic string theory, with one heterotic string shown. The shaded upper and lower faces are boundaries. In the strongly coupled IIA string the upper and lower faces would be periodically identified.

the dilaton diverges a finite distance from the D8-brane. To cure this, one must run into a boundary (orientifold plane) before the divergence: this sets a maximum distance between the D8-brane and the boundary. As one goes to the strongly coupled limit, the initial value for the dilaton is greater and so this distance is shorter. In the strongly coupled limit the D8-branes disappear into the boundary, and in the eleven-dimensional theory there is no way to pull them out. The moduli for their positions just become Wilson lines for the gauge theory in the boundary.

We have now determined the strongly coupled behaviors of all of the ten-dimensional string theories. One can apply the same methods to the compactified theories, and we will do this in detail for toroidal compactifications of the heterotic string in section 19.9. Almost all of that section can be read now; we defer it because to complete the discussion we will need some understanding of strings moving on the smooth manifold K3.

14.6 What is string theory?

What we have learned is shown in figure 14.4. There is a single theory, and all known string theories arise as limits of the parameter space, as does M-theory with 11 noncompact spacetime dimensions. For example, if one starts with the type I theory on T^2, then by varying the two radii, the string coupling, and the Wilson line in one of the compact directions, one can reach the noncompact weakly coupled limit of any of the other string theories, or the noncompact limit of M-theory. Figure 14.4 shows a two-dimensional slice through this four-parameter space. This is only

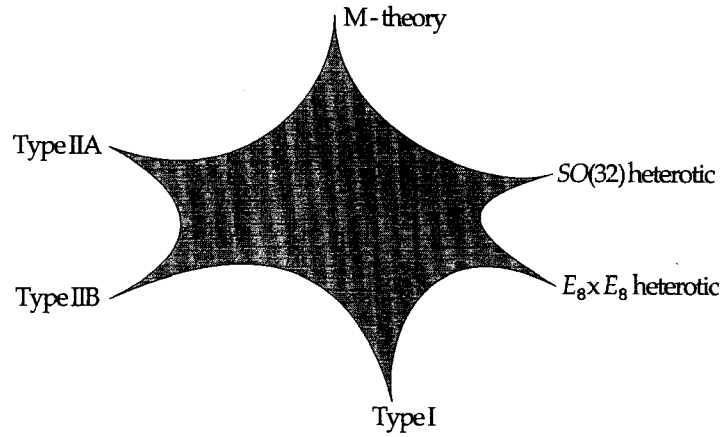


Fig. 14.4. All string theories, and M-theory, as limits of one theory.

one of many branches of the moduli space, and one with a fairly large number of unbroken supersymmetries, 16.

The question is, what is the theory of which all these things are limits? On the one hand we know a lot about it, in that we are able to put together this picture of its moduli space. On the other hand, over most of moduli space, including the M-theory limit, we have only the low energy effective field theory. In the various weakly coupled string limits, we have a description that is presumably valid at all energies but only as an asymptotic expansion in the coupling. This is very far from a complete understanding.

As an example of a question that we do not know how to answer, consider graviton–graviton scattering with center-of-mass energy E. Let us suppose that in moduli space we are near one of the weakly coupled string descriptions, at some small but finite coupling g. The ten-dimensional gravitational constant is of order $G_N \sim g^2\alpha'^4$. The Schwarzchild radius of the system is of order $R \sim (G_N E)^{1/7}$. One would expect that a black hole would be produced provided that R is large compared to the Compton wavelength E^{-1} and also to the string length $\alpha'^{1/2}$. The latter condition is more stringent, giving

$$E \gtrsim \frac{1}{g^2\alpha'^{1/2}} \,. \tag{14.6.1}$$

At this scale, these considerations show that the interactions are strong and string perturbation theory has broken down. Moreover, we do not

even *in principle* have a way to study the scattering, as we should in a complete theory. Of course, this process is so complicated that we would not expect to obtain an analytic description, but a criterion for a complete understanding of the theory is that we could in principle, with a large enough and fast enough computer, answer any question of this sort. In our present state of knowledge we cannot do this. We could only instruct the computer to calculate many terms in the string perturbation series, but each term would be larger than the one before it, and the series would tell us nothing. This particular process is of some interest, because there are arguments that it cannot be described by ordinary quantum mechanics and requires a generalization in which pure states evolve to density matrices. We will briefly discuss this issue in the next section.

Even if one is only interested in physics at accessible energies, it is likely that to understand the nonsupersymmetric state in which we live will require a complete understanding of the dynamics of the theory. In the case of quantum field theory, to satisfy Wilson's criterion of 'computability in principle' required an understanding of the renormalization group, and this in turn gave much more conceptual insight into the dynamics of the theory.

One possibility is that each of the string theories (or perhaps, just some of them) can be given a nonperturbative definition in the form of string field theory, so that each would give a good nonperturbative definition. The various dualities would then amount to changes of variables from one theory to another. However, there are various reasons to doubt this. The most prominent is simply that string field theory has not been successful — it has not allowed us to calculate anything we did not already know how to calculate using string perturbation theory. Notably, all the recent progress in understanding nonperturbative physics has taken place without the aid of string field theory, and no connection between the two has emerged. On the contrary, the entire style of argument in the recent developments has been that there are different effective descriptions, each with its own range of validity, and there is no indication that in general any description has a wider range of validity than it should. That is, a given string theory is a valid effective description only near the corresponding cusp of figure 14.4. And if strings are the wrong degrees of freedom for writing down the full Hamiltonian, no bookkeeping device like string field theory will give a satisfactory description. We should also note that even in quantum field theory, where we have a nonperturbative definition, this idea of understanding dualities as changes of variables seems to work only in simple low dimensional examples. Even in field theory the understanding of duality is likely to require new ideas.

However, there must be some exact definition of the theory, in terms of some set of variables, because the graviton scattering question must have an answer in principle. The term M-theory, originally applied to the eleven-dimensional limit, has now come to denote the complete theory.

14.7 Is M for matrix?

A notable feature of the recent progress has been the convergence of many lines of work, as the roles of such constructions as D-branes, string solitons, and $d = 11$ supergravity have been recognized. It is likely that the correct degrees of freedom for M-theory are already known, but their full significance not appreciated. Indeed, one promising proposal is that D-branes, specifically D0-branes, are those degrees of freedom.

According to our current picture, D-branes give a precise description of part of the spectrum, the R–R charged states, but only near the cusps where the type I, IIA, and IIB strings are weakly coupled — elsewhere their relevance comes only from the usual supersymmetric continuation argument. To extend this to a complete description covering the whole parameter space requires some cleverness. The remainder of this section gives a description of this idea, *matrix theory*.

Consider a state in the IIA theory and imagine boosting it to large momentum in the hidden X^{10} direction. Of course 'boosting' is a deceptive term because the compactification of this dimension breaks Lorentz invariance, but at least at large coupling (and so large R_{10}), we should be able to make sense of this. The energy of a particle with n units of compact momentum is

$$E = (p_{10}^2 + q^2 + m^2)^{1/2} \approx p_{10} + \frac{q^2 + m^2}{2p_{10}} = \frac{n}{R_{10}} + \frac{R_{10}}{2n}(q^2 + m^2) \,. \quad (14.7.1)$$

Here q is the momentum in the other nine spatial dimensions. Recalling the connection between p_{10} and D0-brane charge, this is a state of n D0-branes, and the first term in the action is the D0-brane rest mass. Large boost is large n/R_{10}. In this limit, the second term in the energy is quite small. States that have finite energy in the original frame have

$$E - n/R_{10} = O(R_{10}/n) \quad (14.7.2)$$

in the boosted frame. There are very few string states with the property (14.7.2). For example, even adding massless closed strings would add an energy q, which does not go to zero with R_{10}/n. Excited open strings connected to the D0-branes also have too large an energy. Thus it seems that we can restrict to ground state open strings attached to the D0-branes.

The Hamiltonian for these was given in eq. (13.5.14), which we now write in terms of the M-theory parameters M_{11} and R_{10}:

$$H = R_{10}\mathrm{Tr}\left\{\frac{1}{2}p_i p_i - \frac{M_{11}^6}{16\pi^2}[X^i, X^j]^2 - \frac{M_{11}^3}{4\pi}\lambda\Gamma^0\Gamma^i[X^i,\lambda]\right\} . \qquad (14.7.3)$$

We have dropped higher powers of momentum coming from the Born–Infeld term because all such corrections are suppressed by the boost, just as the square root in the energy (14.7.1) simplifies. Also, we drop the additive term n/R_{10} from H.

The Hamiltonian (14.7.3) is conjectured to be the complete description of systems with $p_{10} = n/R_{10} \gg M_{11}^{-1}$. Now take R_{10} and n/R_{10} to infinity, to describe a highly boosted system in eleven noncompact dimensions. By eleven-dimensional Lorentz invariance, we can put any system in this frame, so this should be a complete description of the whole of M-theory! This is the *matrix theory* proposal. We emphasize that this is a conjecture, not a derivation: we can derive the Hamiltonian (14.7.3) only at weak string coupling, where we know what the theory is. In effect, we are taking a specific result derived at the IIA cusp of figure 14.4 and conjecturing that it is valid over the whole moduli space.

This is a remarkably simple and explicit proposal: the nine $n\times n$ matrices X^i_{ab} are all one needs. As one check, let us recall the observation from the previous chapter that only one length scale appears in this Hamiltonian, $g^{1/3}\alpha'^{1/2}$. This is the minimum distance that can be probed by D0-brane scattering, and now in light of M-theory we see that this scale has another interpretation — it is M_{11}^{-1}, the eleven-dimensional Planck length. This is the fundamental length scale of M-theory, and so the only one that should appear.

At first sight, the normalization of the Hamiltonian (14.7.3) seems to involve another parameter, R_{10}. Recall, however, that the system is boosted and so internal times are dilated. The boost factor is proportional to p_{10}, so the time-scale should be divided by a dimensionless factor $p_{10}/M_{11} = n/M_{11}R_{10}$, and again only the scale M_{11} appears.

The description of the eleven-dimensional spacetime in matrix theory is rather asymmetric: time is the only explicit coordinate, nine spatial dimensions emerge from matrix functions of time, and the last dimension is the Fourier transform of n. This asymmetric picture is similar to the light-cone gauge fixing of a covariant theory.

Now let us discuss some of the physics. As in the discussion of IIA–M-theory duality, a graviton of momentum $p_{10} = n/R_{10}$ is a bound state of n D0-branes. Again, the existence of these bound states is necessary for M-theory to be correct, and has been shown in part. For a bound state of total momentum q_i, the $SU(n)$ dynamics is responsible for the zero-energy

bound state, and the center-of-mass energy from the $U(1)$ part $p_i = q_i I_n/n$ is

$$E = \frac{R_{10}}{2}\mathrm{Tr}(p_i p_i) = \frac{q^2}{2p_{10}} , \qquad (14.7.4)$$

correctly reproducing the energy (14.7.1) for a massless state.

Now let us consider a simple interaction, graviton–graviton scattering. Let the gravitons have 10-momenta $p_{10} = n_{1,2}/R_{10}$ and be at well-separated positions $Y^i_{1,2}$. The total number of D0-branes is n_1+n_2, and the coordinate matrices X^i are approximately block diagonal. Write X^i as

$$X^i = X^i_0 + x^i \qquad (14.7.5a)$$

$$X^i_0 = Y^i_1 I_1 + Y^i_2 I_2 , \quad x^i = x^i_{11} + x^i_{22} + x^i_{12} + x^i_{21} . \qquad (14.7.5b)$$

Here I_1 and I_2 are the identity matrices in the two blocks, which are respectively $n_1 \times n_1$ and $n_2 \times n_2$, and we have separated the fluctuation x^i into a piece in each block plus off-diagonal pieces. First setting the off-diagonal $x^i_{12,21}$ to zero, the blocks decouple because $[x^i_{11}, x^j_{22}] = 0$. The wavefunction is then a product of the corresponding bound state wavefunctions,

$$\psi(x_{11}, x_{22}) = \psi_0(x_{11})\psi_0(x_{22}) . \qquad (14.7.6)$$

Now consider the off-diagonal block. These degrees of freedom are heavy: the commutator

$$[X^i_0, x^j_{12}] = (Y^i_1 - Y^i_2)x^j_{12} \qquad (14.7.7)$$

gives them a mass proportional to the separation of the gravitons. Thus we can integrate them out to obtain the effective interaction between the gravitons.

We would like to use this to test the matrix theory proposal, to see that the effective interaction at long distance agrees with eleven-dimensional supergravity. In fact, we can do this without any further calculation: all the necessary results can be extracted from the cylinder amplitude (13.5.6). At distances small compared to the string scale, the cylinder is dominated by the lightest open strings stretched between the D0-branes, which are precisely the off-diagonal matrix theory degrees of freedom. At distances long compared to the string scale, the cylinder is dominated by the lightest closed string states and so goes over to the supergravity result. This is ten-dimensional supergravity, but it is equivalent to the answer from eleven-dimensional supergravity for the following reason. In the process we are studying, the sizes of the blocks stay fixed at n_1 and n_2, meaning that the values of p_{10} and p'_{10} do not change in the scattering and the p_{10} of the exchanged graviton is zero. This has the effect of averaging over x^{10} and so giving the dimensionally reduced answer.

Finally, we should keep only the leading velocity dependence from the cylinder, because the time dilation from the boost suppresses higher powers as in eq. (14.7.1). The result (13.5.7) for $p = 0$, multiplying by the number of D0-branes in each clump, is

$$V(r,v) = 4\pi^{5/2}\Gamma(7/2)\alpha'^3 n_1 n_2 \frac{v^4}{r^7} = \frac{15\pi^3}{2}\frac{p_{10}p'_{10}}{M_{11}^9 R_{10}}\frac{v^4}{r^7} . \tag{14.7.8}$$

Because the functional form is the same at large and small r, the matrix theory correctly reproduces the supergravity amplitude (in the matrix theory literature, the standard convention is $M_{11} = (2\pi)^{-1/3}M_{11}$ (here), which removes all 2πs from the matrix theory Hamiltonian).

This is an interesting result, but its significance is not clear. Some higher order extensions do not appear to work, and it may be that one must take the large n limit to obtain agreement with supergravity. The loop expansion parameter in the quantum mechanics is then large, so perturbative calculations are not sufficient. Also, the process being studied here, where the p_{10} of the exchanged graviton vanishes, is quite special. When this is not the case, one has a very different process where the sizes of the blocks change, meaning that D0-branes move from one clump to the other; this appears to be much harder to study.

Matrix theory, if correct, satisfies the 'computability' criterion: we can in principle calculate graviton–graviton scattering numerically at any energy. The analytic understanding of the bound states is still limited, but in principle they could be determined numerically to any desired accuracy and then the wavefunction for the two-graviton state evolved forward in time. Of course any simulation is at finite n and R_{10}, and the matrix theory proposal requires that we take these to infinity; but if the proposal is correct then the limits exist and can be taken numerically. For now all this is just a statement in principle, as various difficulties make the numerical calculation impractical. Most notable among these is the difficulty of preserving to sufficient precision the supersymmetric cancellations that are needed for the theory to make sense — for example, along the flat directions of the potential.

The M-theory membrane

If the matrices X^i are a complete set of degrees of freedom, then it must be possible to identify all the known states of M-theory, in particular the membranes. We might have expected that these would require us to add explicit D2-brane degrees of freedom, but remarkably the membranes are already present as excitations of the D0-brane Hamiltonian.

To see this, define the $n \times n$ matrices

$$U = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots \\ 0 & \alpha & 0 & 0 & \\ 0 & 0 & \alpha^2 & 0 & \\ 0 & 0 & 0 & \alpha^3 & \\ \vdots & & & & \ddots \end{bmatrix}, \quad V = \begin{bmatrix} 0 & 0 & 0 & & 1 \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & & 0 \\ 0 & 0 & 1 & & 0 \\ & \vdots & & \ddots & \end{bmatrix}, \tag{14.7.9}$$

where $\alpha = \exp(2\pi i/n)$. These have the properties

$$U^n = V^n = 1 , \quad UV = \alpha VU , \tag{14.7.10}$$

and these properties determine U and V up to change of basis. The matrices $U^r V^s$ for $1 \le r,s \le n$ form a complete set, and so any matrix can be expanded in terms of them. For example,

$$X^i = \sum_{r,s=1}^{n} X^i_{rs} U^r V^s , \tag{14.7.11}$$

and similarly for the fermion λ. To each matrix we can then associate a periodic function of two variables,

$$X^i \to X^i(p,q) = \sum_{r,s=1}^{n} X^i_{rs} \exp(ipr + iqs) . \tag{14.7.12}$$

If we focus on matrices which remain smooth functions of p and q as n becomes large (so that the typical r and s remain finite), then the commutator maps

$$[X^i, X^j] \to \frac{i}{n}(\partial_q X^i \partial_p X^j - \partial_p X^i \partial_q X^j) + O(n^2) \equiv \frac{i}{n}\{X^i, X^j\}_{\rm PB} + O(n^2) . \tag{14.7.13}$$

One can verify this by considering simple monomials $U^r V^s$. Notice the analogy to taking the classical limit of a quantum system, with the Poisson bracket appearing. One can also rewrite the trace as an integral,

$$\mathrm{Tr} = n \int dq\, dp . \tag{14.7.14}$$

The Hamiltonian becomes

$$R_{10} \int dq\, dp \left(\frac{n}{2}\Pi_i\Pi_i + \frac{M_{11}^6}{16\pi^2 n}\{X^i, X^j\}^2_{\rm PB} - i\frac{M_{11}^3}{4\pi}\lambda\Gamma^0\Gamma^i\{X^i,\lambda\}_{\rm PB} \right) . \tag{14.7.15}$$

Since $X^i(p,q)$ is a function of two variables, this Hamiltonian evidently describes the quantum mechanics of a membrane. In fact, it is identical to the Hamiltonian one gets from an eleven-dimensional supersymmetric

membrane action in the light-cone gauge. We do not have space to develop this in detail, but as an example consider the static configuration

$$X^1 = aq\ , \quad X^2 = bp\ ; \tag{14.7.16}$$

since q and p are periodic we must also suppose $X^{1,2}$ to be as well. Then the energy is

$$\frac{M_{11}^6 R_{10} a^2 b^2}{2n} = \frac{M_{11}^6 A^2}{2(2\pi)^4 p_{10}} = \frac{\tau_{M2}^2 A^2}{2p_{10}}\ . \tag{14.7.17}$$

Here $A = 4\pi^2 ab$ is the area of the membrane. The product $\tau_{M2}A$ is the mass of an M-theory membrane of this area, so this agrees with the energy (14.7.1).

There was at one time an effort to define eleven-dimensional supergravity as a theory of fundamental membranes; this was one of the roots of the name M-theory. This had many difficulties, the most immediate being that the world-sheet theory is nonrenormalizable. However, it was noted that the light-cone Hamiltonian (14.7.15) was the large-n limit of dimensionally reduced $d = 10$, $N = 1$ gauge theory (14.7.3), so the finite-n theory could be thought of as regularizing the membrane. Matrix theory puts this idea in a new context. One of the difficulties of the original interpretation was that the potential has flat directions, for example

$$X^i = Y_1^i I_1 + Y_2^i I_2 \tag{14.7.18}$$

as in eq. (14.7.5). This implies a continuous spectrum, which is physically unsatisfactory given the original interpretation of the Hamiltonian as arising from gauge-fixing the action for a single membrane. However, we now interpret the configuration (14.7.18) as a two-particle state. The continuous spectrum is not a problem because the matrix theory is supposed to describe states with arbitrary numbers of particles. We should emphasize that in focusing on matrices that map to smooth functions of p and q we have picked out just a piece of the matrix theory spectrum, namely states of a single membrane of toroidal topology. Other topologies, other branes, and graviton states are elsewhere.

Since matrix theory is supposed to be a complete formulation of M-theory, it must in particular reproduce all of string theory. It is surprising that it can do this starting with just nine matrices, but we now see how it is possible — it contains membranes, and strings are just wrapped membranes. The point is that one can hide a great deal in a large matrix! If we compactify one of the nine X^i dimensions, the membranes wrapped in this direction reproduce string theory; arguments have been given that the string interactions are correctly incorporated.

Finite n and compactification

In arguing for matrix theory we took n to be large. Let us also ask, does the finite-n matrix theory Hamiltonian have any physical relevance? In fact it describes M-theory compactified in a *lightlike* direction,

$$(x^0, x^{10}) \cong (x^0 - \pi\tilde{R}_{10}, x^{10} + \pi\tilde{R}_{10})\ . \tag{14.7.19}$$

To see this — in fact, to define it — let us reach this theory as the limit of spacelike compactification,

$$(x^0, x^{10}) \cong (x^0 - \pi\tilde{R}_{10}, x^{10} + \pi\tilde{R}_{10} + 2\pi\epsilon^2\tilde{R}_{10}) \tag{14.7.20}$$

with $\epsilon \to 0$. The invariant length of the compact dimension is $2\pi\epsilon\tilde{R}_{10}$, so this is Lorentz-equivalent to the spacelike compactification

$$(x'^0, x'^{10}) \cong (x'^0, x'^{10} + 2\pi\epsilon\tilde{R}_{10})\ , \tag{14.7.21}$$

where

$$x'^0 \pm x'^{10} = \epsilon^{\mp 1}(x^0 \pm x^{10})\ . \tag{14.7.22}$$

Unlike the $n \to \infty$ conjecture, the finite-n conjecture can actually be derived from things that we already know. Because the invariant radius (14.7.21) is going to zero, we are in the regime of weakly coupled IIA string theory. Moreover, states that have finite energy in the original frame acquire

$$E',\ p'_{10} \propto O(\epsilon^{-1})\ , \quad E' - p'_{10} \propto O(\epsilon) \tag{14.7.23}$$

under the boost (14.7.22). These are the only states that we are to retain. However, we have already carried out this exercise at the beginning of this section: this is eq. (14.7.2) where

$$R_{10} = \epsilon\tilde{R}_{10}\ . \tag{14.7.24}$$

The derivation of the matrix theory Hamiltonian then goes through just as before, and it is surely correct because we are in weakly coupled string theory. The lightlike theory is often described as the *discrete light-cone quantization (DLCQ)* of M-theory, meaning light-cone quantization with a discrete spectrum of p_-. This idea has been developed in field theory, but one must be careful because the definition there is generally not equivalent to the lightlike limit.

Of course, the physics in a spacetime with lightlike compactification may be rather exotic, so this result does not directly enable us to understand the eleven-dimensional theory which is supposed to be recovered in the large-n limit. However, it has been very valuable in understanding how the matrix theory conjecture is to be extended to the case that some of the additional dimensions are compactified. Let us consider,

for example, the case that k dimensions are periodic. Working in the frame (14.7.21), we are instructed to take R_{10} to zero holding fixed M_{11} and all momenta and distances in the transverse directions (those other than x^{10}). We then keep only states whose energy is $O(R_{10})$ above the BPS minimum. These clearly include the gravitons (and their superpartners) with nonzero p_{10}, which are just the D0-branes. In addition, let us consider M2-branes that are wrapped around the 10-direction and around one of the transverse directions x^m. From the IIA point of view these are F-strings winding in the x^m-direction. They have mass equal to $\tau_{M2}A = 4\pi^2 R_m R_{10}$ and so are candidates to survive in the limit. However, for M2-branes with vanishing p_{10}, $E = (q^2 + m^2)^{1/2}$ and we also need that they have zero momentum in the noncompact directions — this is a point of measure zero. The only membrane states that survive are M2-branes with nonzero p_{10}, which are F-strings that end on D0-branes in the IIA description. These F-strings must be in their ground states, but they can wind any number of times around the transverse compact directions. The lightlike limit now has many more degrees of freedom than in the noncompact case, because there is an additional winding quantum number for each compact dimension. In fact, it is simpler to use the T-dual description, where the D0-branes become Dk-branes and the winding number becomes momentum: the lightlike limit of matrix theory then includes the full $(k+1)$-dimensional $U(n)$ Yang–Mills theory on the branes.

It is notable that the number of degrees of freedom goes up drastically with compactification of each additional dimension, as the dimension of the effective gauge theory increases. A difficulty is that for $k > 3$ the gauge theory on the brane is nonrenormalizable. However, for $k > 3$ our discussion of the lightlike limit is incomplete. In the first place, we have not considered all the degrees of freedom. For $k \geq 4$, an M5-brane that wraps the x^{10}-direction and four of the transverse directions also survives the limit. Moreover, the coupling of the T-dual string theory,

$$\frac{R_{10}}{\alpha'^{1/2}} \prod_m \frac{\alpha'^{1/2}}{R_m} = R_{10}^{(3-k)/2} (M_{11}^3/2\pi)^{(1-k)/2} \prod_m R_m^{-1} , \tag{14.7.25}$$

diverges as $R_{10} \to 0$ for $k \geq 4$. The lightlike limit is then no longer a weakly coupled string theory, and it is necessary to perform further dualities. The various cases $k \geq 4$ are quite interesting, but we must leave the details to the references.

In summary, the various compactifications of matrix theory suggest a deep relation between large-n gauge theory and string theory. Such a relation has arisen from various other points of view, and may lead to a better understanding of gauge theories as well as string theory.

14.8 Black hole quantum mechanics

In the early 1970s it was found that classical black holes obey laws directly analogous to the laws of thermodynamics. This analogy was made sharper by Hawking's discovery that black holes radiate as black bodies at the corresponding temperature. Under this analogy, the entropy of a black hole is equal to the area of its event horizon divided by $\kappa^2/2\pi$. The analogy is so sharp that it has long been a goal to find a statistical mechanical theory associated with this thermodynamics, and in particular to associate the entropy with the density of states of the black hole. Many arguments have been put forward in this direction but until recently there was no example where the states of a black hole could be counted in a controlled way.

This has now been done for some string theory black holes. To see the idea, let us return to the relation between a D-brane and an R–R charged black p-brane. The thermodynamic and other issues are the same for black p-branes as for black holes. The explicit solution for a black p-brane with Q units of R–R charge is (for $p \leq 6$)

$$ds^2 = Z(r)^{-1/2} \eta_{\mu\nu} dx^\mu dx^\nu + Z(r)^{1/2} dx^m dx^m , \tag{14.8.1a}$$

$$e^{2\Phi} = Z(r)^{(3-p)/2} . \tag{14.8.1b}$$

Here x^μ is tangent to the p-brane, x^m is transverse, and

$$Z(r) = 1 + \frac{\rho^{7-p}}{r^{7-p}} , \quad r^2 = x^m x^m , \tag{14.8.2a}$$

$$\rho^{7-p} = \alpha'^{(7-p)/2} g Q (4\pi)^{(5-p)/2} \Gamma\left(\frac{7-p}{2}\right) . \tag{14.8.2b}$$

The numerical constant, which is not relevant to the immediate discussion, is obtained in exercise 14.6. The characteristic length ρ is shorter than the string scale when gQ is less than 1. In this case, the effective low energy field theory that we have used to derive the solution (14.8.1) is not valid. When gQ is greater than 1 the geometry is smooth on the string scale and the low energy field theory should be a good description.

Consider now string perturbation theory in the presence of Q coincident D-branes. The expansion parameter is gQ: each additional world-sheet boundary brings a factor of the string coupling g but also a factor of Q from the sum over Chan–Paton factors. When gQ is small, string perturbation theory is good, but when it is large it breaks down. Thus the situation appears to be very much as with the instanton in section 13.6: in one range of parameters the low energy field theory description is good, and in another range the D-brane description is good.

In the instanton case we can continue from one regime to the other by varying the instanton scale factor. In the black p-brane case we can do the

same by varying the string coupling, as we have often done in the analysis of strongly coupled strings. We can use this to count supersymmetric (BPS) black hole states: we can do the counting at small gQ, where the weakly coupled D-brane description is good, and continue to large gQ, where the black hole description is accurate.

The particular solution (14.8.1) is not useful for a test of the black hole entropy formula because the event horizon, at $r = 0$, is singular. It can be made nonsingular by adding energy to give a nonsupersymmetric black hole, but in the supersymmetric (extremal) limit the area goes to zero. To obtain a supersymmetric black p-brane with a smooth horizon of nonzero area requires at least three nonzero charges. A simple example combines Q_1 D1-branes in the 5-direction with Q_5 D5-branes in the (5,6,7,8,9)-directions. To make the energy finite the (6,7,8,9)-directions are compactified on a T^4 of volume V_4. We also take the 5-direction to be finite, but it is useful to keep its length L large. The third charge is momentum p_5. The solution is

$$ds^2 = Z_1^{-1/2}Z_5^{-1/2}\left[\eta_{\mu\nu}dx^\mu dx^\nu + (Z_n - 1)(dt + dx_5)^2\right] + Z_1^{1/2}Z_5^{1/2}dx^i dx^i + Z_1^{1/2}Z_5^{-1/2}dx^m dx^m , \quad (14.8.3a)$$

$$e^{-2\Phi} = Z_5/Z_1 . \quad (14.8.3b)$$

Here μ, ν run over the (0,5)-directions tangent to all the branes, i runs over the (1,2,3,4)-directions transverse to all branes, and m runs over the (6,7,8,9)-directions tangent to the D5-branes and transverse to the D1-branes. We have defined

$$Z_1 = 1 + \frac{r_1^2}{r^2} , \quad r_1^2 = \frac{(2\pi)^4 g Q_1 \alpha'^3}{V_4} , \quad (14.8.4a)$$

$$Z_5 = 1 + \frac{r_5^2}{r^2} , \quad r_5^2 = g Q_5 \alpha' , \quad (14.8.4b)$$

$$Z_n = 1 + \frac{r_n^2}{r^2} , \quad r_n^2 = \frac{(2\pi)^5 g^2 p_5 \alpha'^4}{L V_4} , \quad (14.8.4c)$$

with $r^2 = x^i x^i$. The event horizon is at $r = 0$; the interior of the black hole is not included in this coordinate system. The integers Q_1, Q_5, and $n_5 = p_5 L/2\pi$ are all taken to be large so that this describes a classical black hole, with horizon much larger than the Planck scale.

The solution (14.8.3) is in terms of the string metric. The black hole area law applies to the Einstein metric $G_{E\mu\nu} = e^{-\Phi/2}G_{\mu\nu}$, whose action is field-independent. This is

$$ds_E^2 = Z_1^{-3/4}Z_5^{-1/4}\left[\eta_{\mu\nu}dx^\mu dx^\nu + (Z_n - 1)(dt + dx_5)^2\right] + Z_1^{1/4}Z_5^{3/4}dx^i dx^i + Z_1^{1/4}Z_5^{-1/4}dx^m dx^m . \quad (14.8.5)$$

Now let us determine the horizon area. The eight-dimensional horizon is a 3-sphere in the transverse dimensions and is extended in the (5,6,7,8,9)-directions. Near the origin the angular metric is

$$\left(\frac{r_1^2}{r^2}\right)^{1/4}\left(\frac{r_5^2}{r^2}\right)^{3/4} r^2 d\Omega^2 = r_1^{1/2} r_5^{3/2} d\Omega^2 , \quad (14.8.6)$$

with total area $2\pi^2 r_1^{3/4} r_5^{9/4}$. From $G_{55} = Z_1^{-3/4}Z_5^{-1/4}Z_n$ it follows that the invariant length of the horizon in the 5-direction is $r_1^{-3/4}r_5^{-1/4}r_n L$. Similarly the invariant volume in the toroidal directions is $r_1 r_5^{-1} V_4$. The area is the product

$$A = 2\pi^2 L V_4 r_1 r_5 r_n = 2^6 \pi^7 g^2 \alpha'^4 (Q_1 Q_5 n_5)^{1/2} = \kappa^2 (Q_1 Q_5 n_5)^{1/2} . \quad (14.8.7)$$

This gives for the black hole entropy

$$S = \frac{2\pi A}{\kappa^2} = 2\pi (Q_1 Q_5 n_5)^{1/2} . \quad (14.8.8)$$

The final result is quite simple, depending only on the integer charges and not on any of the moduli g, L, or V_4. This is a reflection of the classical black hole area law: under adiabatic changes in the moduli the horizon area cannot change.

Now let us consider the same black hole in the regime where the D-brane picture is valid. The dynamics of the $\#_{ND} = 4$ system was discussed in chapter 13, and in particular the potential is

$$V = \frac{1}{(2\pi\alpha')^2}|X_i\chi - \chi Y_i|^2 + \frac{g_1^2}{4}D_1^A D_1^A + \frac{g_5^2}{4V_4}D_5^A D_5^A . \quad (14.8.9)$$

This is generalized from the earlier (13.6.25) because there are multiple D1-branes and D5-branes. Thus in the first term the $Q_1 \times Q_1$ D1-brane collective coordinates X_i act on the left of the $Q_1 \times Q_5$ matrix χ, and the $Q_5 \times Q_5$ D5-brane collective coordinates Y_i act on the right. The black hole is a bound state of D1- and D5-branes, so the χ are nonzero. The first term in the potential then requires that

$$X^i = x^i I_{Q_1} , \quad Y^i = x^i I_{Q_5} , \quad (14.8.10)$$

and the center-of-mass x^i is the only light degree of freedom in the transverse directions. Also, the 1-1 X_m and 5-5 A_m are now charged under the $U(Q_1)$ and $U(Q_5)$ and so contribute to the D-terms in the general form (B.7.3).[4] What is important is the dimension of the moduli space, which can be determined by counting. The X_m contribute $4Q_1^2$ real scalars,

[4] The A^4 term is just a rewriting of the $[A_m, A_n]^2$ term from the dimensional reduction, and similarly for the X^4.

the A_m contribute $4Q_5^2$, and the χ contribute $4Q_1Q_5$. The vanishing of the D-terms imposes $3Q_1^2 + 3Q_5^2$ conditions; since the Qs are large we do not worry about the $U(1)$ parts, which are only $1/Q^2$ of the total. Also, the $U(Q_1)$ and $U(Q_5)$ gauge equivalences remove another $Q_1^2 + Q_5^2$ moduli, leaving $4Q_1Q_5$. This is a generalization of the counting that we did for the instanton in section 13.6.

These moduli are functions of x^5 and x^0. We are treating L as very large, but the counting extends to small L with some subtlety. So we have a two-dimensional field theory with $4Q_1Q_5$ real scalars and by supersymmetry $4Q_1Q_5$ Majorana fermions, and we need its density of states. This is a standard calculation, which in fact we have already done. For a CFT of central charge c, the general relation (7.2.30) between the central charge and the density of states gives

$$\mathrm{Tr}[\exp(-\beta H)] \approx \exp(\pi c L/12\beta) \,. \tag{14.8.11}$$

We have effectively set $\tilde{c} = 0$ because only the left-movers are excited in the supersymmetric states. The earlier result was for a string of length 2π, so we have replaced $H \to LH/2\pi$ by dimensional analysis. The density of states is related to this by

$$\int_0^\infty dE\, n(E) \exp(-\beta E) = \mathrm{Tr}[\exp(-\beta H)] \,, \tag{14.8.12}$$

giving in saddle point approximation

$$n(E) \approx \exp\left[(\pi c E L/3)^{1/2}\right] . \tag{14.8.13}$$

Finally, the central charge for our system is $c = 6Q_1Q_5$, while $E = 2\pi n_5/L$, and so

$$n(E) \approx \exp\left[2\pi(Q_1Q_5n_5)^{1/2}\right] , \tag{14.8.14}$$

in precise agreement with the exponential of the black hole entropy. This is a remarkable result, and another indication, beyond perturbative finiteness, that string theory defines a sensible theory of quantum gravity.

This result has been extended to other supersymmetric black holes, to the entropy of almost supersymmetric black holes, and to decay and absorption rates of almost supersymmetric black holes. In these cases the agreement is somewhat surprising, not obviously a consequence of supersymmetry. Subsequently the 'string' picture of the black holes has been extended to circumstances such as M-theory where there is no D-brane interpretation. These results are suggestive but the interpretation is not clear. We will discuss highly nonsupersymmetric black holes below.

Recalling the idea that D-branes can probe distances below the string scale, one might wonder whether the black p-brane metrics (14.8.1) and (14.8.3) can be seen even in the regime $gQ < 1$ in which the D-brane

picture rather than low energy field theory is relevant. Indeed, in some cases they can; this is developed in exercise 14.7.

The metric simplifies very close to $r = 0$, where the terms 1 in Z_1, Z_5, and Z_n become negligible. Taking for simplicity the case $r_n = 0$, the metric becomes

$$ds^2 = \frac{r^2}{r_1 r_5}\eta_{\mu\nu}dx^\mu dx^\nu + \frac{r_1 r_5}{r^2}dr^2 + r_1 r_5 d\Omega^2 + \frac{r_1}{r_5}dx^m dx^m \,. \tag{14.8.15}$$

This is a product space

$$AdS_3 \times S^3 \times T^4 \,. \tag{14.8.16}$$

Here AdS_3 is three-dimensional anti-de-Sitter space, which is the geometry in the coordinates x^μ and r (to be precise, these coordinates cover only part of anti-de-Sitter space). In a similar way, the metric (14.8.1) near a black 3-brane is

$$AdS_5 \times S^5 \,. \tag{14.8.17}$$

The case $p = 3$ is special because the dilaton remains finite at the horizon $r = 0$, as it does for the D1–D5 metric (14.8.3).

Very recently, a very powerful new duality proposal has emerged. Consider the IR dynamics of a system of N coincident Dp-branes. The bulk closed strings should decouple from the dynamics on the branes because gravity is an irrelevant interaction. The brane dynamics will then be described by the supersymmetric Yang–Mills theory on the brane, even for gN large. On the other hand, when gN is large the description of the system in terms of low energy supergravity should be valid as we have discussed. Thus we have two different descriptions which appear to have an overlapping range of validity. In the Yang–Mills description the effective expansion parameter gN is large, so perturbation theory is not valid. However, for g fixed, N is also very large. Noting that the gauge group on the branes is $U(N)$, this is the limit of a large number of 'colors,' the large-N limit. Field theories simplify in this limit, but it has been a long-standing unsolved problem to obtain any analytic understanding of Yang–Mills theories in this way. Now it appears, at least for theories with enough supersymmetry, that one can calculate amplitudes in the gauge theory by using the dual picture, where at low energy supergravity is essentially classical. If this idea is correct, it is a tremendous advance in the understanding of gauge field theories.

A correspondence principle

To make a precise entropy calculation we had to consider an extremal black hole with a specific set of charges. What of the familiar neutral

Schwarzschild black hole? Here too one can make a quantitative statement, but not at the level of precision of the supersymmetric case.

For a four-dimensional Schwarzschild black hole of mass M, the radius and entropy are

$$R \approx G_N M \,, \tag{14.8.18a}$$

$$S_{bh} \approx \frac{R^2}{G_N} \approx G_N M^2 \,. \tag{14.8.18b}$$

In this section we will systematically ignore numerical constants like 2 and π, for a reason to be explained below; hence the $\approx$. Let us consider what happens as we adiabatically change the dimensionless string coupling g. In four dimensions, dimensional analysis gives

$$G_N \approx g^2 \alpha' \,. \tag{14.8.19}$$

As we vary g the dimensionless combination $G_N M^2$ stays fixed. The simplest way to see this is to appeal to the fact that the black hole entropy (14.8.18) has the same properties as the thermodynamic entropy, and so is invariant under adiabatic changes.

Now imagine making the coupling very weak. One might imagine that for sufficiently weak coupling the black hole will no longer be black. One can see where this should happen from the following argument. The preceding two equations imply that

$$\frac{R}{\alpha'^{1/2}} \approx g S_{bh}^{1/2} \,. \tag{14.8.20}$$

We are imagining that S_{bh} is large so that the thermodynamic picture is good. Until g is very small, the Schwarzschild radius is then large compared to the string length and the gravitational dynamics should not be affected by stringy physics.

However, when g becomes small enough that $gS_{bh}^{1/2}$ is of order 1, stringy corrections to the action become important. If we try to extrapolate past this point, the black hole becomes smaller than a string! It is then unlikely that the field theory description of the black hole continues to be valid. Rather, the system should look like a state in weakly coupled string theory. This is how we can make the comparison: at this point, if the black hole entropy has a statistical interpretation, then the weakly coupled string theory should have the appropriate number of states of the given mass to account for this entropy. Since the entropy is assumed to be large we are interested in highly excited states. For a *single* highly excited string of mass M the density of states can be found as in section 9.8 and exercise 11.12,

$$\exp\left\{\pi M[(c+\tilde{c})\alpha'/3]^{1/2}\right\} \,. \tag{14.8.21}$$

In fact, one can show that with this exponential growth in their number,

the single string states are a significant fraction of the total number of states of given energy. In particular, and in contrast to the R–R case, states with D-branes plus anti-D-branes would have a much lower entropy because of the energy locked in the D-brane rest mass. The entropy of weakly coupled string states is then the logarithm

$$S_s \approx M\alpha'^{1/2} \approx g^{-1} M G_N^{1/2} \,. \tag{14.8.22}$$

This entropy has a different parametric dependence than the black hole entropy (14.8.18). However, they are to be compared only at the point

$$g S_{bh}^{1/2} \approx 1 \,, \tag{14.8.23}$$

where the transition from one picture to the other occurs. Inserting this value for g, the string entropy (14.8.22) becomes $S_{bh}^{1/2} M G_N^{1/2} = S_{bh}$.

We see that the numerical coefficients cannot be determined in this approach, since we do not know the exact coupling where the transition occurs, and corrections are in any case becoming significant on each side. However, *a priori* the entropy could have failed to match by a power of the large dimensionless number in the problem, S_{bh}. One can show that the same agreement holds in any dimension (exercise 14.8) and for black holes with a variety of charges. This is further evidence for the statistical interpretation of the black hole entropy, and that string theory has the appropriate number of states to be a complete theory of quantum gravity.

The information paradox

A closely associated issue is the black hole information paradox. A black hole of given mass and charge can be formed in a very large number of ways. It will then evaporate, and the Hawking radiation is apparently independent of what went into the black hole. This is inconsistent with ordinary quantum mechanics, as it requires pure states to evolve into mixed states.

There are various schools of thought here. The proposal of Hawking is that this is just the way things are: the laws of quantum mechanics need to be changed. There is also strong skepticism about this view, partly because this modification of quantum mechanics is rather ugly and very possibly inconsistent. However, 20 years of investigation have only served to sharpen the paradox. The principal alternative, that the initial state is encoded in subtle correlations in the Hawking radiation, sounds plausible but in fact is even more radical.[5] The problem is that Hawking radiation

[5] A third major alternative is that the evaporation ends in a remnant, a Planck-mass object having an enormous number of internal states. This might be stable or might release its information over an exceedingly long time scale. This has its own problems of aesthetics and possibly consistency, and is generally regarded as less likely.

emerges from the region of the horizon, where the geometry is smooth and so ordinary low energy field theory should be valid. One can follow the Hawking radiation and see correlations develop between the fields inside and outside the black hole; the superposition principle then forbids the necessary correlations to exist strictly among the fields outside. To evade this requires that the locality principle in quantum field theory break down in some long-ranged but subtle way.

The recent progress in string duality suggests that black holes do obey the ordinary rules of quantum mechanics. The multiplets include black holes along with various nonsingular states, and we have continuously deformed a black hole into a system that obeys ordinary quantum mechanics. However, this is certainly not conclusive — we have two descriptions with different ranges of validity, and while the D-brane system has an explicit quantum mechanical description, one could imagine that as the coupling constant is increased a critical coupling is reached where the D-particles collapse to form a black hole. At this coupling there could be a discontinuous change (or a smooth crossover) from ordinary quantum behavior to information loss.

Certainly if matrix theory is correct, the ordinary laws of quantum mechanics are preserved and the information must escape (there are not enough states for the remnant idea). It should be noted that in matrix theory only locality in time is explicit, so the necessary nonlocality may be present. If so, it is important to see in detail how this happens. In particular it may give insight into the cosmological constant problem, which stands in the way of our understanding the vacuum and supersymmetry breaking. This is another place where the continued failure of mundane ideas suggests that we need something new and perhaps nonlocal.

Exercises

14.1 From the supersymmetry algebra (13.2.9), show that an infinite type II F-string with excitations moving in only one direction is a BPS state. Show the same for a D-string.

14.2 Using the multi-NS5-brane solution (14.1.15), (14.1.17) and the D-string action, calculate the mass of a D-string stretched between two NS5-branes. Using IIB S-duality, compare this with the mass of an F-string stretched between D5-branes.

14.3 For compactification of the type II string on T^4, where the U-duality group is $SO(5,5,\mathbf{Z})$ and the T-duality group is $SO(4,4,\mathbf{Z})$, repeat the discussion in section 14.2 of the representations carried by the vector fields.

14.4 (a) For the series of operations TST discussed in section 14.4, deduce the transformation of each gauge field and higher rank form.
(b) Deduce the transformation of each extended object (D-brane, F-string, or NS-brane, with the various possible orientations for each).
(c) In each case compare with the interpretation as a 90° rotation of M-theory compactified on T^2.

14.5 As discussed in section 14.6, consider the type I theory compactified on T^2. In terms of the two radii, the string coupling, and the Wilson line, determine the six limits of parameter space that give the six noncompact theories at the cusps of figure 14.4, with the coupling going to zero in the stringy limits.

14.6 Expand the black p-brane solution (14.8.1) to first order in gQ and compare with the field produced by a Dp-brane, calculated in the linearized low energy field theory as in section 8.7.

14.7 Consider a D1-brane aligned along the 1-direction. Evaluate the D1-brane action in the field (14.8.3) and expand to order v^2. For $n_5 = 0$, compare with the order v^2 interaction between a D1-brane and a collection of D1- and D5-branes as obtained from the annulus.

14.8 Extend the correspondence principle to Schwarzschild black holes in other dimensions. The necessary black hole properties can be obtained by dimensional analysis. The entropy is always equal to the horizon area (with units l^{d-2}) times G_N.

15
Advanced CFT

We have encountered a number of infinite-dimensional symmetry algebras on the world-sheet: conformal, superconformal, and current. While we have used these symmetries as needed to obtain specific physical results, in the present chapter we would like to take maximum advantage of them in determining the form of the world-sheet theory. An obvious goal, not yet reached, would be to construct the general conformal or superconformal field theory, corresponding to the general classical string background.

This subject is no longer as central as it once appeared to be, as spacetime rather than world-sheet symmetries have been the principal tools in recent times. However, it is a subject of some beauty in its own right, with various applications to string compactification and also to other areas of physics.

We first discuss the representations of the conformal algebra, and the constraints imposed by conformal invariance on correlation functions. We then study some examples, such as the minimal models, Sugawara and coset theories, where the symmetries do in fact determine the theory completely. We briefly summarize the representation theory of the $N = 1$ superconformal algebra. We then discuss a framework, rational conformal field theory, which incorporates all these CFTs. To conclude this chapter we present some important results about the relation between conformal field theories and nearby two-dimensional field theories that are *not* conformally invariant, and the application of CFT in statistical mechanics.

15.1 Representations of the Virasoro algebra

In section 3.7 we discussed the connection between classical string backgrounds and general CFTs. In particular, we observed that CFTs corresponding to compactification of the spatial dimensions are unitary and

their spectra are discrete and bounded below. These additional conditions strongly restrict the world-sheet theory, and we will assume them throughout this chapter except for occasional asides.

Because the spectrum is bounded below, acting repeatedly with Virasoro lowering operators always produces a highest weight (primary) state $|h\rangle$, with properties

$$L_0|h\rangle = h|h\rangle \,, \tag{15.1.1a}$$

$$L_m|h\rangle = 0, \quad m > 0 \,. \tag{15.1.1b}$$

Starting from a highest weight state, we can form a representation of the Virasoro algebra

$$[L_m, L_n] = (m-n)L_{m+n} + \frac{c}{12}(m^3 - m)\delta_{m,-n} \tag{15.1.2}$$

by taking $|h\rangle$ together with all the states obtained by acting on $|h\rangle$ with the Virasoro raising operators,

$$L_{-k_1} L_{-k_2} \ldots L_{-k_l}|h\rangle \,. \tag{15.1.3}$$

We will denote this state $|h, \{k\}\rangle$ or $L_{-\{k\}}|h\rangle$ for short. The state (15.1.3) is known as a *descendant* of $|h\rangle$, or a *secondary*. A primary together with all of its descendants is also known as a *conformal family*. The integers $\{k\}$ may be put in the standard order $k_1 \geq k_2 \geq \ldots \geq k_l \geq 1$ by commuting the generators. This process terminates in a finite number of steps, because each nonzero commutator reduces the number of generators by one. To see that this is a representation, consider acting on $|h, \{k\}\rangle$ with any Virasoro generator L_n. For $n < 0$, commute L_n into its standard order; for $n \geq 0$, commute it to the right until it annihilates $|h\rangle$. In either case, the nonzero commutators are again of the form $|h, \{k'\}\rangle$. All coefficients are determined entirely in terms of the central charge c from the algebra and the weight h obtained when L_0 acts on $|h\rangle$; these two parameters completely define the highest weight representation.

It is a useful fact that for unitary CFTs all states lie in highest weight representations — not only can we always get from any state to a primary with lowering operators, but we can always get back again with raising operators. Suppose there were a state $|\phi\rangle$ that could not be expanded in terms of primaries and secondaries. Consider the lowest-dimension state with this property. By taking

$$|\phi\rangle \to |\phi\rangle - |i\rangle\langle i|\phi\rangle \tag{15.1.4}$$

with $|i\rangle$ running over a complete orthonormal set of primaries and secondaries, we may assume $|\phi\rangle$ to be orthogonal to all primaries and secondaries. Now, $|\phi\rangle$ is not primary, so there is a nonzero state $L_n|\phi\rangle$ for

some $n > 0$. Since the CFT is unitary this has a positive norm,

$$\langle\phi|L_{-n}L_n|\phi\rangle > 0 \,. \tag{15.1.5}$$

The state $L_n|\phi\rangle$ lies in a highest weight representation, since by assumption $|\phi\rangle$ is the lowest state that does not, and so therefore does $L_{-n}L_n|\phi\rangle$. Therefore it must be orthogonal to $|\phi\rangle$, in contradiction to eq. (15.1.5).

This need not hold in more general circumstances. Consider the operator ∂X of the linear dilaton theory. Lowering this gives the unit operator, $L_1 \cdot \partial X = -\alpha' V \cdot 1$, but $L_{-1} \cdot 1 = 0$ so we cannot raise this operator back to ∂X. The problem is the noncompactness of X combined with the position dependence of the dilaton, so that even $|1\rangle$ is not normalizable.

Now we would like to know what values of c and h are allowed in a unitary theory. The basic method was employed in section 2.9, using the Virasoro algebra to compute the inner product

$$\mathcal{M}^1 \equiv \langle h|L_1L_{-1}|h\rangle = 2h \,, \tag{15.1.6}$$

implying $h \geq 0$. Consideration of another inner product gave $c \geq 0$. Now look more systematically, level by level. At the second level of the highest weight representation, the two states $L_{-1}L_{-1}|h\rangle$ and $L_{-2}|h\rangle$ have the matrix of inner products

$$\mathcal{M}^2 = \begin{bmatrix} \langle h|L_1^2 \\ \langle h|L_2 \end{bmatrix} \begin{bmatrix} L_{-1}^2|h\rangle & L_{-2}|h\rangle \end{bmatrix} \,. \tag{15.1.7}$$

Commuting the lowering operators to the right gives

$$\mathcal{M}^2 = \begin{pmatrix} 8h^2+4h & 6h \\ 6h & 4h+c/2 \end{pmatrix} \tag{15.1.8}$$

and

$$\det(\mathcal{M}^2) = 32h(h-r_+)(h-r_-) \,, \tag{15.1.9a}$$

$$16r_\pm = (5-c) \pm [(1-c)(25-c)]^{1/2} \,. \tag{15.1.9b}$$

In a unitary theory the matrix of inner products, and in particular its determinant, cannot be negative. The determinant is nonnegative in the region $c \geq 1$, $h \geq 0$, but for $0 < c < 1$ a new region $h_- < h < h_+$ is excluded.

At level N, the matrix of inner products is

$$\mathcal{M}^N_{\{k\},\{k'\}}(c,h) = \langle h, \{k\}|h, \{k'\}\rangle \,, \quad \sum_i k_i = N \,. \tag{15.1.10}$$

Its determinant has been found,

$$\det[\mathcal{M}^N(c,h)] = K_N \prod_{1\leq rs\leq N} (h-h_{r,s})^{P(N-rs)} \tag{15.1.11}$$

with K_N a positive constant. This is the *Kac determinant*. The zeros of the determinant are at

$$h_{r,s} = \frac{c-1}{24} + \frac{1}{4}(r\alpha_+ + s\alpha_-)^2 \,, \tag{15.1.12}$$

where

$$\alpha_\pm = (24)^{-1/2}\left[(1-c)^{1/2} \pm (25-c)^{1/2}\right] \,. \tag{15.1.13}$$

The multiplicity $P(N-rs)$ of each root is the partition of $N-rs$, the number of ways that $N-rs$ can be written as a sum of positive integers (with $P(0) \equiv 1$):

$$\prod_{n=1}^{\infty} \frac{1}{1-q^n} = \sum_{k=0}^{\infty} P(k)q^k \,. \tag{15.1.14}$$

At level 2, for example, the roots are $h_{1,1} = 0$, $h_{2,1} = h_+$, and $h_{1,2} = h_-$, each with multiplicity 1, as found above.

The calculation of the determinant (15.1.11) is too lengthy to repeat here. The basic strategy is to construct all of the null states, those corresponding to the zeros of the determinant, either by direct combinatoric means or using some tricks from CFT. The determinant is a polynomial in h and so is completely determined by its zeros, up to a normalization which can be obtained by looking at the $h \to \infty$ limit. The order of the polynomial is readily determined from the Virasoro algebra, so one can know when one has all the null states. Let us note one particular feature. At level 2, the null state corresponding to $h_{1,1}$ is $L_{-1}L_{-1}|h=0\rangle$. This is a descendant of the level 1 null state $L_{-1}|h=0\rangle$. In general, the zero $h_{r,s}$ appears first at level rs. At every higher level N are further null states obtained by acting with raising operators on the level rs state; the partition $P(N-rs)$ in the Kac determinant is the total number of ways to act with raising operators of total level $N-rs$.

A careful study of the determinant and its functional dependence on c and h shows (the analysis is again too lengthy to repeat here) that unitary representations are allowed only in the region $c \geq 1$, $h \geq 0$ and at a discrete set of points (c,h) in the region $0 \leq c < 1$:

$$c = 1 - \frac{6}{m(m+1)} \,, \quad m = 2,3,\dots \,,$$

$$= 0, \frac{1}{2}, \frac{7}{10}, \frac{4}{5}, \frac{6}{7}, \cdots \,, \tag{15.1.15a}$$

$$h = h_{r,s} = \frac{[r(m+1)-sm]^2 - 1}{4m(m+1)} \,, \tag{15.1.15b}$$

where $1 \leq r \leq m-1$ and $1 \leq s \leq m$. The discrete representations are of great interest, and we will return to them in section 15.3.

For a unitary representation, the Kac determinant also determines whether the states $|h,\{k\}\rangle$ are linearly independent. If it is positive, they are; if it vanishes, some linear combination(s) are orthogonal to all states and so by unitarity must vanish. The representation is then said to be *degenerate*. Of the unitary representations, all the discrete representations (15.1.15) are degenerate, as are the representations $c = 1$, $h = \frac{1}{4}n^2$, $n \in \mathbf{Z}$ and $c > 1$, $h = 0$. For example, at $h = 0$ we always have $L_{-1} \cdot 1 = 0$, but at the next level $L_{-2} \cdot 1 = T_{zz}$ is nonzero.

Let us make a few remarks about the nonunitary case. In the full matter CFT of string theory, the states $L_{-\{k\}}|h\rangle$ obtained from any primary state $|h\rangle$ *are* linearly independent when the momentum is nonzero. This can be seen by using the same light-cone decomposition used in the no-ghost proof of chapter 4. The term in L_{-n} of greatest N^{lc} is $k^- \alpha^+_{-n}$. These manifestly generate independent states; the upper triangular structure then guarantees that this independence holds also for the full Virasoro generators.

A representation of the Virasoro algebra with all of the $L_{-\{k\}}|h\rangle$ linearly independent is known as a *Verma module*. Verma modules exist at all values of c and h. Verma modules are particularly interesting when the dimension h takes a value $h_{r,s}$ such that the Kac determinant vanishes. The module then contains nonvanishing null states (states that are orthogonal to all states in the module). Acting on a null state with a Virasoro generator gives a null state again, since for any null $|v\rangle$ and for any state $|\psi\rangle$ in the module we have $\langle\psi|(L_n|v\rangle) = (\langle\psi|L_n)|v\rangle = 0$. The representation is thus reducible: the subspace of null states is left invariant by the Virasoro algebra.[1] The L_n for $n > 0$ must therefore annihilate the *lowest* null state, so this state is in fact primary, in addition to being a level rs descendant of the original primary state $|h_{r,s}\rangle$. That is, the $h_{r,s}$ Verma module contains an $h = h_{r,s} + rs$ Verma submodule. In some cases, including the special discrete values of c (15.1.15), there is an intricate pattern of nested submodules.

Clearly a Verma module can be unitary only at those values of c and h where nondegenerate unitary representations are allowed. At the (c,h) values with degenerate unitary representations, the unitary representation is obtained from the corresponding Verma module by modding out the null states.

As a final example consider the matter sector of string theory, $c = 26$. From the OCQ, we know that there are many null physical states at $h = 1$. This can be seen from the Kac formula as well. For $c = 26$, $\alpha_+ = 3i/6^{1/2}$, $\alpha_- = 2i/6^{1/2}$, and so

$$h_{r,s} = \frac{25 - (3r + 2s)^2}{24} . \tag{15.1.16}$$

The corresponding null physical state is at

$$h = h_{r,s} + rs = \frac{25 - (3r - 2s)^2}{24} . \tag{15.1.17}$$

Any pair of positive integers with $|3r - 2s| = 1$ leads to a null physical state at $h = 1$. For example, the states $(r,s) = (1,1)$ and $(1,2)$ were constructed in exercise 4.2. With care, one can show that the number of null states implied by the Kac formula is exactly that required by the no-ghost theorem.

[1] By contrast, a unitary representation is always irreducible. The reader can show that the lowest state in any invariant subspace would have to be orthogonal to itself, and therefore vanish.

15.2 The conformal bootstrap

We now study the constraints imposed by conformal invariance on correlation functions on the sphere. In chapter 6 we saw that the Möbius subgroup, with three complex parameters, reduced the n-point function to a function of $n-3$ complex variables. The rest of the conformal symmetry gives further information: it determines all the correlation functions of descendant fields in terms of those of the primary fields.

To begin, consider the correlation function of the energy momentum tensor $T(z)$ with n primary fields $\mathcal{O}$. The singularities of the correlation function as a function of z are known from the $T\mathcal{O}$ OPE. In addition, it must fall as z^{-4} for $z \to \infty$, since in the coordinate patch $u = 1/z$, $T_{uu} = z^4 T_{zz}$ is holomorphic at $u = 0$. This determines the correlation function to be

$$\langle T(z)\mathcal{O}_1(z_1)\dots\mathcal{O}_n(z_n)\rangle_{S_2} = \sum_{i=1}^{n}\left[\frac{h_i}{(z-z_i)^2} + \frac{1}{(z-z_i)}\frac{\partial}{\partial z_i}\right]\langle \mathcal{O}_1(z_1)\dots\mathcal{O}_n(z_n)\rangle_{S_2} . \tag{15.2.1}$$

A possible holomorphic addition is forbidden by the boundary condition at infinity. In addition, the asymptotics of order z^{-1}, z^{-2}, and z^{-3} must vanish; these are the same as the conditions from Möbius invariance, developed in section 6.7. The correlation function with several Ts is of the same form, with additional singularities from the TT OPE. Now make a Laurent expansion in $z - z_1$,

$$T(z)\mathcal{O}_1(z_1) = \sum_{k=-\infty}^{\infty} (z - z_1)^{k-2} L_{-k} \cdot \mathcal{O}_1(z_1) . \tag{15.2.2}$$

Then for $k \geq 1$, matching coefficients of $(z - z_1)^{k-2}$ on the right and left of the correlator (expectation value) (15.2.1) gives

$$\langle\, [L_{-k} \cdot \mathcal{O}_1(z_1)]\, \mathcal{O}_2(z_2) \ldots \mathcal{O}_n(z_n)\, \rangle_{S_2} = \mathscr{L}_{-k} \langle\, \mathcal{O}_1(z_1) \ldots \mathcal{O}_n(z_n)\, \rangle_{S_2}\ , \tag{15.2.3}$$

where

$$\mathscr{L}_{-k} = \sum_{i=2}^{n} \left[\frac{h_i(k-1)}{(z_i - z_1)^k} - \frac{1}{(z_i - z_1)^{k-1}} \frac{\partial}{\partial z_i} \right] . \tag{15.2.4}$$

This extends to multiple generators, and to the antiholomorphic side,

$$\left\langle\, [L_{-k_1} \ldots L_{-k_\ell} \tilde{L}_{-l_1} \ldots \tilde{L}_{-l_m} \cdot \mathcal{O}_1(z_1)] \ldots \mathcal{O}_n(z_n)\, \right\rangle_{S_2} = \mathscr{L}_{-k_\ell} \ldots \mathscr{L}_{-k_1} \tilde{\mathscr{L}}_{-l_m} \ldots \tilde{\mathscr{L}}_{-l_1} \langle\, \mathcal{O}_1(z_1) \ldots \mathcal{O}_n(z_n)\, \rangle_{S_2}\ . \tag{15.2.5}$$

The additional terms from the TT OPE do not contribute when all the k_i and l_i are positive. The correlator of one descendant and $n-1$ primaries is thus expressed in terms of that of n primaries. Clearly this can be extended to n descendants, though the result is more complicated because there are additional terms from the TT singularities.

Earlier we argued that the operator product coefficients were the basic data in CFT, determining all the other correlations via factorization. We see now that it is only the operator product coefficients of primaries that are necessary. It is worth developing this somewhat further for the four-point correlation. Start with the operator product of two primaries, with the sum over operators now broken up into a sum over conformal families i and a sum within each family,

$$\mathcal{O}_m(z,\bar{z})\mathcal{O}_n(0,0) = \sum_{i,\{k,\tilde{k}\}} z^{-h_m-h_n+h_i+N} \bar{z}^{-\tilde{h}_m-\tilde{h}_n+\tilde{h}_i+\tilde{N}} \times c^{i\{k,\tilde{k}\}}{}_{mn} L_{-\{k\}} \tilde{L}_{-\{\tilde{k}\}} \cdot \mathcal{O}_i(0,0)\ , \tag{15.2.6}$$

where N is the total level of $\{k\}$. Writing the operator product coefficient $c^{i\{k,\tilde{k}\}}{}_{mn}$ as a three-point correlator and using the result (15.2.5) to relate this to the correlator of three primaries gives

$$c^{i\{k,\tilde{k}\}}{}_{mn} = \sum_{\{k',\tilde{k}'\}} \mathscr{M}^{-1}_{\{k\},\{k'\}} \mathscr{M}^{-1}_{\{\tilde{k}\},\{\tilde{k}'\}} \times \mathscr{L}_{-\{k'\}} \tilde{\mathscr{L}}_{-\{\tilde{k}'\}} \langle\, \mathcal{O}_m(\infty,\infty)\mathcal{O}_n(1,1)\mathcal{O}_i(z_1,\bar{z}_1)\, \rangle_{S_2}\Big|_{z_1=0}\ . \tag{15.2.7}$$

To relate the operator product coefficient to a correlator we have to raise an index, so the inverse $\mathscr{M}^{-1}$ appears (with an appropriate adjustment for degenerate representations). The right-hand side is equal to the operator product of the primaries times a function of the coordinates and their

derivatives, the latter being completely determined by the conformal invariance. Carrying out the differentiations in $\mathscr{L}_{-\{k'\}}$ and $\tilde{\mathscr{L}}_{-\{\tilde{k}'\}}$ and then summing leaves

$$c^{i\{k,\tilde{k}\}}{}_{mn} = \beta^{i\{k\}}_{mn} \tilde{\beta}^{i\{\tilde{k}\}}_{mn} c^{i}_{mn}\ . \tag{15.2.8}$$

The coefficient $\beta^{i\{k\}}_{mn}$ is a function of the weights h_m, h_n, and h_i and the central charge c, but is otherwise independent of the CFT.

Now use the OPE (15.2.6) to relate the four-point correlation to the product of three-point correlations,

$$\langle\, \mathcal{O}_j(\infty,\infty)\mathcal{O}_l(1,1)\mathcal{O}_m(z,\bar{z})\mathcal{O}_n(0,0)\, \rangle_{S_2} = \sum_i c^{i}_{jl} c_{imn} \mathscr{F}^{jl}_{mn}(i|z) \tilde{\mathscr{F}}^{jl}_{mn}(i|\bar{z})\ , \tag{15.2.9}$$

where

$$\mathscr{F}^{jl}_{mn}(i|z) = \sum_{\{k\},\{k'\}} z^{-h_m-h_n+h_i+N} \beta^{i\{k\}}_{jl} \mathscr{M}_{\{k\},\{k'\}} \beta^{i\{k'\}}_{mn}\ . \tag{15.2.10}$$

This function is known as the conformal block, and is holomorphic except at $z = 0$, 1, and ∞. The steps leading to the decomposition (15.2.9) show that the conformal block is determined by the conformal invariance as a function of h_j, h_l, h_m, h_n, h_i, c, and z. One can calculate it order by order in z by working through the definition.

Recall that the single condition for a set of operator product coefficients to define a consistent CFT on the sphere is duality of the four-point function, the equality of the decompositions (15.2.9) in the $(jl)(mn)$, $(jm)(ln)$, and $(jn)(lm)$ channels. The program of solving this constraint is known as the *conformal bootstrap*. The general solution is not known. One limitation is that the conformal blocks are not known in closed form except for special values of c and h.

Beyond the sphere, there are the additional constraints of modular invariance of the zero-point and one-point functions on the torus. Here we will discuss only a few of the most general consequences. Separating the sum over states in the partition function into a sum over conformal families and a sum within each family yields

$$Z(\tau) = \sum_{i,\{k,\tilde{k}\}} q^{-c/24+h_i+N} \bar{q}^{-\tilde{c}/24+\tilde{h}_i+\tilde{N}} = \sum_i \chi_{c,h_i}(q) \chi_{\tilde{c},\tilde{h}_i}(\bar{q})\ . \tag{15.2.11}$$

Here

$$\chi_{c,h}(q) = q^{-c/24+h} \sum_{\{k\}} q^N \tag{15.2.12}$$

is the *character* of the (c,h) representation of the Virasoro algebra. For a Verma module the states generated by the L_{-k} are in one-to-one correspondence with the excitations of a free boson, generated by α_{-k}. Thus,

$$\chi_{c,h}(q) = q^{-c/24+h} \prod_{n=1}^{\infty} \frac{1}{1-q^n} \tag{15.2.13}$$

for a nondegenerate representation. For degenerate representations it is necessary to correct this expression for overcounting. A generic degenerate representation would have only one null primary, say at level N; the representation obtained by modding out the resulting null Verma module would then have character $(1-q^N)q^{1/24}\eta(q)^{-1}$. For the unitary degenerate representations (15.1.15), with their nested submodules, the calculation of the character is more complicated.

In section 7.2 we found the asymptotic behavior of the partition function for a general CFT,

$$Z(i\ell) \overset{\ell\to 0}{\sim} \exp(\pi c/6\ell)\,, \tag{15.2.14}$$

letting $c = \tilde{c}$. For a single conformal family, letting $q = \exp(-2\pi\ell)$,

$$\chi_{c,h}(q) \leq q^h \eta(i\ell)^{-1} \overset{\ell\to 0}{\sim} \ell^{1/2} \exp(\pi/12\ell)\,. \tag{15.2.15}$$

Then for a general CFT

$$Z(i\ell) \leq \mathcal{N}\ell \exp(\pi/6\ell) \tag{15.2.16}$$

as $\ell \to 0$, with $\mathcal{N}$ the total number of primary fields in the sum (15.2.11). Comparing this bound with the known asymptotic behavior (15.2.14), $\mathcal{N}$ can be finite only if $c < 1$. So, while we have been able to use conformal invariance to reduce sums over states to sums over primaries only, this remains an infinite sum whenever $c \geq 1$. The $c < 1$ theories, to be considered in the next section, stand out as particularly simple.

15.3 Minimal models

For fields in degenerate representations, conformal invariance imposes additional strong constraints on the correlation functions. Throughout this section we take $c \leq 1$, because only in this range do degenerate representations of positive h exist. We will not initially assume the CFT to be unitary, but the special unitary values of c will eventually appear.

Consider, as an example, a primary field $\mathcal{O}_{1,2}$ with weight

$$h = h_{1,2} = \frac{c-1}{24} + \frac{(\alpha_+ + 2\alpha_-)^2}{4}\,. \tag{15.3.1}$$

For now we leave the right-moving weight $\tilde{h}$ unspecified. The vanishing descendant is

$$\mathcal{N}_{1,2} = \left[L_{-2} - \frac{3}{2(2h_{1,2}+1)} L_{-1}^2\right] \cdot \mathcal{O}_{1,2} = 0\,. \tag{15.3.2}$$

Inserting this into a correlation with other primary fields and using the relation (15.2.5) expressing correlations of descendants in terms of those of primaries gives a partial differential equation for the correlations of the degenerate primary,

$$\begin{aligned} 0 &= \left\langle \mathcal{N}_{1,2}(z_1) \prod_{i=2}^{n} \mathcal{O}_i(z_i) \right\rangle_{S_2} \\ &= \left[\mathcal{L}_{-2} - \frac{3}{2(2h_{1,2}+1)} \mathcal{L}_{-1}^2\right] \mathcal{A}_n \\ &= \left[\sum_{i=2}^{n} \frac{h_i}{(z_i - z_1)^2} - \sum_{i=2}^{n} \frac{1}{z_i - z_1} \frac{\partial}{\partial z_i} - \frac{3}{2(2h_{1,2}+1)} \frac{\partial^2}{\partial z_1^2}\right] \mathcal{A}_n\,, \end{aligned} \tag{15.3.3}$$

where

$$\mathcal{A}_n = \left\langle \mathcal{O}_{1,2}(z_1,\bar{z}_1) \prod_{i=2}^{n} \mathcal{O}_i(z_i,\bar{z}_i) \right\rangle_{S_2}\,. \tag{15.3.4}$$

For $n = 4$, the correlation is known from conformal invariance up to a function of a single complex variable, and eq. (15.3.3) becomes an ordinary differential equation. In particular, setting to zero the z^{-1}, z^{-2}, and z^{-3} terms in the $T(z)$ expectation value (15.2.1) allows one to solve for $\partial/\partial z_{2,3,4}$ in terms of $\partial/\partial z_1$, with the result that eq. (15.3.3) becomes

$$\begin{aligned} \left[\sum_{i=2}^{4} \frac{h_i}{(z_i - z_1)^2} - \sum_{2\leq i<j\leq 4} \frac{h_{1,2} - h_2 - h_3 - h_4 + 2(h_i + h_j)}{(z_i - z_1)(z_j - z_1)}\right. \\ \left. + \sum_{i=2}^{4} \frac{1}{z_i - z_1} \frac{\partial}{\partial z_1} - \frac{3}{2(2h_{1,2}+1)} \frac{\partial^2}{\partial z_1^2}\right] \mathcal{A}_4 = 0\,. \end{aligned} \tag{15.3.5}$$

This differential equation is of hypergeometric form. The hypergeometric functions, however, are holomorphic (except for branch cuts at coincident points), while $\mathcal{O}_{1,2}$ has an unknown $\bar{z}_1$ dependence. Now insert the expansion (15.2.9), in which the four-point correlation is written as a sum of terms, each a holomorphic conformal block times a conjugated block. The conformal blocks satisfy the same differential equation (15.3.5) and so are hypergeometric functions. Being second order, the differential equation (15.3.5) has two independent solutions, and each conformal block is a linear combination of these.

This procedure generalizes to any degenerate primary field. The primary $\mathcal{O}_{r,s}$ will satisfy a generalization of the hypergeometric equation. This

differential equation is of maximum order rs, coming from the term L_{-1}^{rs} in the null state $\mathcal{N}_{r,s}$. If the antiholomorphic weight is also degenerate, $\tilde{h} = h_{\tilde{r},\tilde{s}}$, the antiholomorphic conformal blocks satisfy a differential equation of order $\tilde{r}\tilde{s}$, so that

$$\left\langle \mathcal{O}_{r,s;\tilde{r},\tilde{s}}(z_1,\bar{z}_1)\prod_{i=2}^{4}\mathcal{O}_i(z_i,\bar{z}_i)\right\rangle_{S_2} = \sum_{i=1}^{rs}\sum_{j=1}^{\tilde{r}\tilde{s}} a_{ij} f_i(z)\tilde{f}_j(\bar{z}) \,, \tag{15.3.6}$$

where $f_i(z)$ and $\tilde{f}_j(\bar{z})$ are the general solutions of the holomorphic and antiholomorphic equations. The constants a_{ij} are not determined by the differential equation. They are constrained by locality — the holomorphic and antiholomorphic functions each have branch cuts, but the product must be single-valued — and by associativity. We will describe below some theories in which it has been possible to solve these conditions.

Let us see how the differential equation constrains the operator products of $\mathcal{O}_{1,2}$. According to the theory of ordinary differential equations, the points $z_1 = z_i$ are regular singular points, so that the solutions are of the form $(z_1 - z_i)^\kappa$ times a holomorphic function. Inserting this form into the differential equation and examining the most singular term yields the characteristic equation

$$\frac{3}{2(2h_{1,2}+1)}\kappa(\kappa-1) + \kappa - h_i = 0 \,. \tag{15.3.7}$$

This gives two solutions $(z_1 - z_i)^{\kappa_\pm}$ for the leading behavior as $z_1 \to z_i$; comparing this to the OPE gives

$$h_\pm = h_{1,2} + h_i + \kappa_\pm \tag{15.3.8}$$

for the primary fields in the $\mathcal{O}_{1,2}\mathcal{O}_i$ product. Parameterizing the weight by

$$h_i = \frac{c-1}{24} + \frac{\gamma^2}{4} \,, \tag{15.3.9}$$

the two solutions to the characteristic equation correspond to

$$h_\pm = \frac{c-1}{24} + \frac{(\gamma \pm \alpha_-)^2}{4} \,. \tag{15.3.10}$$

These are the only weights that can appear in the operator product, so we have derived the *fusion rule*,

$$\mathcal{O}_{1,2}\mathcal{O}_{(\gamma)} = [\mathcal{O}_{(\gamma+\alpha_-)}] + [\mathcal{O}_{(\gamma-\alpha_-)}] \;; \tag{15.3.11}$$

we have labeled the primary fields other than $\mathcal{O}_{1,2}$ by the corresponding parameter γ. A fusion rule is an OPE without the coefficients, a list of the conformal families that are allowed to appear in a given operator product (though it is possible that some will in fact have vanishing coefficient).

For the operator $\mathcal{O}_{2,1}$, one obtains in the same way the fusion rule

$$\mathcal{O}_{2,1}\mathcal{O}_{(\gamma)} = [\mathcal{O}_{(\gamma+\alpha_+)}] + [\mathcal{O}_{(\gamma-\alpha_+)}] \,. \tag{15.3.12}$$

In particular, for the product of two degenerate primaries this becomes

$$\mathcal{O}_{1,2}\mathcal{O}_{r,s} = [\mathcal{O}_{r,s+1}] + [\mathcal{O}_{r,s-1}] \,, \tag{15.3.13a}$$

$$\mathcal{O}_{2,1}\mathcal{O}_{r,s} = [\mathcal{O}_{r+1,s}] + [\mathcal{O}_{r-1,s}] \,. \tag{15.3.13b}$$

For positive values of the indices, the families on the right-hand side are degenerate. In fact, only these actually appear. Consider the fusion rule for $\mathcal{O}_{1,2}\mathcal{O}_{2,1}$. By applying the rule (15.3.13a) we conclude that only $[\mathcal{O}_{2,2}]$ and $[\mathcal{O}_{2,0}]$ may appear in the product, while the rule (15.3.13b) allows only $[\mathcal{O}_{2,2}]$ and $[\mathcal{O}_{0,2}]$. Together, these imply that only $[\mathcal{O}_{2,2}]$ can actually appear in the product. This generalizes: only primaries $r \geq 1$ and $s \geq 1$ are generated. The algebra of degenerate conformal families thus closes, and iterated products of $\mathcal{O}_{1,2}$ and $\mathcal{O}_{2,1}$ generate all degenerate $\mathcal{O}_{r,s}$. This suggests that we focus on CFTs in which *all* conformal theories are degenerate.

The values of r and s are still unbounded above, so that the operator algebra will generate an infinite set of conformal families. When $\alpha_-/\alpha_+ = -p/q$ is *rational*, the algebra closes on a finite set.[2] In particular, one then has

$$c = 1 - 6\frac{(p-q)^2}{pq} \,, \tag{15.3.14a}$$

$$h_{r,s} = \frac{(rq-sp)^2 - (p-q)^2}{4pq} \,. \tag{15.3.14b}$$

The point is that there is a reflection symmetry,

$$h_{p-r,q-s} = h_{r,s} \,, \tag{15.3.15}$$

so that each conformal family has at least *two* null vectors, at levels rs and $(p-r)(q-s)$, and its correlators satisfy *two* differential equations. The reflection of the conditions $r > 0$ and $s > 0$ is $r < p$ and $s < q$, so the operators are restricted to the range

$$1 \leq r \leq p-1 \,, \quad 1 \leq s \leq q-1 \,. \tag{15.3.16}$$

These theories, with a finite algebra of degenerate conformal families, are known as *minimal models*. They have been solved: the general solution of the locality, duality, and modular invariance conditions is known, and the operator product coefficients can be extracted though the details are too lengthy to present here.

[2] Note that $\alpha_+\alpha_- = -1$, and that $0 > \alpha_-/\alpha_+ > -1$.

Although the minimal models seem rather special, they have received a great deal of attention, as examples of nontrivial CFTs, as prototypes for more general solutions of the conformal bootstrap, as building blocks for four-dimensional string theories, and because they describe the critical behavior of many two-dimensional systems. We will return to several of these points later.

Let us now consider the question of unitarity. A necessary condition for unitarity is that all weights are nonnegative. One can show that this is true of the weights (15.3.14) only for $q = p + 1$. These are precisely the $c < 1$ representations (15.1.15) already singled out by unitarity:

$$p = m\,, \quad q = m + 1\,. \tag{15.3.17}$$

Notice that these theories have been found and solved purely from symmetry, without ever giving a Lagrangian description. This is how they were discovered, though various Lagrangian descriptions are now known; we will mention several later. For $m = 3$, c is $\frac{1}{2}$ and there is an obvious Lagrangian representation, the free fermion. The allowed primaries,

$$h_{1,1} = 0\,, \quad h_{2,1} = \frac{1}{2}\,, \quad h_{1,2} = \frac{1}{16}\,, \tag{15.3.18}$$

are already familiar, being respectively the unit operator, the fermion ψ, and the R sector ground state.

The full minimal model fusion rules can be derived using repeated applications of the $\mathcal{O}_{2,1}$ and $\mathcal{O}_{1,2}$ rules and associativity. They are

$$\mathcal{O}_{r_1,s_1}\mathcal{O}_{r_2,s_2} = \sum [\mathcal{O}_{r,s}]\,, \tag{15.3.19a}$$

$$r = |r_1 - r_2| + 1,\ |r_1 - r_2| + 3,\ \ldots,\ \min(r_1 + r_2 - 1, 2p - 1 - r_1 - r_2)\,, \tag{15.3.19b}$$

$$s = |s_1 - s_2| + 1,\ s_1 + s_2 + 3,\ \ldots,\ \min(s_1 + s_2 - 1, 2q - 1 - s_1 - s_2)\,. \tag{15.3.19c}$$

For $\mathcal{O}_{p-1,1}$ only a single term appears in the fusion with any other field, $\mathcal{O}_{p-1,1}\mathcal{O}_{r,s} = [\mathcal{O}_{p-r,s}]$. A primary with these properties is known as a *simple current*. Simple currents have the useful property that they have definite *monodromy* with respect to any other primary. Consider the operator product of a simple current $J(z)$ of weight h with any primary,

$$J(z)\mathcal{O}_i(0) = z^{h_{i'}-h_i-h}[\mathcal{O}_{i'}(0) + \text{descendants}]\,, \tag{15.3.20}$$

where $J \cdot [\mathcal{O}_i] = [\mathcal{O}_{i'}]$. The terms with descendants bring in only integer powers of z, so all terms on the right pick up a common phase

$$2\pi(h_{i'} - h_i - h) = 2\pi Q_i \tag{15.3.21}$$

when z encircles the origin. The charge Q_i, defined mod 1, is a discrete symmetry of the OPE. Using the associativity of the OPE, the operator

product coefficient c_{ij}^k can be nonzero only if $Q_i + Q_j = Q_k$. Also, by taking repeated operator products of J with itself one must eventually reach the unit operator; suppose this occurs first for J^N. Then associativity implies that NQ_i must be an integer, so this is a $\mathbf{Z}_N$ symmetry. For the minimal models,

$$\mathcal{O}_{p-1,1}\mathcal{O}_{p-1,1} = [\mathcal{O}_{1,1}] \tag{15.3.22}$$

which is the identity, and so the discrete symmetry is $\mathbf{Z}_2$. Evaluating the weights (15.3.21) gives

$$Q_{r,s} = \frac{p(1-s) + q(1-r)}{2} \mod 1\,. \tag{15.3.23}$$

For the unitary case (15.3.17), $\exp(2\pi i Q_{r,s})$ is $(-1)^{s-1}$ for m odd and $(-1)^{r-1}$ for m even.

Feigin–Fuchs representation

To close this section, we describe a clever use of CFT to generate integral representations of the solutions to the differential equations satisfied by the degenerate fields. Define

$$c = 1 - 24\alpha_0^2 \tag{15.3.24}$$

and consider the linear dilaton theory with the same value of the central charge,

$$T = -\frac{1}{2}\partial\phi\partial\phi + 2^{1/2}i\alpha_0\partial^2\phi\,. \tag{15.3.25}$$

The linear dilaton theory is not the same as a minimal model. In particular, the modes α_{-k} generate a Fock space of independent states, so the partition function is of order $\exp(\pi/6\ell)$ as $\ell \to 0$, larger than that of a minimal model. However, the correlators of the minimal model can be obtained from those of the linear dilaton theory. The vertex operator

$$V_\alpha = \exp(2^{1/2}i\alpha\phi) \tag{15.3.26}$$

has weight $\alpha^2 - 2\alpha\alpha_0$, so for

$$\alpha = \alpha_0 - \frac{\gamma}{2} \tag{15.3.27}$$

it is a primary of weight

$$\frac{c-1}{24} + \frac{\gamma^2}{4}\,. \tag{15.3.28}$$

For $\gamma = r\alpha_+ + s\alpha_-$ it is then degenerate, and its correlator satisfies the same differential equation as the corresponding minimal model primary.

There is a complication: the correlator

$$\langle V_{\alpha_1} V_{\alpha_2} V_{\alpha_3} V_{\alpha_4} \rangle \tag{15.3.29}$$

generally vanishes due to the conservation law

$$\sum_i \alpha_i = 2\alpha_0 \tag{15.3.30}$$

(derived in exercise 6.2). There is a trick which enables us to find a nonvanishing correlator that satisfies the same differential equation. The operators

$$J_\pm = \exp(2^{1/2} i\alpha_\pm \phi) \tag{15.3.31}$$

are of weight $(1,0)$, so the line integral

$$Q_\pm = \oint dz\, J_\pm \,, \tag{15.3.32}$$

known as a *screening charge*, is conformally invariant. Inserting $Q_+^{n_+} Q_-^{n_-}$ into the expectation value, the charge conservation condition is satisfied for

$$n_+ = \frac{1}{2}\sum_i r_i - 2 \,, \quad n_- = \frac{1}{2}\sum_i s_i - 2 \,. \tag{15.3.33}$$

Further, since the screening charges are conformally invariant, they do not introduce singularities into $T(z)$ and the derivation of the differential equation still holds. Thus, the minimal model conformal blocks are represented as contour integrals of the correlators of free-field exponentials, which are of course known. This is the *Feigin–Fuchs representation*. It is possible to replace $V_\alpha \to V_{2\alpha_0-\alpha}$ in some of the vertex operators, since this has the same weight; one still obtains integer values of $n_\pm$, but this may reduce the number of screening charges needed. It may seem curious that the charges of the $(1,0)$ vertex operators are just such as to allow for integer $n_\pm$. In fact, one can work backwards, deriving the Kac determinant from the linear dilaton theory with screening charges.

The contours in the screening operators have not been specified — they may be any nontrivial closed contours (but must end on the same Riemann sheet where they began, because there are branch cuts in the integrand), or they may begin and end on vertex operators if the integrand vanishes sufficiently rapidly at those points. By various choices of contour one generates all solutions to the differential equations, as in the theory of hypergeometric functions. As noted before, one must impose associativity and locality to determine the actual correlation functions. The Feigin–Fuchs representation has been a useful tool in solving these conditions.

15.4 Current algebras

We now consider a Virasoro algebra L_k combined with a current algebra j_k^a. We saw in section 11.5 that the Virasoro generators are actually constructed from the currents. We will extend that discussion to make fuller use of the world-sheet symmetry.

Recall that a primary state $|r,i\rangle$ in representation r of g satisfies

$$L_m|r,i\rangle = j_m^a|r,i\rangle \,, \quad m>0 \,, \tag{15.4.1a}$$

$$j_0^a|r,i\rangle = |r,j\rangle t_{r,ji}^a \,. \tag{15.4.1b}$$

As in the case of the Virasoro algebra, we are interested in highest weight representations, obtained by acting on a primary state with the L_m and j_m^a for $m<0$. As we have discussed, a CFT with a current algebra can always be factored into a Sugawara part and a part that commutes with the current algebra. We focus on the Sugawara part, where

$$T(z) = \frac{1}{(k+h(g))\psi^2} :jj(z): \,. \tag{15.4.2}$$

Recall also that the central charge is

$$c^{g,k} = \frac{k\dim(g)}{k+h(g)} \tag{15.4.3}$$

and that the weight of a primary state is

$$h_r = \frac{Q_r}{(k+h(g))\psi^2} \,. \tag{15.4.4}$$

As in the Virasoro case, all correlations can be reduced to those of the primary fields. In parallel to the derivation of eq. (15.2.3), one finds

$$\langle (j_{-m}^a \cdot \mathcal{O}_1(z_1))\mathcal{O}_2(z_2)\ldots\mathcal{O}_n(z_n)\rangle_{S_2} = \mathcal{J}_{-m}^a \langle \mathcal{O}_1(z_1)\ldots\mathcal{O}_n(z_n)\rangle_{S_2} \,, \tag{15.4.5}$$

where

$$\mathcal{J}_{-m}^a = -\sum_{i=2}^n \frac{t^{a(i)}}{(z_i-z_1)^m} \,, \tag{15.4.6}$$

and so on for multiple raising operators. Here, $t^{a(i)}$ acts in the representation r_i on the primary $\mathcal{O}_i$; the representation indices on $t^{a(i)}$ and $\mathcal{O}_i$ are suppressed.

The Sugawara theory is solved in the same way as the minimal models. In particular, *all* representations are degenerate, and in fact contain null descendants of two distinct types. The first follows directly from the Sugawara form of T, which in modes reads

$$L_m = \frac{1}{(k+h(g))\psi^2}\sum_{n=-\infty}^{\infty} j_n^a j_{m-n}^a \,. \tag{15.4.7}$$

For $m = -1$, this implies that any correlator of primaries is annihilated by

$$\mathscr{L}_{-1} - \frac{2}{(k+h(g))\psi^2}\sum_a t^{a(i)}\mathscr{J}^a_{-1} \,. \tag{15.4.8}$$

This is the *Knizhnik–Zamolodchikov (KZ) equation,*

$$\left[\frac{\partial}{\partial z_1} + \frac{2}{(k+h(g))\psi^2}\sum_a\sum_{i=2}^n \frac{t^{a(1)}t^{a(i)}}{z_1 - z_i}\right]\langle\, \mathscr{O}_1(z_1)\ldots\mathscr{O}_n(z_n)\,\rangle_{S_2} = 0 \,. \tag{15.4.9}$$

We have suppressed group indices on the primary fields, but by writing the correlator in terms of g-invariants, the KZ equation becomes a set of coupled first order differential equations — coupled because there is in general more than one g-invariant for given representations r_i. Exercise 15.5 develops one example. For the leading singularity $(z_1 - z_i)^\kappa$ as $z_1 \to z_i$, the KZ equation reproduces the known result (15.4.4) but does not give fusion rules. There is again a free-field representation of the current algebra (exercise 15.6), analogous to the Feigin–Fuchs representation of the Virasoro algebra.

The second type of null descendant involves the currents only, and does constrain the fusion rules. For convenience, let us focus on the case $g = SU(2)$. The results can then be extended to general g by examining the $SU(2)$ subalgebras associated with the various roots α. We saw in chapter 11 that the $SU(2)$ current algebra has at least two interesting $SU(2)$ Lie subalgebras, namely the global symmetry $j_0^\pm$, j_0^3 and the pseudospin

$$j_{-1}^+ \,, \quad j_0^3 - \frac{k}{2} \,, \quad j_1^- \,. \tag{15.4.10}$$

Now consider some primary field

$$|j,m\rangle \,, \tag{15.4.11}$$

which we have labeled by its quantum numbers under the global $SU(2)$. What are its pseudospin quantum numbers (j',m')? Since it is primary, it is annihilated by the pseudospin lowering operator, so $m' = -j'$. We also have $m' = m - k/2$, so $j' = k/2 - m$. Now, the pseudospin representation has dimension $2j'+1$, so if we raise any state $2j'+1$ times we get zero:

$$(j_{-1}^+)^{k-2m+1}|j,m\rangle = 0 \,. \tag{15.4.12}$$

This is the null descendant.

Now take the correlation of this descendant with some current algebra primaries and use the relation (15.4.5) between the correlators of descendants and primaries to obtain

$$\begin{aligned} 0 &= \left\langle (\mathscr{J}_{-1}^+)^{k-2m_1+1}\cdot \mathscr{O}_1(z_1)\ldots\mathscr{O}_n(z_n)\right\rangle_{S_2} \\ &= \left[-\sum_{i=2}^n \frac{t^{+(i)}}{z_i - z_1}\right]^{k-2m_1+1} \langle\, \mathscr{O}_1(z_1)\ldots\mathscr{O}_n(z_n)\,\rangle_{S_2} \,. \end{aligned} \tag{15.4.13}$$

Notice that, unlike the earlier null equations, this one involves no derivatives and is purely algebraic. To see how this constrains the operator products, consider the three-point correlation. By considering the separate z_i dependences in eq. (15.4.13) one obtains

$$0 = \sum_{m_2,m_3} [(t^{+(2)})^{l_2}]_{m_2,n_2}[(t^{+(3)})^{l_3}]_{m_3,n_3}\,\langle\, \mathscr{O}_{j_1,m_1}\mathscr{O}_{j_2,m_2}\mathscr{O}_{j_3,m_3}\,\rangle_{S_2} \,, \tag{15.4.14}$$

where we have now written out the group indices explicitly. This holds for all n_2 and n_3, and for

$$l_2 + l_3 \geq k - 2m_1 + 1 \,. \tag{15.4.15}$$

The matrix elements of $(t^+)^l$ are nonvanishing for at least some $n_{2,3}$ if $m_2 \geq l_2 - j_2$ and $m_3 \geq l_3 - j_3$. Noting the restriction on $l_{2,3}$, we can conclude that the correlation vanishes when $m_2 + m_3 \geq k - 2m_1 + 1 - j_2 - j_3$. Using $m_1 + m_2 + m_3 = 0$ and taking $m_1 = j_1$ (the most stringent case) gives

$$\langle\, \mathscr{O}_{j_1,j_1}\mathscr{O}_{j_2,m_2}\mathscr{O}_{j_3,m_3}\,\rangle_{S_2} = 0 \quad \text{if} \quad j_1 + j_2 + j_3 > k \,. \tag{15.4.16}$$

Although this was derived for $m_1 = j_1$, rotational invariance now guarantees that it applies for all m_1. Applying also the standard result for multiplication of $SU(2)$ representations, we have the fusion rule

$$[j_1] \times [j_2] = [\,|j_1 - j_2|\,] + [\,|j_1 - j_2| + 1] + \ldots + [\min(j_1 + j_2, k - j_1 - j_2)] \,. \tag{15.4.17}$$

Again there is a simple current, the maximum value $j = k/2$:

$$[j_1] \times [k/2] = [k/2 - j_1] \,. \tag{15.4.18}$$

The corresponding $\mathbf{Z}_2$ symmetry is simply $(-1)^{2j}$.

Modular invariance

The spectrum of a $g \times g$ current algebra will contain some number $n_{r\tilde{r}}$ of each highest weight representation $|r,\tilde{r}\rangle$. The partition function is then

$$Z(\tau) = \sum_{r,\tilde{r}} n_{r\tilde{r}}\chi_r(q)\chi_{\tilde{r}}(q)^* \,, \tag{15.4.19}$$

with the character defined by analogy to that for the conformal algebra, eq. (15.2.12). Invariance under $\tau \to \tau + 1$ amounts as usual to level

matching, so $n_{r\tilde{r}}$ can be nonvanishing only when $h_r - h_{\tilde{r}}$ is an integer. Under $\tau \to -1/\tau$ the characters mix,

$$\chi_r(q') = \sum_{r'} S_{rr'}\chi_{r'}(q) , \tag{15.4.20}$$

so the condition for modular invariance is the matrix equation

$$S^\dagger n S = n . \tag{15.4.21}$$

The characters are obtained by considering all states generated by the raising operators, with appropriate allowance for degeneracy. Only the currents need be considered, since by the Sugawara relation the Virasoro generators do not generate any additional states. The calculation is then parallel to the calculation of the characters of finite Lie algebras, and the result is similar to the Weyl character formula. The details are too lengthy to repeat here, and we will only mention one simple classic result: the modular S matrix for $SU(2)$ at level k is

$$S_{jj'} = \left(\frac{2}{k+2}\right)^{1/2} \sin\frac{\pi(2j+1)(2j'+1)}{k+2} . \tag{15.4.22}$$

The general solution to the modular invariance conditions is known. One solution, at any level, is the diagonal modular invariant for which each representation with $j = \tilde{\jmath}$ appears once:

$$n_{j\tilde{\jmath}} = \delta_{j\tilde{\jmath}} . \tag{15.4.23}$$

These are known as the A invariants. When the level k is even, there is another solution obtained by twisting with respect to $(-1)^{2j}$. One keeps the previous states with j integer only, and adds in a twisted sector where $\tilde{\jmath} = k/2 - j$. For k a multiple of 4, j in the twisted sector runs over integers, while for $k+2$ a multiple of 4, j in the twisted sector runs over half-integers:

$$n_{j\tilde{\jmath}} = \delta_{j\tilde{\jmath}}\Big|_{j\in\mathbf{Z}} + \delta_{k/2-j,\tilde{\jmath}}\Big|_{j\in\mathbf{Z}+k/4} . \tag{15.4.24}$$

These are known as the D invariants. For the special values $k = 10, 16, 28$ there are exceptional solutions, the E invariants. The A–D–E terminology refers to the simply-laced Lie algebras. The solutions are in one-to-one correspondence with these algebras, the Dynkin diagrams arising in the construction of the invariants.

Strings on group manifolds

Thus far the discussion has used only symmetry, without reference to a Lagrangian. There is an important Lagrangian example of a current

algebra. Let us start with a simple case, a nonlinear sigma model with a three-dimensional target space,

$$S = \frac{1}{2\pi\alpha'}\int d^2z\,(G_{mn} + B_{mn})\partial X^m \bar{\partial} X^n . \tag{15.4.25}$$

Let G_{mn} be the metric of a 3-sphere of radius r and let the antisymmetric tensor field strength be

$$H_{mnp} = \frac{q}{r^3}\epsilon_{mnp} \tag{15.4.26}$$

for some constant q; ϵ_{mnp} is a tensor normalized to $\epsilon_{mnp}\epsilon^{mnp} = 6$. The curvature is

$$R_{mn} = \frac{2}{r^2}G_{mn} . \tag{15.4.27}$$

To leading order in α', the nonvanishing beta functions (3.7.14) for this nonlinear sigma model are

$$\beta^G_{mn} = \alpha' G_{mn}\left(\frac{2}{r^2} - \frac{q^2}{2r^6}\right) , \tag{15.4.28a}$$

$$\beta^\Phi = \frac{1}{2} - \frac{\alpha' q^2}{4r^6} . \tag{15.4.28b}$$

The first term in β^Φ is the contribution of three free scalars. The theory is therefore conformally invariant to leading order in α' if

$$r^2 = \frac{|q|}{2} + O(\alpha') . \tag{15.4.29}$$

The central charge is

$$c = 6\beta^\Phi = 3 - \frac{6\alpha'}{r^2} + O\left(\frac{\alpha'^2}{r^4}\right) . \tag{15.4.30}$$

A 3-sphere has symmetry algebra $O(4) = SU(2) \times SU(2)$. In a CFT, we know that each current will be either holomorphic or antiholomorphic. Comparing with the $SU(2)$ Sugawara central charge

$$c = 3 - \frac{6}{k+2} , \tag{15.4.31}$$

the sigma model is evidently a Sugawara theory. One $SU(2)$ will be left-moving on the world-sheet and one right-moving.

The general analysis of current algebras showed that the level k is quantized. In the nonlinear sigma model it arises from the Dirac quantization condition. The argument is parallel to that in section 13.3. A nonzero total flux H is incompatible with $H = dB$ for a single-valued B. We can write the dependence of the string amplitude on this background as

$$\exp\left(\frac{i}{2\pi\alpha'}\int_M B\right) = \exp\left(\frac{i}{2\pi\alpha'}\int_N H\right) , \tag{15.4.32}$$

where M is the embedding of the world-sheet in the target space and N is any three-dimensional manifold in S_3 whose boundary is M. In order that this be independent of the choice of N we need

$$1 = \exp\left(\frac{i}{2\pi\alpha'}\int_{S_3} H\right) = \exp\left(\frac{\pi i q}{\alpha'}\right) . \tag{15.4.33}$$

Thus,

$$q = 2\alpha' n , \quad r^2 = \alpha' n \tag{15.4.34}$$

for integer n. More generally, $\int H$ over any closed 3-manifold in spacetime must be a multiple of $4\pi^2\alpha'$.

This is the desired quantization, and $|n|$ is just the level k of the current algebra. In particular, the one-loop central charge (15.4.30) becomes

$$c = 3 - \frac{6}{|n|} + O\left(\frac{1}{n^2}\right) , \tag{15.4.35}$$

agreeing with the current algebra result to this order.

The 3-sphere is the same as the $SU(2)$ group manifold, under the identification

$$g = x^4 + ix^i\sigma^i , \quad \sum_{i=1}^{4}(x^i)^2 = 1 . \tag{15.4.36}$$

The action (15.4.25) can be rewritten as the *Wess–Zumino–Novikov–Witten (WZNW) action*

$$S = \frac{|n|}{4\pi}\int_M d^2z\,\mathrm{Tr}(\partial g^{-1}\bar\partial g) + \frac{in}{12\pi}\int_N \mathrm{Tr}(\omega^3) , \tag{15.4.37}$$

where $\omega = g^{-1}dg$ is the *Maurer–Cartan 1-form*. Here M is the embedding of the world-sheet in the group manifold, and N is any 3-surface in the group manifold whose boundary is M. In this form, the action generalizes to any Lie group g. The second term is known as the *Wess–Zumino* term. The reader can check that

$$d(\omega^3) = 0 . \tag{15.4.38}$$

Therefore, locally on the group $\omega^3 = d\chi$ for some 2-form χ, and the Chern–Simons term can be written as a two-dimensional action

$$\frac{n}{12\pi}\int_M \mathrm{Tr}(\chi) . \tag{15.4.39}$$

As with the magnetic monopole, there is no such χ that is nonsingular on the whole space.

The variation of the WZNW action is[3]

$$\begin{aligned}\delta S &= \frac{|n|}{2\pi}\int d^2z\,\mathrm{Tr}\,[\bar\partial g\, g^{-1}\partial(g^{-1}\delta g)] \\ &= \frac{|n|}{2\pi}\int d^2z\,\mathrm{Tr}\,[g^{-1}\partial g\,\bar\partial(\delta g g^{-1})] .\end{aligned} \tag{15.4.40}$$

As guaranteed by conformal invariance, the global $g \times g$ symmetry

$$\delta g(z,\bar z) = i\epsilon_L g(z,\bar z) - i g(z,\bar z)\epsilon_R \tag{15.4.41}$$

is elevated to a current algebra,

$$\delta g(z,\bar z) = i\epsilon_L(z) g(z,\bar z) - i g(z,\bar z)\epsilon_R(\bar z) . \tag{15.4.42}$$

Left-multiplication is associated with a left-moving current algebra and right-multiplication with a right-moving current algebra. The currents are

$$|n|\mathrm{Tr}(\epsilon_R g^{-1}\partial g) , \quad |n|\mathrm{Tr}(\epsilon_L\bar\partial g g^{-1}) . \tag{15.4.43}$$

Let us check that the Poisson bracket of two currents has the correct c-number piece. To get this, it is sufficient to expand

$$g = 1 + i(2|n|)^{-1/2}\phi_a\sigma^a + \ldots \tag{15.4.44}$$

and keep the leading terms in the Lagrangian density and currents,

$$\mathscr{L} = \frac{1}{4\pi}\partial\phi^a\bar\partial\phi^a + O(\phi^3) , \tag{15.4.45a}$$

$$j_R^a = |n|^{1/2}\partial\phi^a + O(\phi^2) , \tag{15.4.45b}$$

$$j_L^a = |n|^{1/2}\bar\partial\phi^a + O(\phi^2) . \tag{15.4.45c}$$

The higher-order terms do not contribute to the c-number in the Poisson bracket. The kinetic term now has the canonical $\alpha' = 2$ normalization so the level $k = |n|$ follows from the normalization of the currents.

Which states appear in the spectrum? We can make an educated guess by thinking about large k, where the group manifold becomes more and more flat. The currents then approximate free boson modes so the primary states, annihilated by the raising operators, have no internal excitations — the vertex operators are just functions of g. The representation matrices form a complete set of such functions, so we identify

$$D^r_{ij}(g) = \mathscr{O}^r_i(z)\tilde{\mathscr{O}}^r_j(\bar z) . \tag{15.4.46}$$

This transforms as the representation (r,r) under $g\times g$, so summing over all r gives the diagonal modular invariant. Recall that for each k the number of primaries is finite; $D^r_{ij}(g)$ for higher r evidently is not primary. This reasoning is correct for simply connected groups, but otherwise we must

[3] This is for $n > 0$; for $n < 0$ interchange z and $\bar z$.

exclude some representations and add in winding sectors. For example, $SU(2)/\mathbf{Z}(2) = O(3)$ leads to the D invariant. We can understand the restriction to even levels for the D invariant: $\int H$ on $SU(2)/\mathbf{Z}(2)$ is half of $\int H$ on $SU(2)$, so the coefficient must be even to give a well-defined path integral.

The group manifold example vividly shows how familiar notions of spacetime are altered in string theory. If we consider eight flat dimensions with both right- and left-moving momenta compactified on the E_8 root lattice, we obtain an $E_{8L} \times E_{8R}$ current algebra at level one. We get the same theory with *248 dimensions* forming the E_8 group manifold with unit H charge.

15.5 Coset models

A clever construction allows us to obtain from current algebras the minimal models and many new CFTs. Consider a current algebra G, which might be a sum of several factors (g_i, k_i). Let H be some subalgebra. Then as in the discussion of Sugawara theories we can separate the energy-momentum tensor into two pieces,

$$T^G = T^H + T^{G/H} . \tag{15.5.1}$$

The central charge of $T^{G/H}$ is

$$c^{G/H} = c^G - c^H . \tag{15.5.2}$$

For any subalgebra the Sugawara theory thus separates into the Sugawara theory of the subalgebra, and a new *coset* CFT. A notable example is

$$G = SU(2)_k \oplus SU(2)_1 , \quad c^G = 4 - \frac{6}{k+2} , \tag{15.5.3a}$$

$$H = SU(2)_{k+1} , \quad c^H = 3 - \frac{6}{k+3} , \tag{15.5.3b}$$

where the subscripts denote the levels. Here, the H currents are the sums of the currents of the two $SU(2)$ current algebras in G, $j^a = j^a_{(1)} + j^a_{(2)}$. Then the central charges

$$c^{G/H} = 1 - \frac{6}{(k+2)(k+3)} \tag{15.5.4}$$

are precisely those of the unitary minimal models with $m = k + 2$.

A representation of the G current algebra can be decomposed under the subalgebras,

$$\chi^G_r(q) = \sum_{r',r''} n^r_{r'r''} \chi^H_{r'}(q) \chi^{G/H}_{r''}(q), \tag{15.5.5}$$

where r is any representation of G, and r' and r'' respectively run over all H and G/H representations, with $n^r_{r'r''}$ nonnegative integers. For the minimal model coset (15.5.3), all unitary representations can be obtained in this way. The current algebra theories are rather well understood, so this is often a useful way to represent the coset theory. For example, while the Kac determinant gives necessary conditions for a minimal model representation to be unitary, the coset construction is regarded as having provided the existence proof, the unitary current algebra representations having been constructed directly. The minimal model fusion rules (15.3.19) can be derived from the $SU(2)$ current algebra rules (15.4.17), and the minimal model modular transformation

$$S_{rs,r's'} = \left[\frac{8}{(p+1)(q+1)}(-1)^{(r+s)(r'+s')}\right]^{1/2} \sin\frac{\pi r r'}{p} \sin\frac{\pi s s'}{q} \tag{15.5.6}$$

can be obtained from the $SU(2)$ result (15.4.22). Further, the minimal model modular invariants are closely related to the $SU(2)$ A–D–E invariants.

Taking various G and H leads to a wealth of new theories. In this section and the next we will describe only some of the most important examples, and then in section 15.7 we discuss some generalizations. The coset construction can be regarded as gauging the subalgebra H. Conformal invariance forbids a kinetic term for the gauge field, and the equation of motion for this field then requires the H-charge to vanish, leaving the coset theory. This is the gauging of a continuous symmetry; equivalently, one is treating the H currents as constraints. Recall that gauging a discrete symmetry gave the orbifold (twisting) construction.

The *parafermionic theories* are:

$$\frac{SU(2)_k}{U(1)} , \quad c = 2 - \frac{6}{k+2} . \tag{15.5.7}$$

Focusing on the $U(1)$ current algebra generated by j^3, by the OPE we can write this in terms of a left-moving boson H with standard normalization $H(z)H(0) \sim -\ln z$:

$$j^3 = i(k/2)^{1/2}\partial H , \quad T^H = -\frac{1}{2}\partial H \partial H . \tag{15.5.8}$$

Operators can be separated into a free boson part and a parafermionic part. For the $SU(2)$ currents themselves we have

$$j^+ = \exp[iH(2/k)^{1/2}]\psi_1 , \quad j^- = \exp[-iH(2/k)^{1/2}]\psi_1^\dagger , \tag{15.5.9}$$

where ψ_1 is known as the parafermionic current. Subtracting the weight

of the exponential, the current has weight $(k-1)/k$. One obtains further currents

$$:(j^+)^l: = (j^+_{-1})^l \cdot 1 \equiv \exp(ilH(2/k)^{1/2})\psi_l \,, \tag{15.5.10}$$

with ψ_l having weight $l(k-l)/k$. The current algebra null vector (15.4.12) implies that ψ_l vanishes for $l > k$, which could also have been anticipated from its negative weight. The weight also implies that $\psi_0 = \psi_k = 1$, and from this one can also deduce that $\psi_l = \psi^\dagger_{k-l}$. The current algebra primaries similarly separate,

$$\mathcal{O}_{j,m} = \exp[imH(2/k)^{1/2}]\psi^j_m \,, \tag{15.5.11}$$

where ψ^j_m is a primary field of the parafermion algebra, and has weight $j(j+1)/(k+2) - m^2/k$.

Factoring out the OPE of the free boson, the operator products of the parafermionic currents become

$$\psi_l(z)\psi_{l'}(0) \approx z^{-2ll'/k}(\psi_{l+l'} + \ldots) \,. \tag{15.5.12}$$

This algebra is more complicated than those encountered previously, in that the currents have branch cuts with respect to each other. However, it is simple in one respect: each pair of currents has definite *monodromy*, meaning that all terms in the operator product change by the same phase, $\exp(-4\pi i ll'/k)$, when one current circles the other. We will mention an application of the parafermion theories later.

For small k, the parafermion theories reduce to known examples. For $k = 1$, the parafermion central charge is zero and the parafermion theory trivial. In other words, at $k = 1$ the free boson is the whole $SU(2)$ current algebra: this is just the torus at its self-dual radius. For $k = 2$, the parafermion central charge is $\frac{1}{2}$, so the parafermion must be an ordinary free fermion. We recall from section 11.5 that $SU(2)$ at $k = 2$ can be represented in terms of three free fermions. The free boson H is obtained by bosonizing $\psi^{1,2}$, leaving ψ^3 as the parafermion. At $k = 3$ the parafermion central charge is $\frac{4}{5}$, identifying it as the $m = 5$ unitary minimal model.

Although constructed as $SU(2)$ cosets, the minimal models have no $SU(2)$ symmetry nor other any weight 1 primaries. In order for an operator from the G theory to be part of the coset theory, it must be nonsingular with respect to the H currents, and no linear combination of the currents $j^a_{(1)}$ and $j^a_{(2)}$ is nonsingular with respect to $j^a_{(1)} + j^a_{(2)}$. The situation becomes more interesting if we consider the bilinear invariants

$$:j^a_{(1)}j^a_{(1)}: \,, \quad :j^a_{(1)}j^a_{(2)}: \,, \quad :j^a_{(2)}j^a_{(2)}: \,. \tag{15.5.13}$$

In parallel with the calculations in exercise 11.7, the operator product of

the H current with these bilinears is

$$\left[j^b_{(1)}(z) + j^b_{(2)}(z)\right] :j^a_{(i)}j^a_{(j)}(0): = \sum_{k=0}^{\infty} \frac{1}{z^{k+1}} \left[j^b_{k(1)} + j^b_{k(2)}\right] j^a_{-1(i)} j^a_{-1(j)} \cdot 1 \,. \tag{15.5.14}$$

The $k = 0$ term vanishes because the bilinear is G-invariant. For $k = 1$, commuting the lowering operator to the right gives a linear combination of $j^b_{-1(1)}$ and $j^b_{-1(2)}$. All higher poles vanish. Thus, there are three bilinear invariants and only two possible singularities, so one linear combination commutes with the H current and lies entirely within the coset theory. This is just the coset energy-momentum tensor $T^{G/H}$, which we already know.

For $SU(2)$ cosets that is the end of the story, but let us consider the generalization

$$G = SU(n)_{k_1} \oplus SU(n)_{k_2} \,, \quad c^G = (n^2-1)\left[\frac{k_1}{k_1+n} + \frac{k_2}{k_2+n}\right] \,, \tag{15.5.15a}$$

$$H = SU(n)_{k_1+k_2}, \quad c^H = (n^2-1)\frac{k_1+k_2}{k_1+k_2+n} \,. \tag{15.5.15b}$$

For $n \geq 3$ there is a symmetric cubic invariant

$$d^{abc} \propto \mathrm{Tr}(t^a\{t^b, t^c\}) \,, \tag{15.5.16}$$

which vanishes for $n = 2$. Similarly, for $n \geq 4$ there is an independent symmetric quartic invariant, and so forth. Using the cubic invariant, we can construct the four invariants $d^{abc}\,:j^a_{(i)}j^b_{(j)}j^c_{(k)}:$. The operator product with the H current has three possible singularities, $z^{-2}d^{abc}\,:j^b_{(j)}j^c_{(k)}:$, so there must be one linear combination $W(z)$ that lies in the coset theory. That is, the coset theory has a conserved spin-3 current. The states of the coset theory fall in representations of an *extended chiral algebra*, consisting of the Laurent modes of $T(z)$, $W(z)$, and any additional generators needed to close the algebra.

In general, the algebra contains higher spin currents as well. For example, the operator product $W(z)W(0)$ contains a spin-4 term involving the product of four currents. For the special case $n = 3$ and $k_2 = 1$, making use of the current algebra null vectors, the algebra of $T(z)$ and $W(z)$ actually closes without any new fields. It is the W_3 *algebra*, which in OPE form is

$$\begin{aligned} W(z)W(0) \sim{}& \frac{c}{3z^6} + \frac{2}{z^4}T(0) + \frac{1}{z^3}\partial T(0) + \frac{3}{10z^2}\partial^2 T(0) + \frac{1}{15z}\partial^3 T(0) \\ &+ \frac{16}{220+50c}\left(\frac{2}{z^2} + \frac{1}{z}\partial\right)[10 :T^2(0): -3\partial^2 T(0)] \,. \end{aligned} \tag{15.5.17}$$

In contrast to the various algebras we have encountered before, this one is *nonlinear*: the spin-4 term involves the square of $T(z)$. This is the only

closed algebra containing only a spin-2 and spin-3 current and was first discovered by imposing closure directly. It has a representation theory parallel to that of the Virasoro algebra, and in particular has a series of unitary degenerate representations of central charge

$$c = 2 - \frac{24}{(k+3)(k+4)} . \tag{15.5.18}$$

The $(k_1, k_2, n) = (k, 1, 3)$ cosets produce these representations. As it happens, the first nontrivial case is $k = 1$, $c = \frac{4}{5}$, which as we have seen also has a parafermionic algebra. The number of extended chiral algebras is enormous, and they have not been fully classified.

15.6 Representations of the $N = 1$ superconformal algebra

All the ideas of this chapter generalize to the superconformal algebras. In this section we will describe only the basics: the Kac formula, the discrete series, and the coset construction.

A highest weight state, of either the R or NS algebra, is annihilated by L_n and G_n for $n > 0$. The representation is generated by L_n for $n < 0$ and G_n for $n \leq 0$. Each G_n acts at most once, since $G_n^2 = L_{2n}$. The Kac formula for the R and NS algebras can be written in a uniform way,

$$\det(\mathcal{M}^N)_{\mathrm{R,NS}} = (h - \epsilon\hat{c}/16)K_N \prod_{1 \leq rs \leq 2N} (h - h_{r,s})^{P_{\mathrm{R,NS}}(N - rs/2)} . \tag{15.6.1}$$

Here, ϵ is 1 in the R sector and 0 in the NS sector. The zeros are at

$$h_{r,s} = \frac{\hat{c} - 1 + \epsilon}{16} + \frac{1}{4}(r\hat{\alpha}_+ + s\hat{\alpha}_-)^2 , \tag{15.6.2}$$

where $r - s$ must be even in the R sector and odd in the NS sector. We have defined $\hat{c} = 2c/3$ and

$$\hat{\alpha}_\pm = \frac{1}{4}\left[(1 - \hat{c})^{1/2} \pm (9 - \hat{c})^{1/2}\right] . \tag{15.6.3}$$

The multiplicity of each zero is again the number of ways a given level can be reached by the raising operators of the theory,

$$\prod_{n=1}^{\infty} \frac{1 + q^{n-1}}{1 - q^n} = \sum_{k=0}^{\infty} P_{\mathrm{R}}(k) q^k , \tag{15.6.4a}$$

$$\prod_{n=1}^{\infty} \frac{1 + q^{n-1/2}}{1 - q^n} = \sum_{k=0}^{\infty} P_{\mathrm{NS}}(k) q^k . \tag{15.6.4b}$$

Unitary representations are allowed at

$$\hat{c} \geq 1 , \quad h \geq \epsilon\frac{\hat{c}}{16} , \tag{15.6.5}$$

and at the discrete series

$$c = \frac{3}{2} - \frac{12}{m(m+2)} , \quad m = 2, 3, \ldots ,$$
$$= 0, \frac{7}{10}, 1, \frac{81}{70}, \ldots , \tag{15.6.6a}$$

$$h = h_{r,s} \equiv \frac{[r(m+2) - sm]^2 - 4}{8m(m+2)} + \frac{\epsilon}{16} , \tag{15.6.6b}$$

where $1 \leq r \leq m - 1$ and $1 \leq s \leq m + 1$.

A coset representation for the $N = 1$ unitary discrete series is

$$G = SU(2)_k \oplus SU(2)_2 , \quad H = SU(2)_{k+2} . \tag{15.6.7}$$

The central charge is correct for $m = k + 2$. The reader can verify that the coset theory has $N = 1$ world-sheet supersymmetry: using the free fermion representation of the $k = 2$ factor, one linear combination of the $(\frac{3}{2}, 0)$ fields $j^a_{(1)}\psi^a$ and $i\epsilon^{abc}\psi^a\psi^b\psi^c$ is nonsingular with respect to the H current and is the supercurrent of the coset theory.

For small m, some of these theories are familiar. At $m = 2$, c vanishes and we have the trivial theory. At $m = 3$, $c = \frac{7}{10}$, which is the $m = 4$ member of the Virasoro unitary series. At $m = 4$, $c = 1$; this is the free boson representation discussed in section 10.7.

15.7 Rational CFT

We have seen that holomorphicity on the world-sheet is a powerful property. It would be useful if a general local operator of weight $(h, \tilde{h})$ could be divided in some way into a holomorphic $(h, 0)$ field times an antiholomorphic $(0, \tilde{h})$ field, or a sum of such terms. The conformal block expression (15.2.9) shows the sense in which this is possible: by organizing intermediate states into conformal families, the correlation function is written as a sum of terms, each holomorphic times antiholomorphic. While this was carried out for the four-point function on the sphere, it is clear that the derivation can be extended to n-point functions on arbitrary Riemann surfaces. For example, the conformal blocks of the zero-point function on the torus are just the characters,

$$Z(\tau) = \sum_{i,j} n_{ij} \chi_i(q) \chi_j(q)^* , \tag{15.7.1}$$

where n_{ij} counts the number of times a given representation of the left and right algebras appears in the spectrum.

When the sum is infinite this factorization does not seem particularly helpful, but when the sum is finite it is. In fact, in all the examples discussed in this section, and in virtually all known exact CFTs, the sum

is finite. What is happening is that the spectrum, though it must contain an infinite number of Virasoro representations for $c \geq 1$, consists of a finite number of representations of some larger extended chiral algebra. This is the definition of a *rational conformal field theory (RCFT)*.

It has been conjectured that all rational theories can be represented as cosets, and that *any* CFT can be arbitrarily well approximated by a rational theory (see exercise 15.9 for an example). If so, then we are close to constructing the general CFT, but the second conjecture in particular seems very optimistic.

We will describe here a few of the general ideas and results. The basic objects in RCFT are the conformal blocks and the *fusion rules*, nonnegative integers N_{ij}^k which count the number of ways the representations i and j can be combined to give the representation k. For the Virasoro algebra, we know that two representations can be combined to give a third in a unique way: the expectation value of the primaries determines those of all descendants. For other algebras, N_{ij}^k may be greater than 1. For example, even for ordinary Lie algebras there are two ways to combine two adjoint **8**s of $SU(3)$ to make another adjoint, namely d^{abc} and f^{abc}. As a result, the same holds for the corresponding current algebra representations: $N_{\mathbf{88}}^{\mathbf{8}} = 2$.

Repeating the derivation of the conformal blocks, for a general algebra the number of independent blocks $\mathcal{F}_{ij}^{kl}(r|z)$ is

$$N_{ijkl} = N_{ij}^r N_{rkl} \ , \tag{15.7.2}$$

where the repeated index is summed. Indices are lowered with $N_{ij}^0 = N_{ij}$, zero denoting the identity representation. One can show that for each i, N_{ij} is nonvanishing only for a single j. This defines the conjugate representation, $N_{i\bar{i}} = 1$. In the minimal models and $SU(2)$ current algebra, all representations are self-conjugate, but for $SU(n)$, $n > 2$ for example, they are not. By associativity, the s-channel conformal blocks $\mathcal{F}_{ij}^{kl}(r|z)$ are linearly related to the t-channel blocks $\mathcal{F}_{il}^{jk}(r|1-z)$. The number of independent functions must be the same in each channel, so the fusion rules themselves satisfy an associativity relation,

$$N_{ij}^r N_{rkl} = N_{ik}^r N_{rjl} = N_{il}^r N_{rjk} \ . \tag{15.7.3}$$

We will now derive two of the simpler results in this subject, namely that the weights and the central charge must in fact be rational numbers in an RCFT. First note that the conformal blocks are not single-valued on the original Riemann surface — they have branch cuts — but they are single-valued on the covering space, where a new sheet is defined whenever one vertex operator circles another. Any series of moves that brings the vertex operators back to their original positions and sheets must leave the conformal blocks invariant. For example,

$$\tau_1\tau_2\tau_3\tau_4 = \tau_{12}\tau_{13}\tau_{23} \ , \tag{15.7.4}$$

where $\tau_{i\cdots j}$ denotes a *Dehn twist*, cutting open the surface on a circle containing the indicated vertex operators, rotating by 2π and gluing. To see this, examine for example vertex operator 1. On the right-hand side, the combined effect of τ_{12} and τ_{13} is for this operator to circle operators 2 and 3 and to rotate by 4π. On the left, this is the same as the combined effect of τ_4 (which on the sphere is the same as τ_{123}) and τ_1. Eq. (15.7.4) is an N_{ijkl}-dimensional matrix equation on the conformal blocks. For example,

$$\tau_1 : \quad \mathcal{F}_{ij}^{kl}(r|z) \to \exp(2\pi i h_i)\mathcal{F}_{ij}^{kl}(r|z) \ , \tag{15.7.5a}$$

$$\tau_{12} : \quad \mathcal{F}_{ij}^{kl}(r|z) \to \exp(2\pi i h_r)\mathcal{F}_{ij}^{kl}(r|z) \ . \tag{15.7.5b}$$

On the other hand, τ_{13} is not diagonal in this basis, but rather in the dual basis $\mathcal{F}_{il}^{jk}(r|1-z)$.

In order to get a basis-independent statement, take the determinant of eq. (15.7.4) and use (15.7.5) to get

$$N_{ijkl}(h_i + h_j + h_k + h_l) - \sum_r (N_{ij}^r N_{rkl} + N_{ik}^r N_{rjl} + N_{il}^r N_{rjk})h_r \in \mathbf{Z} \ . \tag{15.7.6}$$

This step is possible only when the number $\mathcal{N}$ of primaries is finite.

There are many more equations than weights. Focusing on the special case $i = j = k = l$ gives

$$\sum_r N_{ii}^r N_{rii}(4h_i - 3h_r) \in \mathbf{Z} \ . \tag{15.7.7}$$

This is $\mathcal{N} - 1$ equations for $\mathcal{N} - 1$ weights, where $\mathcal{N}$ is the number of primaries; the weight h_0 is always 0, and the $i = 0$ equation is trivial. Let us consider the example of $SU(2)$ current algebra at level 3, where there are four primaries, $j = 0, \frac{1}{2}, 1, \frac{3}{2}$. From the general result (15.4.17), the nonzero fusion rules of the form N_{ii}^r are

$$N_{00}^0 = N_{1/2,1/2}^0 = N_{1/2,1/2}^1 = N_{11}^0 = N_{11}^1 = N_{3/2,3/2}^0 = 1 \ . \tag{15.7.8}$$

Thus we find that

$$8h_{1/2} - 3h_1 \ , \quad 5h_1 \ , \quad 4h_{3/2} \tag{15.7.9}$$

are all integers, which implies that the weights are all rational. These results are consistent with the known weights $j(j+1)/(k+2)$. The reader can show that eqs. (15.7.7) are always nondegenerate and therefore require the weights to be rational.[4]

[4] We are assuming that all the N_{iiii} are nonzero. More generally, one can derive a similar relation with $N_{ii\bar{i}\bar{i}}$, which is always positive.

For the central charge, consider the zero-point function on the torus. The covering space here is just Teichmüller space, on which one may check that

$$S^4 = (ST)^3 = 1 . \tag{15.7.10}$$

The determinant of this implies that

$$1 = [(\det S)^4]^{-3}[(\det S \det T)^3]^4 = (\det T)^{12} . \tag{15.7.11}$$

The transformation T acts on the characters as

$$T : \quad \chi_i(q) \to \exp[2\pi i(h_i - c/24)]\chi_i(q) . \tag{15.7.12}$$

Thus,

$$\frac{\mathcal{N}c}{2} - 12\sum_i h_i \in \mathbf{Z} , \tag{15.7.13}$$

and the rationality of c follows from that of the weights.

The consistency conditions for RCFT have been developed in a systematic way. Let us just mention some of the most central results. The first is the *Verlinde formula*, which determines the fusion rules in terms of the modular transformation S:

$$N_{jk}^i = \sum_r \frac{S_j^r S_k^r S_r^{\dagger i}}{S_0^r} . \tag{15.7.14}$$

Indices are lowered with N_{ij}^0. The second is *naturalness*: any operator product coefficient that is allowed by the full chiral algebra is actually nonzero.[5] The third result describes all possible modular invariants (15.7.1): either $n_{i\bar{\jmath}} = \delta_{i\bar{\jmath}}$ (the diagonal invariant), or $n_{i\bar{\jmath}} = \delta_{i\omega(\bar{\jmath})}$, where $\omega(\bar{\jmath})$ is some permutation symmetry of the fusion rules. The latter two results are not quite as useful as they sound, because they only hold with respect to the full chiral algebra of the theory. As we have seen in the W algebra coset example, this may be larger than one realizes.

Finally, let us mention a rather different generalization of the coset idea. Suppose we have a current algebra G, and we consider all $(2,0)$ operators formed from bilinears in the currents,

$$T' = L_{ab} :j^a j^b: . \tag{15.7.15}$$

The condition that the TT OPE has the correct form for an energy-momentum tensor, and therefore that the modes of T form a Virasoro algebra, is readily found. It is the *Virasoro master equation*,

$$L_{ab} = 2L_{ac}k^{cd}L_{db} - L_{cd}L_{ef}f_a^{ce}f_b^{df} - L_{cd}f_f^{ce}f_{(a}^{df}L_{b)e} , \tag{15.7.16}$$

[5] This precise statement holds only when the N_{jk}^i are restricted to the values 0 and 1; otherwise, it requires some refinement.

where k^{ab} is the coefficient $1/z^2$ in the current–current OPE. The central charge is

$$c = 2k^{ab}L_{ab} . \tag{15.7.17}$$

We already know some solutions to this: the Sugawara tensor for G, or for any subalgebra H of G. Remarkably, the set of solutions is very much larger: for $G = SU(3)_k$, the number has been estimated as $\frac{1}{4}$ *billion* for each k. For each solution the G theory separates into two decoupled theories, with energy-momentum tensors T' and $T^G - T'$. Some of these may be equivalent to known theories, but others are new and many have irrational central charge.

15.8 Renormalization group flows

Consistent string propagation requires a conformally invariant world-sheet theory, but there are several reasons to consider the relation of CFTs to the larger set of all two-dimensional field theories. First, CFT also has application to the description of critical phenomena, where the parameters can be varied away from their critical values. Second, there is a rich mathematical and physical interplay between conformal theories and nearby nonconformal ones, each illuminating the other. Third, conformally invariant theories correspond to string backgrounds that satisfy the classical equations of motion. One might then guess that the proper setting for quantum string theory would be a path integral over all background field configurations — that is, over all two-dimensional quantum field theories. This last is more speculative; it is related to other formulations of string field theory, a subject discussed briefly in chapter 9.

In this section we will develop some general results relating conformal and nonconformal theories. In the next we will discuss some examples and applications. Once again, this is an enormous subject and we can only sketch a few of the central ideas and results.

Scale invariance and the renormalization group

Consider the scale transformation

$$\delta_s z = \epsilon z \tag{15.8.1}$$

on a world-sheet with flat metric $g_{ab} = \delta_{ab}$. Alternatively we could keep the coordinates fixed and scale up the metric,

$$\delta_s g_{ab} = 2\epsilon g_{ab} . \tag{15.8.2}$$

In either form the net change (3.4.6) in the action and measure is

$$-\frac{\epsilon}{2\pi}\int d^2\sigma\, T^a{}_a(\sigma)\,. \tag{15.8.3}$$

A flat world-sheet theory will therefore be scale-invariant provided that

$$T^a{}_a = \partial_a \mathscr{K}^a, \tag{15.8.4}$$

for some local operator $\mathscr{K}^a$.

Scale invariance plays an important role in many parts of physics. One expects that the extreme low energy limit of any quantum field theory will approach a scale-invariant theory. This has not been proven in general, but seems to be true in all examples. The scale-invariant theory may be trivial: if all states are massive then at low enough energy nothing is left. Consider for example a statistical mechanical system. The Boltzmann sum is the same as the Euclidean path integral in quantum field theory. This may have an energy gap for generic values of the parameters and so be trivial at long distance, but when the parameters are tuned to send the gap to zero (a second order phase transition) it is described by a nontrivial scale-invariant theory.

The term *nontrivial* in this context is used in two different ways. The broad usage (which is applied in the previous paragraph) means any field theory without an energy gap, so that there are states of arbitrarily small nonzero energy. A narrower usage reserves the term for scale-invariant theories with interactions that remain important at all distances, as opposed to those whose low energy limit is equivalent to that of a free field theory.

Scale and conformal invariances are closely related. The scale transformation rescales world-sheet distances by a constant factor, leaving angles and ratios of lengths invariant. A conformal transformation rescales world-sheet distances by a position-dependent factor; on a very small patch of the world-sheet it looks like a scale transformation. In particular, conformal transformations leave angles of intersection between curves invariant. Comparing the condition (15.8.4) with the condition $T^a{}_a = 0$ for conformal invariance, one sees that it is possible in principle for a theory to be scale-invariant without being conformally-invariant. However, it is difficult to find examples. Later in the section we will prove that for compact unitary CFTs in two dimensions scale invariance does imply conformal invariance. Exercise 15.12 gives a nonunitary counterexample.

This is of some importance in dimensions greater than two. In the previous chapter we encountered two nontrivial (in the narrow sense) scale-invariant theories. The first was the $d = 4$, $N = 4$ gauge theory. The second was the $d = 6$ $(2,0)$ tensionless string theory, which arose on

coincident IIA or M-theory 5-branes. Both are believed to be conformally invariant.

In quantum field theory, the behavior of matrix elements under a rigid scale transformation is governed by a differential equation, the *renormalization group equation*. Let us derive such an equation. Consider a general quantum field theory in d-dimensional spacetime; spacetime here corresponds to the string world-sheet, which is the case $d = 2$. The scale transformation of a general expectation value is

$$\epsilon^{-1}\delta_s\left\langle\prod_m \mathscr{A}_{i_m}(\sigma_m)\right\rangle = -\frac{1}{2\pi}\int d^2\sigma\left\langle T^a{}_a(\sigma)\prod_m \mathscr{A}_{i_m}(\sigma_m)\right\rangle - \sum_n \Delta_{i_n}{}^j\left\langle \mathscr{A}_j(\sigma_n)\prod_{m\neq n}\mathscr{A}_{i_m}(\sigma_m)\right\rangle, \tag{15.8.5}$$

where $\mathscr{A}_i$ is a complete set of local operators. The second term is from the action of the scale transformation on the operators,

$$\delta_s \mathscr{A}_i(\sigma) = -\Delta_i{}^j \mathscr{A}_j(\sigma)\,. \tag{15.8.6}$$

The integrated trace of the energy-momentum tensor can be expanded in terms of the complete set,

$$\int d^d\sigma\, T^a{}_a = -2\pi{\sum_i}' \int d^d\sigma\, \beta^i(g)\mathscr{A}_i\,. \tag{15.8.7}$$

The prime on the sum indicates that it runs only over operators with dimension less than or equal to d, because this is the dimension of the energy-momentum tensor. We can similarly write a general renormalizable action as a sum over all such terms

$$S = {\sum_i}' g^i \int d^d x\, \mathscr{A}_i(\sigma)\,. \tag{15.8.8}$$

Here g^i is a general notation that includes the interactions as well as the masses and the kinetic term normalizations. The expansions (15.8.7) and (15.8.8) can be used to rewrite the scale transformation (15.8.5) as the *renormalization group equation*,

$$\epsilon^{-1}\delta_s\left\langle\prod_m \mathscr{A}_{i_m}(\sigma_m)\right\rangle = -{\sum_i}'\beta^i(g)\frac{\partial}{\partial g^i}\left\langle\prod_m \mathscr{A}_{i_m}(\sigma_m)\right\rangle - \sum_n \Delta_{i_n}{}^j\left\langle \mathscr{A}_j(\sigma_n)\prod_{m\neq n}\mathscr{A}_{i_m}(\sigma_m)\right\rangle. \tag{15.8.9}$$

There may also be contact terms between $T^a{}_a$ and the other operators, and terms from the g_i-derivative acting on the local operators. These are dependent on definitions (the choice of renormalization scheme) and can all be absorbed into the definition of $\Delta_i{}^j$. Eq. (15.8.9) states that a scale

transformation is equivalent to a change in the coupling plus a mixing of operators. As one looks at longer distances the couplings and operators flow.

The Zamolodchikov c-theorem.

Without conformal invariance, T_{zz} is not holomorphic, its modes do not generate a Virasoro algebra, and the central charge c is not defined. Nevertheless, c has a useful extension to the space of all two-dimensional field theories.

Define

$$F(r^2) = z^4 \langle T_{zz}(z,\bar{z})T_{zz}(0,0)\rangle \,, \tag{15.8.10a}$$

$$G(r^2) = 4z^3\bar{z} \langle T_{zz}(z,\bar{z})T_{z\bar{z}}(0,0)\rangle \,, \tag{15.8.10b}$$

$$H(r^2) = 16z^2\bar{z}^2 \langle T_{z\bar{z}}(z,\bar{z})T_{z\bar{z}}(0,0)\rangle \,. \tag{15.8.10c}$$

Rotational invariance implies that these depend only on $r^2 = z\bar{z}$, as indicated. From conservation, $\bar{\partial}T_{zz} + \partial T_{z\bar{z}} = 0$, one finds that

$$4\dot{F} + \dot{G} - 3G = 0 \,, \quad 4\dot{G} - 4G + \dot{H} - 2H = 0 \,, \tag{15.8.11}$$

where a dot denotes differentiation with respect to r^2. The *Zamolodchikov C function* is the combination

$$C = 2F - G - \frac{3}{8}H \,. \tag{15.8.12}$$

This has the property

$$\dot{C} = -\frac{3}{4}H. \tag{15.8.13}$$

In a unitary theory H can be written as a sum of absolute squares by inserting a complete set of states, and so is nonnegative. The result (15.8.13) shows that the physics changes in a monotonic way as we look at longer and longer distances. Also, C is stationary if and only if the two-point function of $T_{z\bar{z}}$ with itself is zero, implying (by a general result in unitary quantum field theory) that $T_{z\bar{z}}$ itself vanishes identically. The theory is then conformally invariant and C becomes precisely c.

The monotonicity property also implies that the theory at long distance will approach a stationary point of C and therefore a CFT. Again, this is intuitively plausible: at long distances the theory should forget about underlying distance scales. In general this is likely to happen in the trivial sense that all fields are massive and only the empty $c = 0$ theory remains. However, if massless degrees of freedom are present due to some combination of symmetry and the tuning of parameters, the c-theorem implies that their interactions will be conformally invariant. We should

emphasize that the unitarity and compactness are playing a role; in the more general case there do exist counterexamples (exercise 15.12).

Like c, the C function seems to represent some generalized measure of the density of states. The monotonicity is then very plausible: a massive field would contribute to the number of degrees of freedom measured at short distance, but drop out at distances long compared to its Compton wavelength. In spite of this intuitive interpretation, there seems to be no simple generalization of the C function to $d > 2$. However, the principle that the long distance limit of any quantum field theory is conformally invariant still seems to hold under broad conditions.

Conformal perturbation theory

Now let us consider adding small conformally-noninvariant terms to the action of a CFT,

$$S = S_0 + \lambda^i \int d^2z\, \mathcal{O}_i \,, \tag{15.8.14}$$

where S_0 is the action of the CFT. For convenience we focus on the case that the perturbations are primary fields, but the results are easily generalized. The λ^i are the earlier couplings g^i minus the value at the conformal point.

The main question is how the physics in the perturbed theory depends on scale. Consider the following operator product, which arises in first order perturbation theory for correlations of the energy-momentum tensor:

$$-T_{zz}(z,\bar{z})\,\lambda^i \int d^2w\, \mathcal{O}_i(w,\bar{w}) \,. \tag{15.8.15}$$

We have

$$\begin{aligned}
&\partial_{\bar{z}} T_{zz}(z)\mathcal{O}_i(w,\bar{w}) \\
&\quad = \partial_{\bar{z}}\Big[(z-w)^{-2}h_i + (z-w)^{-1}\partial_w\Big]\mathcal{O}_i(w,\bar{w}) \\
&\quad = -2\pi h_i \partial_z \delta^2(z-w)\mathcal{O}_i(w,\bar{w}) + 2\pi\delta^2(z-w)\partial_w\mathcal{O}_i(w,\bar{w}) \,.
\end{aligned} \tag{15.8.16}$$

Integrating this, the first order perturbation (15.8.15) implies that perturbation leads to

$$\partial_{\bar{z}} T_{zz}(z,\bar{z}) = 2\pi\lambda_i(h_i - 1)\partial_z \mathcal{O}_i(z,\bar{z}) \,. \tag{15.8.17}$$

As expected, the energy-momentum tensor is no longer holomorphic, unless the perturbation is of weight $h_i = 1$. The energy-momentum tensor must still be conserved,

$$\partial_{\bar{z}} T_{zz} + \partial_z T_{\bar{z}z} = 0 \,. \tag{15.8.18}$$

Inspection of the divergence (15.8.17) thus identifies

$$T_{\bar{z}z} = 2\pi\lambda^i(1-h_i)\mathcal{O}_i(z,\bar{z}) \,. \tag{15.8.19}$$

We assume that the perturbations are rotationally invariant, $h_i = \tilde{h}_i$, so that T_{ab} remains symmetric.

Referring back to the renormalization group, we have

$$\beta^i = 2(h_i - 1)\lambda^i \,, \tag{15.8.20}$$

so that a rescaling of lengths by ϵ is equivalent to a rescaling of the couplings,

$$\delta\lambda^i = 2\epsilon(1-h_i)\lambda^i \,. \tag{15.8.21}$$

A perturbation with $h_i > 1$ is thus termed *irrelevant*, because its effect drops away at long distance and we return to the conformal theory. A perturbation with $h_i < 1$ is termed *relevant*. It grows more important at low energies, and we move further from the original conformal theory. A perturbation with $h_i = 1$ is termed *marginal*.

Now let us go to the next order in g. Consider first the case that the perturbations $\mathcal{O}_i$ are all of weight $(1,1)$, marginal operators. Second order perturbation theory will then involve the operator product

$$\frac{1}{2}\int d^2z\, \mathcal{O}_i(z,\bar{z}) \int d^2w\, \mathcal{O}_j(w,\bar{w}) \,, \tag{15.8.22}$$

the factor of $\frac{1}{2}$ coming from the expansion of $\exp(-S)$. The part of the OPE that involves only marginal operators is

$$\mathcal{O}_i(z,\bar{z})\mathcal{O}_j(w,\bar{w}) \sim \frac{1}{|z-w|^2} c^k{}_{ij}\mathcal{O}_k(w,\bar{w}) \,, \tag{15.8.23}$$

so the second order term (15.8.22) will have a logarithmic divergence when $z \to w$,

$$2\pi \int \frac{dr}{r} c^k{}_{ij} \int d^2w\, \mathcal{O}_k(w,\bar{w}) \,. \tag{15.8.24}$$

The divergence must be cut off at the lower end, introducing a scale into the problem and breaking conformal invariance. At the upper end, the scale is set by the distance at which we are probing the system. We can read off immediately the scale dependence: if we increase the scale of measurement by a factor $1+\epsilon$, the log increases by ϵ. This is equivalent to shifting the couplings by

$$\delta\lambda^k = -2\pi\epsilon c^k{}_{ij}\lambda^i\lambda^j \,. \tag{15.8.25}$$

In other words,

$$\beta^k = 2\pi c^k{}_{ij}\lambda^i\lambda^j \,. \tag{15.8.26}$$

As an application, suppose that we are interested in perturbations that preserve conformal invariance. We have the familiar necessary condition that the perturbation be a (1,1) tensor, but now we see that there are further conditions: conformal invariance will be violated to second order in λ unless

$$c^k{}_{ij}\lambda^i\lambda^j = 0 \tag{15.8.27}$$

for all (1,1) operators k.

Now we wish to go to second order in λ for perturbations that are not marginal. At weak coupling, the order λ^2 term is important only if the first order term is small — that is, if the coupling is nearly marginal. To leading order in $h_i - 1$, we can just carry over our result for $O(\lambda^2)$ in the marginal case. Combining the contributions (15.8.20) and (15.8.26), we then have

$$\beta^i = 2(h_i-1)\lambda^i + 2\pi c^k{}_{ij}\lambda^i\lambda^j \,, \tag{15.8.28}$$

with corrections being higher order in $h_i - 1$ or λ^i. Let us also work out the C function. With $T_{\bar{z}z} = -\pi\beta^i\mathcal{O}_i$, the result (15.8.13) for the C function becomes to leading order

$$\dot{C} = -12\pi^2\beta^i\beta^j G_{ij} \,, \tag{15.8.29}$$

where

$$G_{ij} = z^2\bar{z}^2 \langle\, \mathcal{O}_i(z,\bar{z})\mathcal{O}_j(0,0) \,\rangle \tag{15.8.30}$$

is evaluated at $\lambda^i = 0$. Observe that

$$\beta^i = \frac{\partial}{\partial\lambda_i} U(g) \,, \tag{15.8.31a}$$

$$U(g) = (h_i - 1)\lambda^i\lambda_i + \frac{2\pi}{3} c_{ijk}\lambda^i\lambda^j\lambda^k \,, \tag{15.8.31b}$$

indices being lowered with G_{ij}. Using this and $\beta^i = -2\dot{\lambda}^i$ gives

$$\dot{C} = 24\pi^2\beta_j\dot{\lambda}^j = 24\pi^2\dot{U} \,. \tag{15.8.32}$$

This integrates to

$$C = c + 24\pi^2 U \tag{15.8.33}$$

with c being the central charge at the conformal point $\lambda^i = 0$.

Now let us apply this to the case of a single slightly relevant operator,

$$\dot{\lambda} = (1-h)\lambda - \pi c_{111}\lambda^2, \tag{15.8.34}$$

normalized so that $G_{11} = 1$. If λ starts out positive it grows, but not indefinitely: the negative second order term cuts off the growth. At long

distance we arrive at a *new* conformal theory, with coupling

$$\lambda' = \frac{1-h}{\pi c_{111}} . \tag{15.8.35}$$

From the string spacetime point of view, we can interpret $U(\lambda)$ as a potential energy for the light field corresponding to the world-sheet coupling λ, and the two conformal theories correspond to the two stationary points of the cubic potential. Note that $\lambda = 0$ is a local maximum: relevant operators on the world-sheet correspond to tachyons in spacetime. The central charge of the new fixed point is

$$c' = c - 8\frac{(1-h)^3}{c_{111}^2} . \tag{15.8.36}$$

15.9 Statistical mechanics

The partition function in classical statistical mechanics is

$$Z = \int [dq] \exp(-\beta H) , \tag{15.9.1}$$

where the integral runs over configuration space, β is the inverse temperature, and the Hamiltonian H is the integral of a local density. This has a strong formal similarity to the path integral for Euclidean quantum theory,

$$Z = \int [d\phi] \exp(-S/\hbar) . \tag{15.9.2}$$

In the statistical mechanical case, the configuration is a function of the spatial dimensions only, so that statistical mechanics in *d spatial* dimensions resembles quantum field theory in *d spacetime dimensions*. An obvious difference between the two situations is that in the statistical mechanical case there is generally an underlying discrete structure, while in relativistic field theory and on the string world-sheet we are generally interested in a continuous manifold.

There is a context in statistical mechanics in which one essentially takes the continuum limit. This is in *critical phenomena*, in which some degrees of freedom have correlation lengths very long compared to the atomic scale, and the discrete structure is no longer seen. In this case, the statistical ensemble is essentially identical to a relativistic field theory. Let us discuss the classic example, the Ising model. Here one has an array of spins on a square lattice in two dimensions, each spin σ_i taking the values ± 1. The energy is

$$H = -\sum_{\text{links}} \sigma_i \sigma_{i'} . \tag{15.9.3}$$

The sum runs over all nearest-neighbor pairs (links). The energy favors adjacent pairs being aligned. When β is small, so that the temperature is large, the correlations between spins are weak and short-range,

$$\langle \sigma_i \sigma_j \rangle \sim \exp[-|i-j|/\xi(\beta)] \tag{15.9.4}$$

as the distance $|i-j|$ goes to infinity. For sufficiently large β the $\mathbf{Z}_2$ symmetry $\sigma_i \to -\sigma_i$ is broken and there is long-range order,

$$\langle \sigma_i \sigma_j \rangle \sim v^2(\beta) + \exp[-|i-j|/\xi'(\beta)] . \tag{15.9.5}$$

For both small and large β the fluctuations are short-range. However, the transition between these behaviors is second order, both $\xi(\beta)$ and $\xi'(\beta)$ going to infinity at the critical value β_c. At the critical point the falloff is power law rather than exponential,

$$\langle \sigma_i \sigma_j \rangle \sim |i-j|^{-\eta} , \quad \beta = \beta_c . \tag{15.9.6}$$

The long-wavelength fluctuations at this point should be described by a continuum path integral. The value of the *critical exponent* η is known from the exact solution of the Ising model to be $\frac{1}{4}$. This cannot be deduced from any classical reasoning, but depends in an essential way on the nonlinear interactions between the fluctuations.

To deduce the CFT describing the critical theory, note the global symmetry of the Ising model, the $\mathbf{Z}_2$ symmetry $\sigma_i \to -\sigma_i$. We have a whole family of CFTs with this symmetry, the minimal models. For reasons to be explained below, the correct minimal model is the first nontrivial one, $m = 3$ with $c = \frac{1}{2}$. The nontrivial primary fields of this theory, taking into account the identification (15.3.15), are

$$\mathcal{O}_{1,1} : h = 0 , \quad \mathcal{O}_{1,2} : h = \frac{1}{16} , \quad \mathcal{O}_{1,3} : h = \frac{1}{2} . \tag{15.9.7}$$

Under the $\mathbf{Z}_2$ (15.3.23), $\mathcal{O}_{1,2}$ is odd and the other two are even. In particular, the Ising spins, being odd under $\mathbf{Z}_2$, should evidently be identified as

$$\sigma_i \to \sigma(z,\bar{z}) = \mathcal{O}_{1,2}(z)\tilde{\mathcal{O}}_{1,2}(\bar{z}) . \tag{15.9.8}$$

The left- and right-moving factors must be the same to give a rotationally invariant operator. There are separate $\mathbf{Z}_2$s acting on the left- and right-moving theories, but all operators have equal left and right charges so we can take either one. The expectation value

$$\langle \sigma(z,\bar{z})\sigma(0,0) \rangle \propto (z\bar{z})^{-2h} = (z\bar{z})^{-1/8} \tag{15.9.9}$$

agrees with the exact solution for the critical exponent η.

The $m = 3$ minimal model is equivalent to the free massless Majorana fermion. Indeed, Onsager solved the Ising model by showing that it could be rewritten in terms of a free fermion on a lattice, which in general is massive but which becomes massless at β_c. Note that $\mathcal{O}_{1,3}$ has the correct

dimension to be identified with the fermion field, and $\mathcal{O}_{1,2}$ has the correct dimension to be the R sector ground state vertex operator for a single Majorana fermion.

Incidentally, the solubility of the Ising model for general β can be understood directly from the CFT. Changing the temperature is equivalent to adding

$$\mathcal{O}_{1,3}(z)\tilde{\mathcal{O}}_{1,3}(\bar{z}) \tag{15.9.10}$$

to the action. This is the only relevant perturbation that is invariant under the $\mathbf{Z}_2$ symmetry. This perturbation breaks the conformal invariance, but it can be shown from the OPEs of the CFT that a spin-4 current constructed from T_{zz}^2 is still conserved. The existence of a symmetry of spin greater than 2 in a *massive* theory is sufficient to allow a complete solution. Of course, in the present case the perturbation (15.9.10) is just a mass for the free fermion, but for other CFTs without such a simple Lagrangian description this more abstract approach is needed.

The requirement that operators have integer spin means that we can only pair the same conformal family on the right and left. For the theory quantized on the circle, this corresponds to the A modular invariant discussed earlier,

$$[\mathcal{O}_{1,1}\tilde{\mathcal{O}}_{1,1}] + [\mathcal{O}_{1,2}\tilde{\mathcal{O}}_{1,2}] + [\mathcal{O}_{1,3}\tilde{\mathcal{O}}_{1,3}] \ . \tag{15.9.11}$$

In terms of the free fermion theory this is the diagonal GSO projection.

For two-dimensional critical theories with few enough degrees of freedom that the central charge is less than one, the classification of unitary representations of the Virasoro algebra completely determines the possible critical exponents: they must be given by one of the minimal models.[6] For this reason this same set of CFTs arises from many different short-distance theories. Let us mention one such context, which illustrates the relation among all the unitary minimal models through the $\mathbf{Z}_2$ symmetry they share. We noted that the $m = 3$ theory has only one relevant perturbation that is invariant under $\mathbf{Z}_2$. We therefore identified this with a variation of the temperature away from the critical point. The operator $\mathcal{O}_{1,1}\tilde{\mathcal{O}}_{1,1}$ is just the identity and adding it to the action has a trivial effect. The operator $\mathcal{O}_{1,2}\tilde{\mathcal{O}}_{1,2}$ is odd under $\mathbf{Z}_2$ and corresponds to turning on a magnetic field that breaks the $\sigma_i \to -\sigma_i$ symmetry. For the minimal model at general m there are $m-2$ nontrivial relevant $\mathbf{Z}_2$-invariant operators. This corresponds to *multicritical behavior*. To reach such a model one must tune $m-2$ parameters precisely.

[6] There is a caveat: the CFTs that arise in statistical physics need not be unitary. Unitarity in that context is related to a property known as reflection positivity, which holds in most but not all systems of interest.

For example, take the Ising model with thermally equilibrated (annealed) vacancies, so that each spin σ can take values ± 1 or 0, the last corresponding to an empty site. When the density ρ of vacancies is small, the behavior is much like the Ising model, with the same critical behavior at some point $\beta_c(\rho)$. However, when the vacancy density reaches a critical value ρ_c, then at $\beta_c(\rho_c)$ there are independent long-range fluctuations of the spin and *density*. This is known as the *tricritical Ising model*, tricritical referring to the need to adjust two parameters to reach the critical point. Since there are more long-range degrees of freedom than in the Ising model, we might expect the critical theory to have a greater central charge. The tricritical Ising model has been identified with the next minimal model, $m = 4$ with $c = \frac{7}{10}$. This generalizes: with spins (also called 'heights') taking $m-1$ values, there is a multicritical point obtained by adjusting $m-2$ parameters which is described by the corresponding minimal model. In fact, every CFT we have described in this chapter can be obtained as the critical limit of a lattice theory, and indeed of a solvable lattice theory. It is quite likely that every rational theory can be obtained from a solvable lattice theory.

A different generalization of the Ising model is the $\mathbf{Z}_k$ Ising model (the clock model). Here the spins take k values $\sigma_i = \exp(2\pi i n/k)$ for $n = 0, 1, \ldots, k-1$, and there is a $\mathbf{Z}_k$ symmetry $\sigma_i \to \exp(2\pi i/k)\sigma_i$. The energy is

$$H = -\sum_{\text{links}} \mathrm{Re}(\sigma_i \sigma_j^*) \ . \tag{15.9.12}$$

Again there is a critical point at a value β_c. The critical behavior is described by the $\mathbf{Z}_k$ parafermion theory. The $\mathbf{Z}_k$ parafermions describe a generic critical system in which the fluctuations transform under a $\mathbf{Z}_k$ symmetry.

Several of the low-lying minimal models can be realized in different ways. The $m = 5$ theory is obtained as a four-height $\mathbf{Z}_2$ model or a $\mathbf{Z}_3$ Ising model. It is also known as the three-state Potts model, referring to a different generalization of the Ising model (spins taking k values with a permutation symmetry S_k) which happens to be the same as the $\mathbf{Z}_k$ generalization when $k = 3$. The $m = 6$ model can be obtained as a five-height $\mathbf{Z}_2$ model or as a tricritical point of the $\mathbf{Z}_3$ Potts/Ising model with vacancies. In fact the $m = 3, 4, 5, 6$ theories have all been realized experimentally, usually in systems of atoms adsorbed on surfaces. Since the $m = 4$ model is also the $m = 3$ minimal model of the $N = 1$ supersymmetric series, this is in a sense the first experimental realization of supersymmetry. (Some atomic and nuclear systems have an approximate Fermi/Bose symmetry, but this is a nonrelativistic algebra whose closure does not involve the translations.)

Landau–Ginzburg models

To complete this section, we will give a slightly different Lagrangian description of the minimal models. To study the long-wavelength behavior of the Ising model, we can integrate out the individual spins and work with a field $\phi(z,\bar{z})$ representing the average spin over a region of many sites. This field takes essentially continuous values, rather than the original discrete ones. The first few terms in the Lagrangian density for ϕ would be

$$\mathscr{L} = \partial\phi\bar{\partial}\phi + \lambda_1\phi^2 + \lambda_2\phi^4 \,. \tag{15.9.13}$$

At $\lambda_1 = 0$ the tree-level mass of the field ϕ is zero. We thus identify λ_1 as being proportional to $\beta_c - \beta$, with $\lambda_1 = 0$ being the critical theory, the $m = 3$ minimal model.

This is the *Landau–Ginzburg* description. The original idea was that the classical potential for ϕ represented the free energy of the system. Now one thinks of this as the effective Lagrangian density for a full quantum (or thermal) path integral. The quantum or thermal fluctuations cannot be neglected. In some systems, though not here, they change the transition from continuous to discontinuous, so that there is no critical behavior. In general they significantly modify the scaling properties (critical exponents).

Now add a $\lambda_3\phi^6$ term and tune λ_1 and λ_2 to zero. We might expect a different critical behavior — the potential is flatter than before, so will have more states below a given energy, but it is still positive so there will be fewer states than for a free scalar. In other words, we guess that c is more than $\frac{1}{2}$ and less than 1. It is natural to identify this with the next minimal model, the $m = 4$ tricritical Ising model, since the number of relevant $\mathbf{Z}_2$-invariant perturbations is two. Similarly, we guess that the Landau–Ginzburg model whose leading potential is ϕ^{2m-2} represents the mth minimal model.

Representing the minimal models by a strongly interacting quantum field theory seems to have little quantitative value, but it gives an intuitive picture of the operator content. To start we guess that ϕ corresponds to the operator of lowest dimension, namely $\mathcal{O}_{2,2}$. Also, we guess that we have the diagonal theory, so the left-moving representation is the same as the right-moving one, and we indicate only the latter. Now, to find ϕ^2, use the fusion rule

$$\mathcal{O}_{2,2}\mathcal{O}_{2,2} = [\mathcal{O}_{1,1}] + [\mathcal{O}_{3,1}] + [\mathcal{O}_{3,3}] + [\mathcal{O}_{1,3}] \,. \tag{15.9.14}$$

The first term is the identity; we guess that ϕ^2 is the remaining operator of lowest dimension, namely $\mathcal{O}_{3,3}$. Taking further products with $\mathcal{O}_{2,2}$, we identify

$$\phi^n = \mathcal{O}_{n+1,n+1} \,, \quad 0 \le n \le m-2 \,. \tag{15.9.15}$$

This terminates due to the upper bound (15.3.16), $r \le m-1$. The lowest term in $\phi \cdot \phi^{m-2}$ is $\mathcal{O}_{m,m-2}$ which by reflection is $\mathcal{O}_{1,2}$. We then continue

$$\phi^{m-1+n} = \mathcal{O}_{n+1,n+2} \,, \quad 0 \le n \le m-3 \,. \tag{15.9.16}$$

All this guesswork can be checked in various ways. One check is that the $\mathbf{Z}_2$ symmetry assignment (15.3.23), namely $(-1)^s$ for m odd and $(-1)^r$ for m even, matches that of ϕ^n. As another check, where is the next monomial ϕ^{2m-3}? The product $\phi \cdot \phi^{2m-4}$ leads to no new primaries. This is just right: the equation of motion is

$$m\lambda_m\phi^{2m-3} = \partial\bar{\partial}\phi = L_{-1}\tilde{L}_{-1}\cdot\phi \,, \tag{15.9.17}$$

so this operator is a descendant. The powers (15.9.15) and (15.9.16) are all the relevant primary operators.

What happens if we add a relevant perturbation to the Lagrangian for the mth minimal model? The Landau–Ginzburg picture indicates that adding ϕ^{2k-2} causes the theory to flow to the kth minimal model. Let us consider in particular ϕ^{2m-4} for m large. This is

$$\mathcal{O}_{m-1,m-2} = \mathcal{O}_{1,3} \,, \quad h = 1 - \frac{2}{m+1} \,, \tag{15.9.18}$$

which is nearly marginal. Thus we can apply the formalism of the previous section. From the fusion rule

$$\mathcal{O}_{1,3}\mathcal{O}_{1,3} = [\mathcal{O}_{1,1}] + [\mathcal{O}_{1,3}] + [\mathcal{O}_{1,5}] \,, \tag{15.9.19}$$

the only nearly marginal operator in $\mathcal{O}_{1,3}\mathcal{O}_{1,3}$ is $\mathcal{O}_{1,3}$ itself, so we are in precisely the single-operator situation worked out in the last paragraph of the previous section. Thus, we can construct a new conformal theory by a small $\mathcal{O}_{1,3}$ perturbation of the minimal model. The Landau–Ginzburg picture indicates that this is the next minimal model down. We can compute the central charge from the c-theorem. Taking from the literature the value $c_{111} = 4/3^{1/2}$ for the large-m minimal model yields

$$c' = c - \frac{12}{m^3} \,. \tag{15.9.20}$$

For large m this is indeed the difference between the central charges of successive minimal models.

Exercises

15.1 Evaluate $\det(\mathscr{M}^3)$ and compare with the Kac formula.

15.2 Derive eqs. (15.2.3) and (15.2.5) for the expectation value of a descendant.

15.3 Work out the steps outlined in the derivation of eq. (15.2.9) to find explicitly the $N = 0$ and $N = 1$ terms in $\mathscr{F}^{jl}_{mn}(i|z)$.

15.4 Verify that the discrete symmetries associated with the simple currents are as asserted below eqs. (15.3.23) and (15.4.18) for the unitary minimal models and the $SU(2)$ WZNW models.

15.5 (a) For the $SU(n)$ current algebra at level k, consider the four-point function with two insertions in the representation $(\mathbf{n}, \mathbf{n})$ and two in the representation $(\bar{\mathbf{n}}, \bar{\mathbf{n}})$. Find the KZ equation for the $SU(n)$ invariants.
(b) Find the general solution for $k = 1$ and determine the coefficients using associativity and locality. Compare this with the free-boson representation.
(c) Do the same for general k; the solution involves hypergeometric functions.

15.6 The *Wakimoto representation* is a free-field representation for the $SU(2)$ current algebra, analogous to the Feigin–Fuchs representation of the minimal models. Show that the following currents form an $SU(2)$ current algebra of level $k = q^2 - 2$:

$$J^+ = iw/2^{1/2}\,, \quad J^3 = iq\partial\phi/2^{1/2} - w\chi\,,$$
$$J^- = i[w\chi^2 + (2-q^2)\partial\chi]/2^{1/2} + q\chi\partial\phi\,.$$

Here w, χ are a commuting $\beta\gamma$ system and ϕ is a free scalar. Show that the Sugawara energy momentum tensor corresponds to the $\beta\gamma$ theory with $h_w = 1$ and $h_\chi = 0$, and with ϕ being a linear dilaton theory of appropriate central charge.

15.7 For the coset construction of the minimal models, combine primary fields from the two factors in G to form irreducible representations of $SU(2)$. Subtract the weight of the corresponding primary of H and show that the resulting weight is one of the allowed weights for the minimal model. Not all minimal model primaries are obtained in this way; some are excited states in the current algebras.

15.8 Repeat the previous exercise for the coset construction of the minimal $N = 1$ superconformal theories.

15.9 For the periodic scalar at any radius, the analysis in section 15.2 shows that the spectrum contains an infinite number of conformal families. Show, however, that if R^2/α' is rational, the partition function is a sum of a finite number of factors, each one holomorphic times antiholomorphic in τ. Show that at these radii there is an enlarged chiral algebra.

15.10 Apply the result (15.7.7) to the $SU(2)$ current algebra at $k = 4$. Show that the resulting relations are consistent with the actual weights of the $SU(2)$ primaries.

15.11 Verify the Verlinde formula (15.7.14) for the $SU(2)$ modular transformation (15.4.22). In this case indices are raised with the identity matrix.

15.12 For the general massless closed string vertex operator, we found the condition for Weyl invariance in section 3.6. Find the weaker condition for invariance under rigid Weyl transformations, and find solutions that have only this smaller invariance.

16
Orbifolds

In the final four chapters we would like to see how compactification of string theory connects with previous ideas for unifying the Standard Model. Our primary focus is the weakly coupled $E_8 \times E_8$ heterotic string, whose compactification leads most directly to physics resembling the Standard Model. At various points we consider other string theories and the effects of strong coupling. In addition, compactified string theories have interesting nonperturbative dynamics, beyond that which we have seen in ten dimensions. In the final chapter we discuss some of the most interesting phenomena.

The two main issues are specific constructions of four-dimensional string theories and general results derived from world-sheet and spacetime symmetries. Our approach to the constructions will generally be to present only the simplest examples of each type, in order to illustrate the characteristic physics of compactified string theories. On the other hand, we have collected as many of the general results as possible.

String compactifications fall into two general categories. The first are based on free world-sheet CFTs, or on CFTs like the minimal models that are solvable though not free. For these one can generally determine the exact tree-level spectrum and interactions. The second category is compactification in the geometric sense, taking the string to propagate on a smooth spacetime manifold some of whose dimensions are compact. In general one is limited to an expansion in powers of α'/R_c^2, with R_c being the characteristic radius of compactification. This is in addition to the usual expansion in the string coupling g. Commonly in a moduli space of smooth compactifications there will be special points (or subspaces) described by free CFTs. Thus the two approaches are complementary, one giving a very detailed picture at special points and the other giving a less detailed but global picture. Some of the solvable compactifications have no such geometric interpretation.

In this chapter we discuss free CFTs and in the next geometric compactification. Again, the literature in each case is quite large and a full account is far beyond the scope of this book.

16.1 Orbifolds of the heterotic string

In section 8.5 we discussed orbifolds, manifolds obtained from flat spacetime by identifying points under a discrete group H of symmetries. Although these manifolds generally have singularities, the resulting string theories are well behaved. The effect of the identification is to add twisted closed strings to the Hilbert space and to project onto invariant states.

We start with the ten-dimensional $E_8 \times E_8$ string, with H a subgroup of the Poincaré $\times$ gauge group. An element of H will act on the coordinates as a rotation θ and translation v,

$$X^m \to \theta^{mn} X^n + v^n \,, \tag{16.1.1}$$

where $m, n = 4, \ldots, 9$. For a four-dimensional theory H will act trivially on X^μ for $\mu = 0, \ldots, 3$. In order to preserve world-sheet supersymmetry the twist must commute with the supercurrent, and so its action on the right-moving fermions is

$$\tilde{\psi}^m \to \theta^{mn} \tilde{\psi}^n \,. \tag{16.1.2}$$

In addition it acts on the current algebra fermions as a gauge rotation γ,

$$\lambda^A \to \gamma^{AB} \lambda^B \,. \tag{16.1.3}$$

Here we are considering gauge rotations γ^{AB} which are in the manifest $SO(16) \times SO(16)$ subgroup of $E_8 \times E_8$. The full element is denoted $(\theta, v; \gamma)$. Just as the fixed points can be thought of as points of singular spacetime curvature, a nontrivial γ can be thought of as singular gauge curvature at the fixed points.

Ignoring the gauge rotation, the set of all elements (θ, v) forms the *space group* S. In the twisted theory the strings are propagating on the space $M^4 \times K$, where

$$K = R^6/S \,. \tag{16.1.4}$$

Because the elements of S in general have fixed points, this space is an orbifold.

Ignoring the translation as well as the gauge rotation leaves the *point group* P, the set of all rotations θ appearing in the elements of the twist group. An orbifold is called Abelian or non-Abelian according to whether the point group is Abelian or non-Abelian.

The subgroup of S consisting of pure translations $(1, v)$ is an Abelian group Λ. An alternative description of the orbifold is to twist first by Λ

to form a particular 6-torus,

$$T^6 = R^6/\Lambda \,. \tag{16.1.5}$$

The space group multiplication law

$$(\theta, w)\cdot(1, v)\cdot(\theta, w)^{-1} = (1, \theta v)\,, \tag{16.1.6}$$

implies that the group

$$\bar{P} \equiv S/\Lambda \tag{16.1.7}$$

is a symmetry of the 6-torus. This is the same as the point group P except that some elements include translations. One can now twist the torus by $\bar{P}$ to form the orbifold

$$K = T^6/\bar{P}\,. \tag{16.1.8}$$

We can assume that the identity element in spacetime appears only with the identity in the gauge group, as $e = (1, 0; 1)$. This is no loss of generality, because if there were additional elements of the form $(1, 0; \gamma)$, one could first twist on the subgroup consisting of these pure gauge twists to obtain a different ten-dimensional theory, or perhaps a different description of the same theory, and then twist this theory under the remaining group which has no pure gauge twists. By closure it follows that each element (θ, v) of the space group appears with a unique gauge element $\gamma(\theta, v)$, and that these have the multiplication law

$$\gamma(\theta_1, v_1)\gamma(\theta_2, v_2) = \gamma((\theta_1, v_1)\cdot(\theta_2, v_2))\,. \tag{16.1.9}$$

That is, there is a homomorphism from the space group to the gauge group.

Modular invariance

Modular invariance requires that the projection onto H-invariant states be accompanied by the addition of twisted states for each $h \in H$:

$$\varphi(\sigma^1 + 2\pi) = h\cdot\varphi(\sigma^1)\,, \tag{16.1.10}$$

where φ stands for a generic world-sheet field. The resulting sum over path integral sectors is naively modular-invariant. However, we know from the example of the superstring in chapter 10 that modular invariance can be spoiled by phases in the path integral. In particular, the phase under $\tau \to \tau+1$ is determined by the level mismatch, the difference $L_0 - \tilde{L}_0$ mod 1. In fact, for Abelian orbifolds it has been shown that this is the only potential obstruction to modular invariance.

To see how this works, consider the spectrum in the sector with twist h. Let N be the smallest integer such that $h^N = 1$; we then call this a

Z_N twist. We can always choose the axes so that the rotation is of the form

$$\theta = \exp[2\pi i(\phi_2 J_{45} + \phi_3 J_{67} + \phi_4 J_{89})]\,. \tag{16.1.11}$$

Define the complex linear combinations

$$Z^i = 2^{-1/2}(X^{2i} + iX^{2i+1})\,, \quad i = 2, 3, 4\,, \tag{16.1.12}$$

with

$$Z^{\bar{i}} \equiv \overline{Z^i} = 2^{-1/2}(X^{2i} - iX^{2i+1})\,. \tag{16.1.13}$$

The periodicity is then

$$Z^i(\sigma + 2\pi) = \exp(2\pi i\phi_i)Z^i(\sigma)\,. \tag{16.1.14}$$

Taking the same complex basis for the $\tilde{\psi}^m$ gives

$$\tilde{\psi}^i(\sigma + 2\pi) = \exp[2\pi i(\phi_i + v)]\tilde{\psi}^i(\sigma) \tag{16.1.15}$$

with $v = 0$ in the R sector and $v = \frac{1}{2}$ in the NS sector. The supercurrent is then periodic or antiperiodic in the usual way depending on v. The oscillators have the following mode numbers:

$$\alpha^i : n + \phi_i\,, \quad \alpha^{\bar{i}} : n - \phi_i\,, \tag{16.1.16a}$$

$$\tilde{\alpha}^i : n - \phi_i\,, \quad \tilde{\alpha}^{\bar{i}} : n + \phi_i\,, \tag{16.1.16b}$$

$$\tilde{\psi}^i : n - \phi_i\ (\mathrm{R})\,,\ n - \phi_i + \tfrac{1}{2}\ (\mathrm{NS})\,, \tag{16.1.16c}$$

$$\tilde{\psi}^{\bar{i}} : n + \phi_i\ (\mathrm{R})\,,\ n + \phi_i + \tfrac{1}{2}\ (\mathrm{NS})\,. \tag{16.1.16d}$$

For a single element, the gauge twist can always be taken in the block-diagonal $U(1)^{16}$ subgroup,

$$\gamma = \mathrm{diag}[\exp(i\beta_1), \ldots, \exp i(\beta_{16})]\,. \tag{16.1.17}$$

This acts on the complex linear combinations $\lambda^{K\pm} = 2^{-1/2}(\lambda^{2K-1} \pm i\lambda^{2K})$ as

$$\lambda^{K\pm} \to \exp(\pm 2\pi i\beta_K)\lambda^{K\pm}\,. \tag{16.1.18}$$

The oscillators $\lambda^{K\pm}$ thus have mode numbers $n \mp \beta_K$ in the R sector of the current algebra, and $n \mp \beta_K + \frac{1}{2}$ in the NS sector.

Because $h^N = 1$ we can write

$$\phi_i = \frac{r_i}{N}\,, \quad \beta_K = \frac{s_K}{N}\,, \tag{16.1.19}$$

for integers r_i and s_K. Actually, we can say a bit more, because the various R sectors are in spinor representations and so contain eigenvalues

$$\frac{1}{2}\sum_{i=2}^{4}\phi_i\,, \quad \frac{1}{2}\sum_{K=1}^{8}\beta_K\,, \quad \frac{1}{2}\sum_{K=9}^{16}\beta_K\,. \tag{16.1.20}$$

Thus we have the mod 2 conditions

$$\sum_{i=2}^{4} r_i = \sum_{K=1}^{8} s_K = \sum_{K=9}^{16} s_K = 0 \mod 2 . \tag{16.1.21}$$

To be precise, if these are not satisfied then h^N is a nontrivial twist of the ten-dimensional theory, and so just changes the starting point.

Consider first the sector (R,R,R), labeled by the periodicities of the two sets of current algebra fermions and the supercurrent. Recall the general result that a complex boson with mode numbers $n+\theta$ has zero-point energy

$$\frac{1}{24} - \frac{1}{8}(2\theta - 1)^2 , \tag{16.1.22}$$

and a complex fermion has the negative of this. The above discussion of modes then gives the level mismatch as

$$L_0 - \tilde{L}_0 = -\sum_{i=2}^{4}(N^i + \tilde{N}^i + \tilde{N}^i_{\tilde{\psi}})\phi_i - \sum_{K=1}^{16} N^K \beta_K - \frac{1}{2}\sum_{i=2}^{4}\phi_i(1-\phi_i) + \frac{1}{2}\sum_{K=1}^{16}\beta_K(1-\beta_K) \mod 1 . \tag{16.1.23}$$

Here N^i counts the number of α^i excitations minus the number of $\alpha^{\bar{I}}$ excitations, and so on.

The oscillator part of $L_0 - \tilde{L}_0$ is a multiple of $1/N$, and the zero-point part a multiple of $1/2N^2$, so that in general there are *no* states for which $L_0 - \tilde{L}_0$ is an integer. Suppose, however, that the zero-point contribution is actually a multiple of $1/N$,

$$-\frac{1}{2}\sum_{i=2}^{4}\phi_i(1-\phi_i) + \frac{1}{2}\sum_{K=1}^{16}\beta_K(1-\beta_K) = \frac{m}{N} \tag{16.1.24}$$

for integer m. Then imposing on the excitation numbers the condition

$$\sum_{i=2}^{4}(N^i + \tilde{N}^i + \tilde{N}^i_{\tilde{\psi}})\phi_i + \sum_{K=1}^{16} N^K \beta_K = \frac{m}{N} \mod 1 \tag{16.1.25}$$

leaves only states with integer $L_0 - \tilde{L}_0$. The left-hand side is just the transformation of the oscillators under h, so this condition is the projection onto h-invariant states. In particular, the phase of h in the twisted sector is determined by the zero-point energy (16.1.24).

Now consider the sector (R,R,NS). The $\tilde{\psi}$ modes are shifted by one-half,

so the level mismatch is equal to the earlier value (16.1.23) plus

$$\delta = \frac{1}{2}\left(-\tilde{N}_{\tilde{\psi}} - 1 + \sum_{i=2}^{4}\phi_i\right) , \tag{16.1.26}$$

where the first term is from the excitations and the last two are from the change in the zero-point energy. Level matching again requires that

$$\delta = k/N \tag{16.1.27}$$

for some integer k. The first two terms add to an integer due to the GSO projection, and then (16.1.27) follows from the mod 2 conditions (16.1.21). Level matching in all other sectors follows in the same way from conditions (16.1.21) and (16.1.24). The latter can also be rephrased

$$\sum_{i=2}^{4} r_i^2 - \sum_{K=1}^{16} s_K^2 = 0 \mod 2N. \tag{16.1.28}$$

For Abelian orbifolds, as long as there are any states for which the level mismatch (16.1.23) is an integer, then by imposing the projection (16.1.25) one obtains a consistent theory. For non-Abelian orbifolds there are additional conditions.

Other free CFTs

The orbifolds above can be thought of as arising from the ten-dimensional theory in one step, twisting by the full space group, or in two, twisting first by the translations to make a toroidal theory and then twisting by the point group. The second construction can be made more general as follows. Represent the current algebra in bosonic form, so the toroidal theory has a momentum lattice of signature (22,6). Many lattices have symmetries that rotate the left and the right momenta independently, as opposed to the above construction in which $\theta_L = \theta_R$ on the (6,6) spacetime momenta. These more general theories are known as *asymmetric orbifolds*. Though there is no longer a geometric interpretation in terms of propagation on a singular space, the construct is consistent in CFT and in string theory.

Another construction is to fermionize all the internal coordinates, giving 44 left-movers and 18 right-movers. Since the Lorentz invariance is broken one can take arbitrary combinations of independent R and NS boundary conditions on the 62 fermions, subject to the constraints of modular invariance, locality of the OPE, and so on. Alternatively, join the real fermions into $22+9$ complex fermions and take sectors with independent aperiodicities $\exp(2\pi i v)$ for each fermion. In spite of appearances this is not strictly more general, because in the first case one can have combinations of boundary conditions such that the fermions cannot be put into pairs

having the same boundary conditions in all sectors. The ten-dimensional E_8 theory from section 11.3 is an example with such essentially real fermions. One can also take some fermions of each type. The general consistent theory is known.

The supercurrent $\tilde{T}_F$ is now written purely in terms of fermions. For example, a single $X\tilde{\psi}$ CFT becomes a theory of three fermions with $\tilde{T}_F = i\tilde{\chi}_1\tilde{\chi}_2\tilde{\chi}_3$. The boundary conditions must be correlated so that all terms in the supercurrent are simultaneously R or NS. It is interesting to ask what is the most general $\tilde{T}_F$ that can be constructed from free fermions alone. A general $(0, \frac{3}{2})$ tensor would be

$$\tilde{T}_F = i \sum_{I,J,K=1}^{18} \tilde{\chi}_I \tilde{\chi}_J \tilde{\chi}_K c_{IJK} \ . \tag{16.1.29}$$

The conditions for the $\tilde{T}_F \tilde{T}_F$ OPE to generate a superconformal algebra are easily solved. The requirement that there be no four-fermi term in the OPE is

$$c_{IJM}c_{KLM} + c_{JKM}c_{ILM} + c_{KIM}c_{JLM} = 0 \ . \tag{16.1.30}$$

This is the Jacobi identity, requiring c_{IJK} to be the structure constants of a Lie algebra. The condition that the $\bar{z}^{-1}$ term in the OPE be precisely $2\tilde{T}_B$ is then

$$18 c_{IKL}c_{JKL} = \delta_{IJ} . \tag{16.1.31}$$

This fixes the normalization of c_{IJK}, and requires the algebra to be semisimple (no Abelian factors). The dimension of the group is the number of fermions, 18. There are three semisimple groups of dimension 18, namely $SU(2)^6$, $SU(3) \times SO(5)$, and $SU(4) \times SU(2)$.

Another construction is to bosonize all fermions including the $\tilde{\psi}^\mu$ to form a lattice of signature (22,9), and then to make a Narain-like construction. Again $\tilde{T}_F$ can be generalized, to a sum of terms of the form

$$e^{ik\cdot X_R} \ , \ k^2 = 6/\alpha' \ ; \quad e^{il\cdot X_R} \bar{\partial} X_R \ , \ l^2 = 2/\alpha' \ . \tag{16.1.32}$$

Obviously there are overlaps among these constructions, though often one or the other description is more convenient. The fermionic construction in particular has been employed by a number of groups. We will be able to see a great deal of interesting spacetime physics even in the simplest orbifold models, so we will not develop these generalizations further.

16.2 Spacetime supersymmetry

We have seen that in consistent string theories there is a symmetry that relates fermions to bosons. The important question is whether in the real world this symmetry is spontaneously broken at very high energy, or whether part of it survives down to the weak interaction scale, the energy that can be reached by particle accelerators. In fact there is a strong argument, independent of string theory, for expecting that exactly one $d = 4$ supersymmetry survives and is spontaneously broken near the weak scale.

The argument has to do with the self-energies of elementary particles. The energy in the field of a charged point particle diverges at short distance. If we suppose that this is cut off physically at some distance l then naively the self-energy is

$$\delta m \approx \frac{\alpha}{l} \ , \tag{16.2.1}$$

with $\alpha = e^2/4\pi$ the fine structure constant. The electron is known to be pointlike down to at least 10^{-16} cm, implying that the energy (16.2.1) is more than 10^3 times the actual electron mass. However, it has been known since the 1930s that relativistic quantum effects reduce the simple classical estimate (16.2.1) to

$$\delta m \approx \alpha m \ln \frac{1}{ml} \ . \tag{16.2.2}$$

Taking l to be near the Planck scale, the logarithm is of order 50 and the self-energy, taking into account numerical factors, is roughly 20% of the actual mass of the electron. For quarks the effect is larger due to the larger $SU(3)$ coupling, so that the self-energy is of order the mass itself. In simple grand unified theories the bottom quark and tau lepton are in the same multiplet and have equal 'bare' masses, but the inclusion of the self-energies accounts to good accuracy for the observed ratio

$$\frac{m_b}{m_\tau} \approx 3 \ . \tag{16.2.3}$$

This is a successful test of grand unification, though less impressive than the unification of the gauge couplings because it is more model-dependent and because the ratio is not known with the same precision.

This leaves one problem in the Standard Model, the Higgs boson. This is the only scalar, and the only particle for which the estimate (16.2.1) is not reduced by relativistic quantum effects. If the Higgs boson remains pointlike up to energies near the Planck scale as in ordinary grand unified theories, then the self-energy is roughly 15 orders of magnitude larger than the actual mass. We have to suppose that the bare mass cancels this correction to an accuracy of roughly one part in 10^{30}, because it

is actually the mass-squared that adds. This seems quite unsatisfactory, especially in light of the very physical way we are able to think about the other self-energies.

One possible resolution of this *naturalness problem* is that the Higgs scalar is not pointlike but actually composite on a scale not far from the weak scale. This is the idea of technicolor theories; it has not been ruled out but has not led to convincing models. A second is that there is some other effect that cancels the self-energy. Indeed, this is the case in supersymmetric theories. The Higgs mass-squared comes from the superpotential, and as discussed in section B.2 this is not renormalized: the self-energy is canceled by a fermionic loop amplitude, at least down to the scale of supersymmetry breaking.

For this reason theories with supersymmetry broken near the weak scale have received a great deal of attention, both in particle phenomenology and in string theory. The $d = 4$ supersymmetry algebra must be $N = 1$ because the gauge-couplings in the Standard Model are chiral. As discussed in section B.2, the $N = 2$ and larger algebras do not allow this.

Supersymmetric string theories are also attractive because as we will see later supersymmetry in spacetime implies a much-enlarged symmetry on the world-sheet, and so the construction and solution of these CFTs has gone much farther than for the nonsupersymmetric theories. Also, nonsupersymmetric string theories usually, though not always, have tachyons in their spectra. Finally, the order-by-order supersymmetric cancellation of the vacuum energy means that there are no tadpole divergences and the perturbation theory is finite at each order.

It is still a logical possibility that all the supersymmetry of string theory is broken at the string scale, and even that the low energy limit of string theory is a technicolor theory. Low energy supersymmetry and string theory are independent ideas: either might be right and the other wrong. However, the discovery of low energy supersymmetry would be an encouraging sign that these ideas are in the right direction. Also, the measurement of the many new masses and couplings of the superpartners would give new windows onto higher energy physics. Given the important role that supersymmetry plays at short distance, and the phenomenological reasons for expecting supersymmetry near the weak scale, it is reasonable to hope that of all the new phenomena that accompany string theory supersymmetry will be directly visible.

What then are the conditions for an orbifold compactification to have an unbroken $N = 1$ supersymmetry? Let us consider first the case that the point group is $\mathbf{Z}_N$ so that it is generated by a single element of the form (16.1.11). This acts on the supersymmetries as

$$Q_\alpha \to D(\phi)_{\alpha\beta} Q_\beta \,, \tag{16.2.4}$$

where $D(\phi)$ is the spinor representation of the rotation. In the usual s-basis this is

$$Q_{\mathbf{s}} \to \exp(i\mathbf{s}\cdot\phi) Q_{\mathbf{s}} \,. \tag{16.2.5}$$

The (s_2, s_3, s_4) run over all combinations of $\pm\frac{1}{2}$, each combination appearing twice. Thus if

$$\phi_2 + \phi_3 + \phi_4 = 0 \tag{16.2.6}$$

with the ϕs otherwise generic, there will be four unbroken supersymmetries, namely those with $s_2 = s_3 = s_4$. Three-quarters of the original 16 supersymmetries of the heterotic string are broken. Other possibilities such as $\phi_2 + \phi_3 - \phi_4 = 0$ give equivalent physics.

Note that this discussion is quite similar to the discussion of the supersymmetry of rotated D-branes in section 13.4. As there, we can express the result in a more general way. Since the rotation takes the Z^i into linear combinations of themselves, it lies in a $U(3)$ subgroup of the $SO(6)$ rotational symmetry of the six orbifold dimensions. The condition (16.2.6) states that the rotation actually lies in $SU(3)$. Under

$$SO(9,1) \to SO(3,1) \times SO(6) \to SO(3,1) \times SU(3) \,, \tag{16.2.7}$$

the **16** decomposes as derived in section B.1,

$$\mathbf{16} \to (\mathbf{2},\mathbf{4}) + (\bar{\mathbf{2}},\bar{\mathbf{4}}) \to (\mathbf{2},\mathbf{3}) + (\mathbf{2},\mathbf{1}) + (\bar{\mathbf{2}},\bar{\mathbf{3}}) + (\bar{\mathbf{2}},\mathbf{1}) \,. \tag{16.2.8}$$

If $P \subset SU(3) \subset SO(6)$, the generators $(\mathbf{2},\mathbf{1})$ and $(\bar{\mathbf{2}},\mathbf{1})$ will survive the orbifold projection and there will be unbroken $N = 1$ supersymmetry. Similarly the stricter condition

$$\phi_2 + \phi_3 = \phi_4 = 0 \tag{16.2.9}$$

implies that

$$P \subset SU(2) \subset SU(3) \subset SO(6) \,. \tag{16.2.10}$$

In this case there will be unbroken $N = 2$ supersymmetry.

16.3 Examples

The main example we will consider is based on a $\mathbf{Z}_3$ orbifold of the torus. The lattice Λ for the $\mathbf{Z}_3$ orbifold is generated by the six translations

$$t_i : \; Z^i \to Z^i + R_i \,, \tag{16.3.1a}$$

$$u_i : \; Z^i \to Z^i + \alpha R_i \,, \quad \alpha = \exp(2\pi i/3) \tag{16.3.1b}$$

The lattice in one complex plane is shown in figure 16.1, with R_i the lattice spacing. For $R_i = \alpha'^{1/2}$ this is the root lattice of $SU(3)$, so up to rescaling

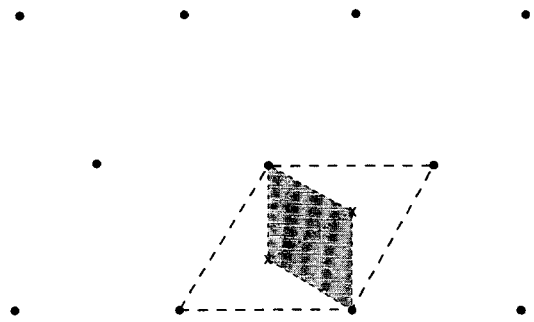


Fig. 16.1. A two-dimensional lattice invariant under rotations by $\pi/3$. A unit cell is indicated. The two points indicated by $\times$ are invariant under the combination of a $2\pi/3$ rotation and a lattice translation, as are the corner points of the unit cell. A fundamental region for the orbifold identification is shaded. One can think of the orbifold space as formed by folding the shaded region on the dotted line and identifying the edges.

of the Z^i, Λ is the root lattice of $SU(3)\times SU(3)\times SU(3)$. This is invariant under independent six-fold rotations of each $SU(3)$ lattice.

For the $\mathbf{Z}_3$ orbifold, the point group consists of a simultaneous three-fold rotation of all three lattices, the $\mathbf{Z}_3$ group $\{1, r, r^2\}$ generated by

$$r: \quad Z^2 \to \alpha Z^2\,, \quad Z^3 \to \alpha Z^3\,, \quad Z^4 \to \alpha^{-2} Z^4\,. \tag{16.3.2}$$

In the notation (16.1.11) this is

$$\phi_i = (\tfrac{1}{3}, \tfrac{1}{3}, -\tfrac{2}{3})\,, \tag{16.3.3}$$

which satisfies the mod 2 condition and leaves $N = 1$ supersymmetry unbroken.

Initially we will consider the simple case that there are no Wilson lines. That is, the translations Λ are not accompanied by gauge twists:[1] they are of the form $g = (1, v; 1)$. The gauge twist must satisfy the mod 2 and level-matching conditions. An easy way to do this is to have the gauge rotation act on the gauge fermions in exactly the same way as the spacetime rotation (16.1.11) acts on the $\tilde{\psi}$,

$$\beta_K = (\phi_2, \phi_3, \phi_4, 0^5; 0^8) = (\tfrac{1}{3}, \tfrac{1}{3}, -\tfrac{2}{3}, 0^5; 0^8)\,. \tag{16.3.4}$$

This is called *embedding the spin connection in the gauge connection.* The two terms in the level-matching condition (16.1.24) then cancel automatically. One way to think about this is to note that the nontrivial part of the world-sheet theory is parity-invariant, which allows a coordinate-invariant Pauli–Villars regulator. We will analyze the spectrum of this model, and discuss more general gauge twists later.

[1] The gauge twists accompanying transformations with fixed points are not referred to as Wilson lines, because in a sense they *do* produce a local field strength, a delta function at the fixed point.

Examining the untwisted sector first, we have to impose the $\mathbf{Z}_3$ projection on the states of the toroidal compactification. We assume that none of the R_i are $\alpha'^{1/2}$, to avoid extra massless states from the $SU(3)$ roots. The eigenvalues of $h = (r, 1; \gamma)$ are powers of α. We first classify the massless left- and right-moving states by their eigenvalues.

On the left-moving side are

$$\alpha^0: \quad \alpha^\mu_{-1}|0\rangle\,, \quad |a\rangle \in (\mathbf{8},\mathbf{1},\mathbf{1}) + (\mathbf{1},\mathbf{78},\mathbf{1}) + (\mathbf{1},\mathbf{1},\mathbf{248})\,, \tag{16.3.5a}$$

$$\alpha^1: \quad \alpha^i_{-1}|0\rangle\,, \quad |a\rangle \in (\mathbf{3},\mathbf{27},\mathbf{1})\,, \tag{16.3.5b}$$

$$\alpha^2: \quad \alpha^{\bar{\imath}}_{-1}|0\rangle\,, \quad |a\rangle \in (\bar{\mathbf{3}},\overline{\mathbf{27}},\mathbf{1})\,. \tag{16.3.5c}$$

For the states with an α_{-1} oscillator excited the eigenvalue comes from the rotation r. The states from the current algebra have been denoted by their group index $|a\rangle$, without reference to a specific (fermionic or bosonic) representation. These states have been decomposed according to their transformation under

$$SU(3)\times E_6 \times E_8 \subset E_8 \times E_8\,. \tag{16.3.6}$$

This decomposition was given in section 11.4 and a derivation outlined in exercise 11.5. The $SU(3)$ acts on the first three complex gauge fermions $\lambda^{1+,2+,3+}$. The gauge rotation (16.3.4) acts on any state as

$$\exp[2\pi i(q_1 + q_2 - 2q_3)/3]\,, \tag{16.3.7}$$

where the q_K are the eigenvalues of the state under $U(1)^{16}$. This is an element of $SU(3)$, in fact of the center of $SU(3)$, acting as α on any element of the $\mathbf{3}$ and α^2 on any element of the $\bar{\mathbf{3}}$.

On the right-moving side h acts only through the rotation r, giving

$$\alpha^0: \quad \tilde{\psi}^\mu_{-1/2}|0\rangle_{\rm NS}\,, \quad |\tfrac{1}{2},\mathbf{1}\rangle_{\rm R}\,, \quad |-\tfrac{1}{2},\bar{\mathbf{1}}\rangle_{\rm R}\,, \tag{16.3.8a}$$

$$\alpha^1: \quad \tilde{\psi}^i_{-1/2}|0\rangle_{\rm NS}\,, \quad |\tfrac{1}{2},\mathbf{3}\rangle_{\rm R}\,, \tag{16.3.8b}$$

$$\alpha^2: \quad \tilde{\psi}^{\bar{\imath}}_{-1/2}|0\rangle_{\rm NS}\,, \quad |-\tfrac{1}{2},\bar{\mathbf{3}}\rangle_{\rm R}\,. \tag{16.3.8c}$$

Here μ runs over the noncompact transverse dimensions 2, 3. We have labeled the fermionic states by their four-dimensional helicity s_1 and by their $SU(3) \subset SO(6)$ transformation. In terms of the spins (s_2, s_3, s_4) the $\mathbf{1}$, $\bar{\mathbf{3}}$, $\mathbf{3}$, and $\bar{\mathbf{1}}$ consist of states with zero, one, two, or three $-\frac{1}{2}$s respectively.

Now pair up left- and right-moving states, looking first at the bosons. In the sector $\alpha^0 \cdot \alpha^0$ are

$$\alpha^\mu_{-1}\tilde{\psi}^\nu_{-1/2}|0\rangle_{\rm NS}\,, \tag{16.3.9}$$

which are the four-dimensional graviton, dilaton, and axion, as well as

$$\tilde{\psi}^{\mu}_{-1/2}|a\rangle_{\rm NS}\ , \quad a \in (\mathbf{8},\mathbf{1},\mathbf{1}) + (\mathbf{1},\mathbf{78},\mathbf{1}) + (\mathbf{1},\mathbf{1},\mathbf{248})\ . \tag{16.3.10}$$

These are $SU(3) \times E_6 \times E_8$ gauge bosons. The gauge group is just the subgroup left invariant by the twist. In the sector $\alpha^1 \cdot \alpha^2$ there are neutral scalars of the form

$$\alpha^{i}_{-1}\tilde{\psi}^{\bar{\jmath}}_{-1/2}|0,0\rangle_{\rm NS} \tag{16.3.11}$$

and scalars

$$\tilde{\psi}^{\bar{\jmath}}_{-1/2}|a\rangle_{\rm NS}\ , \quad a \in (\mathbf{3},\mathbf{27},\mathbf{1})\ . \tag{16.3.12}$$

The sector $\alpha^2 \cdot \alpha^1$ contributes a conjugate set of states. The neutral scalars are from the internal modes of the graviton and antisymmetric tensor. In particular, the symmetric combinations are the moduli for a flat internal metric of the form

$$G_{i\bar{\jmath}}dZ^i dZ^{\bar{\jmath}}. \tag{16.3.13}$$

A metric of this form, with no $dZ^i dZ^j$ or $dZ^{\bar{\imath}}dZ^{\bar{\jmath}}$ components, is known as *Hermitean.*

The fermions are the superpartners of these. The bosonic states in each line of the right-moving spectrum (16.3.8) are replaced by the fermionic states in the same line. In the sector $\alpha^0 \cdot \alpha^0$ are the states

$$\alpha^{2\pm i3}_{-1}|s_1,\mathbf{1}\rangle_{\rm R}\ , \tag{16.3.14}$$

which are the gravitinos with helicity $\pm\frac{3}{2}$ and the dilatinos with helicity $\pm\frac{1}{2}$. The other components of the ten-dimensional gravitino are in the sectors $\alpha^0 \cdot \alpha^{1,2}$ and are removed by the projection, consistent with the earlier deduction that the theory has $N = 1$ supersymmetry. The other spinors with helicity $\frac{1}{2}$ are from the sector $\alpha^2 \cdot \alpha$:

$$|a,\tfrac{1}{2},\mathbf{3}\rangle_{\rm R}\ , \quad a \in (\bar{\mathbf{3}},\overline{\mathbf{27}},\mathbf{1})\ , \tag{16.3.15a}$$

$$\alpha^{\bar{\imath}}_{-1}|\tfrac{1}{2},\mathbf{3}\rangle_{\rm R}\ . \tag{16.3.15b}$$

Now consider the twisted sectors. There are 27 equivalence classes with rotation r, corresponding to the elements

$$h = r t_2^{n_2} t_3^{n_3} t_4^{n_4}\ , \quad n_i \in \{0,1,2\}. \tag{16.3.16}$$

The inverses of these give 27 classes with rotation r^2. The classes are in one-to-one correspondence with the fixed points,[2] which are at

$$Z^i = \frac{\exp(i\pi/6)}{3^{1/2}}(n_2R_2, n_3R_3, n_4R_4)\ . \tag{16.3.17}$$

[2] This one-to-one correspondence does not hold for more complicated space groups.

These are all related by translation, and so give 27 copies of the same spectrum. Thus we need only analyze the class $h = (r, 1; \gamma)$.

We analyze strings twisted by the rotation r. Starting on the right-moving side, in the R sector the μ oscillators are all integer moded, while the i oscillators have mode numbers $n+\frac{2}{3}$. The zero-point energy vanishes by the usual cancellation between Bose and Fermi contributions in the R sector. The only fermionic zero modes are from the spacetime fermions, $\tilde{\psi}_0^{2\pm i3}$, so there are two ground states $|\pm\frac{1}{2}\rangle_{h,\rm R}$. To figure out which survives the GSO projection we look at the bosonized vertex operators. As in eq. (10.3.25), states of a spinor field in a sector with periodicity $\varphi(\sigma_1+2\pi) = \exp(2\pi i\zeta)\varphi(\sigma_1)$ have vertex operators

$$e^{isH}\ , \ s = \tfrac{1}{2} - \zeta \bmod 1\ , \quad e^{i\tilde{s}\tilde{H}}\ , \ \tilde{s} = -\tfrac{1}{2} + \zeta \bmod 1\ , \tag{16.3.18}$$

for left- or right-movers respectively. This follows from the OPE of the bosonized spinor with the vertex operator. The vertex operators for the R sector twisted states then have

$$\exp(i\tilde{s}_a\tilde{H}_a)\ , \quad \tilde{\mathbf{s}} = (\pm\tfrac{1}{2}, -\tfrac{1}{6}, -\tfrac{1}{6}, -\tfrac{1}{6})\ . \tag{16.3.19}$$

The GSO projection as defined in chapter 10 is

$$\exp[\pi i(\tilde{s}_1 + \tilde{s}_2 + \tilde{s}_3 + \tilde{s}_4)] = 1\ . \tag{16.3.20}$$

Thus it is the state

$$|+\tfrac{1}{2}\rangle_{h,\rm R} \tag{16.3.21}$$

that remains.

In the NS sector, the fermionic modes are shifted by $\frac{1}{2}$ to $n+\frac{1}{6}$. The zero-point energy is $\frac{1}{36}$ for a complex boson of shift $\frac{1}{3}$ or $\frac{2}{3}$, and $-\frac{1}{72}$ for a complex boson of shift $\frac{1}{6}$ or $\frac{5}{6}$, and the negative in either case for a fermion, giving

$$-\frac{2}{24} - \frac{2}{48} + \frac{3}{36} + \frac{3}{72} = 0\ . \tag{16.3.22}$$

The only massless state is then the ground state

$$|0\rangle_{h,\rm NS}\ . \tag{16.3.23}$$

On the left-moving side, we will figure out the spectrum in the fermionic formulation. In the (R,NS) sector the three twisted complex fermions have mode numbers $n+\frac{1}{3}$, and the zero-point energy is

$$-\frac{2}{24} + \frac{3}{36} - \frac{3}{36} + \frac{10}{24} - \frac{16}{48} = 0\ . \tag{16.3.24}$$

There are ten fermionic zero modes, $\lambda_0^{I\pm}$ for $I = 7,\ldots,16$, so there are 32 ground states forming a $\mathbf{16}$ and $\overline{\mathbf{16}}$ of $SO(10)$. Again examining the vertex

operator as in eq. (16.3.19), the current algebra GSO projection is

$$\sum_{K=1}^{8} q_K \in 2\mathbf{Z} \tag{16.3.25}$$

in terms of the $U(1)^8$ charges. The vertex operators (16.3.18) imply that $q_1 = q_2 = q_3 = \frac{1}{6}$, and the projection then picks out the $\overline{\mathbf{16}}$.

The other current algebra sector with massless states is (NS,NS), with zero-point energy

$$-\frac{2}{24}+\frac{3}{36}+\frac{3}{72}-\frac{10}{48}-\frac{16}{48}=-\frac{1}{2}\,. \tag{16.3.26}$$

There are several massless states,

$$\lambda^{1+}_{-1/6}\lambda^{2+}_{-1/6}\lambda^{3+}_{-1/6}|0\rangle_{\mathrm{NS,NS}}\,,\quad \lambda^{I}_{-1/2}|0\rangle_{\mathrm{NS,NS}}\,,\ 7\le I\le 16\,, \tag{16.3.27a}$$

$$\lambda^{K+}_{-1/6}\alpha^{\bar{J}}_{-1/3}|0\rangle_{\mathrm{NS,NS}}\,,\ K=1,2,3\,. \tag{16.3.27b}$$

The states in the first line are a singlet and a **10** of $SO(10)$, combining with the $\overline{\mathbf{16}}$ from the (R,NS) sector to form a $\overline{\mathbf{27}}$ of E_6. The nine states in the second line transform as three **3**s of the gauge $SU(3)$, distinguished from one another by the index $\bar{\jmath}$. In all, the left-moving spectrum contains the massless states

$$(\mathbf{1},\overline{\mathbf{27}},\mathbf{1})+(\mathbf{3},\mathbf{1},\mathbf{1})^3\,. \tag{16.3.28}$$

As we have seen from the discussion of modular invariance, the h-projection is equivalent to level matching, so we can match either right-moving state (16.3.21) or (16.3.23) with any left-moving state (16.3.28). The classes (16.3.16) with rotation r give 27 copies of this spectrum, while the twisted sectors with rotation r^2 give the antiparticles.

Connection with grand unification

One of the factors in the low energy gauge group is E_6. As discussed in section 11.4, this is a possible grand unified group for the Standard Model. In E_6 unification, a generation of quarks and leptons is in the **27** or $\overline{\mathbf{27}}$ of E_6. Which representation we call the **27** and which the $\overline{\mathbf{27}}$ is a matter of convention. These are precisely the representations appearing in the $\mathbf{Z}_3$ orbifold: the helicity $\frac{1}{2}$ states that are charged under E_6 are all in the $\overline{\mathbf{27}}$ of E_6. The untwisted states (16.3.15a) comprise nine generations, forming a triplet of the gauge $SU(3)$ and a triplet of $SU(3)\in SO(6)$, and each twisted sector (16.3.28) with rotation r contributes one $\overline{\mathbf{27}}$, for 36 in all. Notice in particular that the matter is chiral, the helicity $+\frac{1}{2}$ and $-\frac{1}{2}$ states carrying different representations of the gauge group. The GSO projection correlates the spacetime helicity with the internal components of the spin,

while the twist contains both a spacetime and a gauge rotation and so correlates the internal spin with the gauge quantum numbers.

Of course this model has too many generations to be realistic, but just the same it is interesting to look at how the gauge symmetry would be reduced to the Standard Model $SU(3)\times SU(2)\times U(1)$. As we will discuss later, the $SU(3)$ symmetry can be broken by the twisted sector states $(\mathbf{3},\mathbf{1},\mathbf{1})$. There are no light states that carry both the E_6 and E_8 gauge quantum numbers, so if the Standard Model is embedded in the former the latter is *hidden*, detectable only through gravitational strength interactions. We will see later that this can have important effects, but for now we can ignore it. This leaves the E_6 factor. From experience with grand unified theories, one might expect that this could be broken to the Standard Model gauge group by the Higgs mechanism, the expectation value of a scalar field. However, that is not possible here. All scalars with E_6 charge are in the **27** representation or its conjugate, and it is not possible to break E_6 to the Standard Model gauge group with this representation. Consulting the decomposition (11.4.25), there are two components of the **27** that are neutral under $SU(3)\times SU(2)\times U(1)$, but even if both have expectation values the gauge symmetry is broken only to $SU(5)$. To break $SU(5)$ to $SU(3)\times SU(2)\times U(1)$, the smallest possible representation is the adjoint **24**, but this is not contained in the **27** of E_6. We will see in chapter 18 that this is a general property of level one current algebras. The current algebras here are at level one just as in ten dimensions, because the orbifold projection does not change their OPEs.

There are still several ways to break to the Standard Model gauge group. One is to include Wilson lines on the original torus. The full twist group of the orbifold is generated by the four elements

$$h_1=(r,0;\gamma)\,,\quad h_2=(1,t_2;\gamma_2)\,,\quad h_3=(1,t_3;\gamma_3)\,,\quad h_4=(1,t_4;\gamma_4)\,, \tag{16.3.29}$$

where the translations are now accompanied by gauge rotations. The gauge twists are highly constrained. For example, $t_2t_3=t_3t_2$ implies that $\gamma_2\gamma_3=\gamma_3\gamma_2$ by the homomorphism property (no pure gauge twists). Also, $r^3=1$ implies that $\gamma^3=1$, while $(rt_i)^3=1$ implies that $(\gamma\gamma_i)^3=1$, and so on. Further, all the elements

$$h_1h_2^{n_2}h_3^{n_3}h_4^{n_4} \tag{16.3.30}$$

must satisfy the mod 2 and level-matching conditions. Unlike the simple toroidal compactification, the general solution is not known; the number of inequivalent solutions has been estimated to be at least 10^6. Various examples resembling the Standard Model have been found. We will give one below.

We should note that if the low energy $SU(3)\times SU(2)\times U(1)$ is embedded

Table 16.1. *Allowed gauge twists for the $\mathbf{Z}_3$ orbifold, and the resulting gauge groups.*

	β_K	gauge group
(*i*)	$(\frac{1}{3},\frac{1}{3},-\frac{2}{3},0^5;0^8)$	$E_8\times E_8$
(*ii*)	$(0^8;0^8)$	$E_8\times E_8$
(*iii*)	$(\frac{1}{3},\frac{1}{3},0^6;-\frac{2}{3},0^7)$	$E_7\times U(1)\times SO(14)\times U(1)$
(*iv*)	$(\frac{1}{3},\frac{1}{3},-\frac{2}{3},0^5;\frac{1}{3},\frac{1}{3},-\frac{2}{3},0^5)$	$E_6\times SU(3)\times E_6\times SU(3)$
(*v*)	$(\frac{1}{3},\frac{1}{3},\frac{1}{3},\frac{1}{3},\frac{2}{3},0^3;\frac{2}{3},0^7)$	$SU(9)\times SO(14)\times U(1)$

in the standard way in E_6, then the usual grand unified prediction $\sin^2\theta_w = \frac{3}{8}$ still holds with Wilson line breaking even though there is no scale at which the theory looks like a four-dimensional unified theory. The reason is the inheritance principle that orbifold projections do not change the couplings of untwisted states such as the gauge bosons.

A different route to symmetry breaking is to use higher level current algebras. One way to construct an orbifold model of this type is to start with an orbifold that has two copies of the same group. For example, embedding the spin connection in each E_8 (twist (*iv*) in table 16.1), leaves an unbroken $SU(3)\times E_6\times SU(3)\times E_6$. Add a twist that has the effect of interchanging the two E_6s so that only the diagonal E_6

$$j^a = j^a_{(1)} + j^a_{(2)} \tag{16.3.31}$$

survives. The z^{-2} term in the OPE is additive, so the level is now $k=2$. The resulting model has larger representations which can break the unified group down to the Standard Model. Realistic models of this type have been constructed. The higher level and Wilson line breakings have an important difference in terms of the scale of symmetry breaking, as we discuss further in chapter 18.

Generalizations

Staying with the $\mathbf{Z}_3$ orbifold but considering more general gauge twists, there are five inequivalent solutions to the mod 2 and level-matching conditions. These are shown in table 16.1, the first twist being the solution (16.3.4) with gauge rotation equal to spacetime rotation. One realistic model with Wilson lines uses the twist (*v*) in the table, with

$$\gamma_2 = (0^7,\tfrac{2}{3};0,\tfrac{1}{3},\tfrac{1}{3},0^6)\,, \tag{16.3.32a}$$
$$\gamma_3 = 0\,, \tag{16.3.32b}$$
$$\gamma_4 = (\tfrac{1}{3},\tfrac{1}{3},\tfrac{1}{3},\tfrac{2}{3},\tfrac{1}{3},0,\tfrac{1}{3},\tfrac{1}{3};\tfrac{1}{3},\tfrac{1}{3},0^6)\,. \tag{16.3.32c}$$

With Wilson lines the gauge twist (16.3.30) is in general different at each fixed point, so the spectra in the different twisted sectors are no longer the same. Because $\gamma_3 = 0$ in this example, the 27 fixed points fall into nine sets of three, and in fact the Wilson lines have been chosen so as to reduce the number of generations to three. The gauge group is $SU(3)\times SU(2)\times U(1)^5$, with a hidden $SO(10)\times U(1)^3$. The chiral matter comprises precisely three generations, accompanied by a number of nonchiral $SU(2)$ doublets (Higgs fields) and $\mathbf{3}+\bar{\mathbf{3}}$s of $SU(3)$. There are also some massless fields coupling to the hidden gauge group, and some singlet fields. The obvious problems with this model are the extra $U(1)$ gauge symmetries, and the extra color triplets which can mediate baryon decay. Some of the singlets are moduli, and in certain of the flat directions the extra $U(1)$s are broken and the triplets heavy. Of course, given the enormous number of consistent CFTs, as well as the large number of free parameters (moduli) in each, string theory will not have real predictive power until the dynamics that selects the vacuum is understood. We will say more about this later.

Another orbifold is a square lattice in each plane with the $\mathbf{Z}_4$ rotation

$$r'\,:\quad Z^2\to iZ^2\,,\quad Z^3\to iZ^3\,,\quad Z^4\to i^{-2}Z^4\,. \tag{16.3.33}$$

Let us again embed the spin connection in the gauge connection. This will share certain features with the $\mathbf{Z}_3$ orbifold. In particular, the gauge twist is again in $SU(3)$ so the unbroken group will include an $E_6\times E_8$ factor, and the spin-$\frac{1}{2}$ states will again be in the **27** and the $\overline{\mathbf{27}}$. This will hold for any model with rotation in $SU(3)$ and with spin connection embedded in gauge connection. The extra gauge factor depends on the model; here it is $SU(2)\times U(1)$ rather than $SU(3)$. Another difference is that the modulus

$$dZ^4 dZ^4 \tag{16.3.34}$$

now survives the twist, in addition to the mixed components

$$dZ^i dZ^j\,. \tag{16.3.35}$$

This corresponds to a change in the complex structure of the compactified dimensions: when this modulus is turned on the metric is no longer Hermitean, though it becomes Hermitean again by redefining the Z^i. The moduli (16.3.34) are thus known as complex structure moduli. The moduli (16.3.35) are known as Kähler moduli, for reasons to be explained in the next chapter.

A third difference is that one finds that the helicity-$\frac{1}{2}$ states include *both* **27**s and the $\overline{\mathbf{27}}$s. We will see in later chapters that this is correlated with the appearance of the two kinds of moduli. These are generations and antigenerations, the latter having the opposite chirality. In the Standard Model there are no antigenerations, but these can obtain mass by pairing with some of the generations when some scalar fields are given expectation

values. The net number of generations is the difference, in which by definition the generations are whichever of the **27**s and $\overline{\mathbf{27}}$s are more numerous.

World-sheet supersymmetries

There is an important general pattern which will apply beyond the orbifold example. The supercurrent for the compact CFT can be separated into two pieces that separately commute with the twist,

$$i\sum_{i=2}^{4}\bar{\partial}Z^i\tilde{\psi}^{\bar{\imath}}\,,\quad i\sum_{i=2}^{4}\bar{\partial}Z^{\bar{\imath}}\tilde{\psi}^i\,. \tag{16.3.36}$$

These, together with the energy-momentum tensor and the current

$$\sum_{i=2}^{4}\tilde{\psi}^i\tilde{\psi}^{\bar{\imath}}\,, \tag{16.3.37}$$

form a right-moving $N = 2$ superconformal algebra. This is a global symmetry of the internal CFT. We will see in chapter 19 that there is a close connection between (0,2) supersymmetry on the world-sheet and $N = 1$ supersymmetry in spacetime.

When in addition the gauge twist is equal to the spacetime twist, we can do the same thing with the $\lambda^{K\pm}$ for $K = 1,2,3$, forming the left-moving supercurrents

$$i\sum_{i=2}^{4}\partial Z^i\lambda^{(i-1)-}\,,\quad i\sum_{i=2}^{4}\partial Z^{\bar{\imath}}\lambda^{(i-1)+}\,. \tag{16.3.38}$$

In this case the compact part of the world-sheet theory separates into 26 free current algebra fermions and a $(c,\tilde{c}) = (9,9)$ CFT which has (2,2) world-sheet supersymmetry. String theories of this type are highly constrained, as we will see in chapter 19.

16.4 Low energy field theory

It is interesting to look in more detail at the low energy field theory resulting from the $\mathbf{Z}_3$ orbifold.

Untwisted states

For the untwisted fields of the orbifold compactification, we can determine the low energy effective action without a stringy calculation. The action

for these fields follows directly from the ten-dimensional action via the inheritance principle. The ten-dimensional bosonic low energy action

$$S_{\text{het}} = \frac{1}{2\kappa_{10}^2}\int d^{10}x\,(-G)^{1/2}e^{-2\Phi}\Big[R + 4\partial_\mu\Phi\partial^\mu\Phi - \frac{1}{2}|\tilde{H}_3|^2 - \frac{\alpha'}{4}\mathrm{Tr}_{\mathrm{v}}(|F_2|^2)\Big] \tag{16.4.1}$$

is determined entirely by supersymmetry, with

$$\tilde{H}_3 = dB_2 - \frac{\alpha'}{4}\mathrm{Tr}_{\mathrm{v}}(A_1\wedge dA_1 + 2A_1\wedge A_1\wedge A_1/3)\,. \tag{16.4.2}$$

The trace is normalized to the vector representation of $SO(16)$.

It is very instructive to carry out this exercise. Insert into the action those fields that survive the $\mathbf{Z}_3$ projection,

$$G_{\mu\nu}\,,\ B_{\mu\nu}\,,\ \Phi\,,\ G_{i\bar{\jmath}}\,,\ B_{i\bar{\jmath}}\,,\ A_\mu^a\,,\ A_{i\bar{\jmath}\bar{k}}\,,\ A_{\bar{\imath}jk}\,. \tag{16.4.3}$$

We have subdivided the gauge generators into a in the adjoint of $SU(3)\times E_6\times E_8$, jk in the $(\mathbf{3},\mathbf{27},\mathbf{1})$, and $\bar{\jmath}\bar{k}$ in the $(\bar{\mathbf{3}},\overline{\mathbf{27}},\mathbf{1})$. Now dimensionally reduce by requiring the fields to be slowly varying functions of the x^μ and to be independent of the x^m.

Let us first ignore the ten-dimensional gauge field. The reduction is then a special case of that for the bosonic string in chapter 8,

$$S = \frac{1}{2\kappa_4^2}\int d^4x\,(-G)^{1/2}\Big[R - 2\partial_\mu\Phi_4\partial^\mu\Phi_4 - \frac{1}{2}e^{-4\Phi_4}|H_3|^2 - \frac{1}{2}G^{i\bar{\jmath}}G^{k\bar{l}}(\partial_\mu G_{i\bar{l}}\partial^\mu G_{\bar{\jmath}k} + \partial_\mu B_{i\bar{l}}\partial^\mu B_{\bar{\jmath}k})\Big]\,. \tag{16.4.4}$$

We have defined the four-dimensional dilaton

$$\Phi_4 = \Phi - \frac{1}{4}\det G_{mn}\,. \tag{16.4.5}$$

We have also made a Weyl transformation to the four-dimensional Einstein metric

$$G_{\mu\nu\text{Einstein}} = e^{-2\Phi_4}G_{\mu\nu}\ ; \tag{16.4.6}$$

henceforth in this chapter this metric is used implicitly. This action differs from the bosonic reduction (8.4.2) in that the projection has removed the Kaluza–Klein and antisymmetric tensor gauge bosons and the ij and $\bar{\imath}\bar{\jmath}$ components of the internal metric and antisymmetric tensor. Note that

$$G_{i\bar{\jmath}} = G_{\bar{\jmath}i} = G^*_{j\bar{\imath}} = G^*_{\bar{\imath}j}\,, \tag{16.4.7a}$$

$$B_{i\bar{\jmath}} = -B_{\bar{\jmath}i} = -B^*_{j\bar{\imath}} = B^*_{\bar{\imath}j}\,. \tag{16.4.7b}$$

The action must be of the general form (B.2.28) required by $N = 1$ supersymmetry. To make the comparison we must first convert the

antisymmetric tensor to a scalar as in section B.4,[3]

$$-\frac{1}{2}\int d^4x(-G)^{1/2}e^{-4\Phi_4}|H_3|^2+\int a\,dH_3$$
$$\to -\frac{1}{2}\int d^4x(-G)^{1/2}e^{4\Phi_4}\partial_\mu a\partial^\mu a\,. \qquad (16.4.8)$$

The action then takes the form

$$\frac{1}{2\kappa_4^2}\int d^4x(-G)^{1/2}\left[R-\frac{2\partial_\mu S^*\partial^\mu S}{(S+S^*)^2}-\frac{1}{2}G^{i\bar{j}}G^{k\bar{l}}\partial_\mu T_{i\bar{l}}\partial^\mu T_{\bar{j}k}\right], \qquad (16.4.9)$$

where

$$S=e^{-2\Phi_4}+ia\,,\quad T_{i\bar{j}}=G_{i\bar{j}}+B_{i\bar{j}}\,. \qquad (16.4.10)$$

This is of the supergravity form (B.2.28) with the Kähler potential

$$\kappa_4^2K=-\ln(S+S^*)-\ln\det(T_{i\bar{j}}+T_{i\bar{j}}^*)\,. \qquad (16.4.11)$$

The index i in eq. (B.2.28) is the same as the pair $i\bar{j}$ in eq. (16.4.9).

Now add the four-dimensional gauge field. In addition to its kinetic term, this appears in the Bianchi identity for the field strength $\tilde{H}$, so that the left-hand side of eq. (16.4.8) becomes

$$-\frac{1}{2}\int d^4x(-G)^{1/2}e^{-4\Phi_4}|\tilde{H}_3|^2+\int a\left[d\tilde{H}_3+\frac{\alpha'}{4}\mathrm{Tr}_v(F_2\wedge F_2)\right]. \qquad (16.4.12)$$

After Poincaré duality the additional terms in the action are

$$-\frac{1}{4g_4^2}\int e^{-2\Phi_4}\mathrm{Tr}_v(|F_2|^2)+\frac{1}{2g_4^2}\int a\mathrm{Tr}_v(F_2\wedge F_2) \qquad (16.4.13)$$

with $g_4^2=4\kappa_4^2/\alpha'$. This is of the supergravity form (B.2.28) with the gauge kinetic term

$$f_{ab}=\frac{\delta_{ab}}{g_4^2}S\,. \qquad (16.4.14)$$

Finally add the scalars coming from the ten-dimensional gauge field. The calculations are a bit longer and are left to the references. The final result is that the Kähler potential is modified to

$$\kappa_4^2K=-\ln(S+S^*)-\ln\det\left[T_{i\bar{j}}+T_{i\bar{j}}^*-\alpha'\mathrm{Tr}_v(A_iA_j^*)\right], \qquad (16.4.15)$$

there is a superpotential

$$W=\epsilon^{ijk}\mathrm{Tr}_v(A_i\,[A_j,A_k])\,, \qquad (16.4.16)$$

[3] One could instead use Poincaré duality to write the supergravity action using an antisymmetric tensor. This is known as the *linear multiplet* formalism and appears often in the string literature.

and the gauge kinetic term is unchanged. The superpotential accounts for the potential energy from the reduction of $F_{mn}F^{mn}$. We have kept the scalars in matrix notation. In components this becomes

$$W=\epsilon^{ijk}\epsilon^{\bar{l}\bar{m}\bar{n}}d^{\bar{x}\bar{y}\bar{z}}A_{i\bar{l}\bar{x}}A_{j\bar{m}\bar{y}}A_{k\bar{n}\bar{z}}\,. \qquad (16.4.17)$$

Here $d^{\bar{x}\bar{y}\bar{z}}$ is the $\mathbf{27}^3$ invariant of E_6, which has just the right form to give rise to the quark and lepton masses.

We will see in chapter 18 that several features found in this example actually apply to the tree-level effective action of every four-dimensional heterotic string theory.

T-duality

The original toroidal compactification had T-duality $O(22,6,\mathbf{Z})$. The subgroup of this that commutes with the $\mathbf{Z}_3$ twist will survive as a T-duality of the orbifold theory. In this case it is an $SU(3,3,\mathbf{Z})$ subgroup. It is interesting to look at the special case that $T_{i\bar{j}}$ is diagonal,

$$T_{i\bar{j}}=T_i\delta_{i\bar{j}}\,,\quad \text{no sum on } i\,, \qquad (16.4.18)$$

and work only to second order in the A_i. The Kähler potential becomes

$$\kappa_4^2K=-\ln(S+S^*)-\sum_i\ln(T_i+T_i^*)+\alpha'\sum_i\frac{\mathrm{Tr}_v(A_iA_i^*)}{T_i+T_i^*}\,. \qquad (16.4.19)$$

For this form of $T_{i\bar{j}}$ the lattice is a product of three two-dimensional lattices. We analyzed the T-duality of a two-dimensional toroidal lattice in section 8.4, finding it to be essentially $PSL(2,\mathbf{Z})\times PSL(2,\mathbf{Z})$. The first factor acts on τ, which characterizes the shape (complex structure) of the torus, while the second acts on ρ, which characterizes the size of the torus and the $B_{i\bar{i}}$ background. In the $\mathbf{Z}_3$ orbifold the twist fixes the shape, so $\tau=\exp(\pi i/3)$, while $\rho=iT_i$ in each plane. Thus there is a $PSL(2,\mathbf{Z})^3$ T-duality subgroup that acts as

$$T_i\to\frac{a_iT_i-ib_i}{ic_iT_i+d_i}\,,\quad a_id_i-b_ic_i=1\,. \qquad (16.4.20)$$

This takes

$$T_i+T_i^*\to\frac{T_i+T_i^*}{|ic_iT_i+d_i|^2}\,. \qquad (16.4.21)$$

The second term in the Kähler potential (16.4.19) is not invariant under this, changing by

$$\kappa_4^2K\to\kappa_4^2K+\mathrm{Re}\left[\sum_i\ln(ic_iT_i+d_i)\right]. \qquad (16.4.22)$$

This does not affect the kinetic terms because it is the real part of a holomorphic function; in other words, this is a Kähler transformation (B.2.32). The final term in the Kähler potential is invariant provided that

$$A_i \to \frac{A_i}{ic_iT_i + d_i} . \tag{16.4.23}$$

The superpotential (16.4.16) then transforms as

$$W \to \frac{W}{\prod_{i=2}^{4}(ic_iT_i + d_i)} . \tag{16.4.24}$$

This is consistent with the general Kähler transformation (B.2.33).

The space of untwisted moduli is the subspace of the toroidal moduli space that is left invariant by $\mathbf{Z}_3$. For the moduli $T_{i\bar{j}}$ this is

$$\frac{SU(3,3)}{SU(3) \times SU(3) \times SU(3,3,\mathbf{Z})} . \tag{16.4.25}$$

There are also flat directions for the matter fields A_i, giving a larger coset in all. The full moduli space for the untwisted fields is the product of this space with the dilaton–axion moduli space

$$\frac{SU(1,1)}{U(1) \times PSL(2,\mathbf{Z})} . \tag{16.4.26}$$

For orbifolds having complex structure moduli (16.3.34), the T-duality group would contain an additional $PSL(2,\mathbf{Z})$ acting on the complex structure moduli U. Various subsequent expressions are appropriately generalized. In particular the moduli space is a product of three cosets: one for the dilaton, one for the Kähler moduli, and one for the complex structure moduli.

Twisted states

For the untwisted states we were able to learn a remarkable amount from general arguments, without detailed calculations. To find the effective action for the twisted states it is necessary to do some explicit calculations with twisted state vertex operators. These methods are well developed but are too detailed for the scope of this book, so we will simply cite a few of the most interesting results.

The main one has to do with the E_6 singlet states in each twisted sector,

$$\lambda^{K+}_{-1/6}\alpha^{\bar{j}}_{-1/3}|0\rangle_{\mathrm{NS,NS}} , \quad K = 1,2,3 , \tag{16.4.27}$$

transforming as three triplets of the gauge $SU(3)$. The result is that these do not appear in the superpotential, and as a consequence the potential has a flat direction with an interesting geometric interpretation. The potential

for these modes comes only from the $SU(3)$ D-term. Defining the field $M_{K\bar{J}}$ associated with these states, the D-term (B.2.20) is

$$D^a \propto M^*_{K\bar{J}}t^a_{KL}M_{L\bar{J}} = \mathrm{Tr}(M^\dagger t^a M) , \tag{16.4.28}$$

where t^a are the fundamental $SU(3)$ matrices. Since the t^a run over a complete set of traceless matrices, D^a can vanish for all a only if

$$MM^\dagger = \rho^2 I \;\Rightarrow\; M = \rho U , \tag{16.4.29}$$

with I the identity, ρ a real constant, and U unitary. The matrix U can be taken to the identity by an $SU(3)$ gauge rotation. Thus there is a one-parameter family of vacua, along which the $SU(3)$ symmetry is completely broken.

These vacua can be understood as compactification on manifolds in which the orbifold singularity has been smoothed out (*blown up*); ρ is the radius of curvature. Thus the orbifold is a limit of the smooth spaces that we will discuss in the next chapter. Indeed, it is known that for some values of the moduli these spaces have orbifold singularities. The orbifold construction shows that the physics remain well-behaved even when the geometry appears to be singular.

The existence of the flat direction ρ can be understood as a general consequence of (2,2) world-sheet supersymmetry, the subject of chapter 19. For compactifications with less world-sheet supersymmetry similar results often hold but they are more model-dependent. We noted above that the number of consistent solutions for orbifolds with Wilson lines is a large number, of order 10^6. This is typical for free CFT constructions. However, when one takes into account that these are embedded in a larger space of smooth compactifications, many of them lie within the same moduli space and the number of distinct moduli spaces is much smaller. As we will discuss in chapter 19, with the inclusion of nonperturbative effects the number of disconnected vacua becomes smaller still.

The moduli spaces for the smooth geometries are in general more complicated and less explicitly known than the cosets that parameterize the orbifolds. The CFT corresponding to a general background of the twisted moduli is not free, because the twisted vertex operators are rather complicated. Expanding in powers of the twisted fields, the first few terms can be determined by considering string scattering amplitudes. For example, denoting a general twisted field (modulus or generation) by C_α, the leading correction to the Kähler potential takes the form

$$C_\alpha C^*_\alpha \prod_{i=2}^{4}(T_j + T^*_j)^{n^i_\alpha} . \tag{16.4.30}$$

The constants n^i_α, known as *modular weights*, can be determined from the scattering amplitudes, and for general orbifold theories are given in

the references. The invariance of the Kähler metric implies the T-duality transformation

$$C_\alpha \to C_\alpha \prod_{i=2}^{4} (ic_i T_i + d_i)^{n_\alpha^i} . \tag{16.4.31}$$

Threshold corrections

The effective action obtained above receives corrections from string loops. The most important of these is the one-loop correction to the gauge coupling, the *threshold corrections* from loops of heavy particles. The Standard Model gauge couplings are known to sufficient accuracy that predictions from unification are sensitive to this correction. Also, the dependence of f_{ab} on fields other than S comes only from one loop, and we will see later that this has an important connection with supersymmetry breaking.

To one-loop accuracy the physical gauge coupling at a scale μ can be written

$$\frac{1}{g_a^2(\mu)} = \frac{Sk_a}{g_4^2} + \frac{b_a}{16\pi^2} \ln \frac{m_{SU}^2}{\mu^2} + \frac{1}{16\pi^2} \tilde{\Delta}_a , \tag{16.4.32}$$

where SU stands for *string unification*. The subscript on g_a denotes a specific factor in the gauge group, whereas that on g_4 denotes the dimension. The first term on the right is the tree-level coupling; in the present case the current algebra level is $k_a = 1$, but for future reference we give the more general form, to be discussed in chapter 18. The second is due to the running of the coupling below the string scale, with the coefficient b_a being related to the renormalization group beta function by

$$\beta_a = \frac{b_a g_a^3}{16\pi^2} . \tag{16.4.33}$$

The final term $\tilde{\Delta}_a$ is the threshold correction. It depends on the masses of all the string states, and therefore on the moduli.

A great deal is known about $\tilde{\Delta}_a$. To calculate it directly one considers the torus amplitude in a constant background field $F^a_{\mu\nu}$, which appears in the world-sheet action in the form $F^a_{\mu\nu} j^a X^\mu \bar{\partial} X^\nu$. This can be simplified by the same sort of manipulations as we used in section 12.6 to obtain explicit loop amplitudes, though the details are longer and we just sketch the results. It is useful to separate the threshold correction as follows:

$$\tilde{\Delta}_a = \Delta_a + 16\pi^2 k_a Y . \tag{16.4.34}$$

The second term has the same dependence k_a on the gauge group as does the tree-level term. It therefore does not affect the predictions for ratios of couplings or for the unification scale, though it is important for some purposes as we will mention below.

The term Δ_a is given by an integral over moduli space,

$$\Delta_a = \int_\Gamma \frac{d^2\tau}{\tau_2} [\mathscr{B}_a(\tau, \bar{\tau}) - b_a] . \tag{16.4.35}$$

Here the function $\mathscr{B}_a(\tau, \bar{\tau})$ is related to a trace over the string spectrum weighted by Q_a^2, with Q_a the gauge charge. The limit of $\mathscr{B}_a(\tau, \bar{\tau})$ as $\tau \to i\infty$ is just b_a, so this integral converges; the term b_a was subtracted out by the matching onto the low energy field theory behavior. Also,

$$m_{SU} = \frac{2\exp[(1-\gamma)/2]}{3^{3/4}(2\pi\alpha')^{1/2}} , \tag{16.4.36}$$

where $\gamma \approx 0.577$ is Euler's constant. We will discuss the physical meaning of this scale in chapter 18. The correction Y is also given by an integral over moduli space; the calculation and final expression are somewhat more complicated than for Δ_a, due in part to the need to separate IR divergences.

For orbifold compactifications, Δ_a can be evaluated in closed form. Let us point out one important general feature. The path integral on the torus includes a sum over the twists h_1 and h_2 in the two directions. If these are generic, so that they lie in $SU(3)$ but not in any proper subgroup (in other words, if they leave only $N = 1$ supersymmetry unbroken), then they effectively force the fields in the path integral to lie near some fixed point. The path integral is therefore insensitive to the shape of the *spacetime* torus and so is independent of the untwisted moduli. If on the other hand h_1 and h_2 lie in $SU(2) \subset SU(3)$, leaving $N = 2$ unbroken, then the amplitude can depend on the moduli. An example is the $\mathbf{Z}_4$ orbifold (16.3.33), in a sector in which $h_1 = 1$ and $h_2 = r'^2$. In particular h_2 acts as

$$r'^2 : \quad Z^2 \to -Z^2 , \quad Z^3 \to -Z^3 , \quad Z^4 \to +Z^4 . \tag{16.4.37}$$

The field Z^4 is completely untwisted and so can wander over the whole spacetime torus. The threshold correction correspondingly depends on both the Kähler and complex structure moduli, T_4 and U_4. Finally, if $h_1 = h_2 = 1$ so that they leave $N = 4$ unbroken, then the threshold correction vanishes due to the $N = 4$ supersymmetry.

The actual form of the threshold correction is

$$\Delta_a = c_a - \sum_i \frac{b_a^i |P^i|}{|P|} \left\{ \ln\left[(T_i + T_i^*) |\eta(T_i)|^4 \right] + \ln\left[(U_i + U_i^*) |\eta(U_i)|^4 \right] \right\} , \tag{16.4.38}$$

with c_a independent of the moduli and η the Dedekind eta function. The sum runs over all pairs (h_1^i, h_2^i) that leave $N = 2$ unbroken. Here P is

the orbifold point group, P^i is the discrete group generated by h_1^i and h_2^i, and $|P^i|$ and $|P|$ are the orders of these groups. Also, b_a^i is the beta function coefficient for the $N = 2$ theory on the T^6/P^i orbifold, and T_i and U_i are the moduli for the fixed plane. For the $\mathbf{Z}_3$ orbifold there are no $N = 2$ sectors and the result is a constant whose value is quite small, of order 5%.

At tree level, g_a^{-2} is the real part of the holomorphic function f in the gauge kinetic term. Noting that

$$\ln\left[(T_i + T_i^*)|\eta(T_i)|^4\right] = \ln(T_i + T_i^*) + 4\,\mathrm{Re}[\ln\eta(T_i)]\,, \qquad (16.4.39)$$

the same is not true for the one-loop coupling. The second term can arise from a holomorphic one-loop contribution $4\ln\eta(T_i)$ to the function f in the effective local action obtained by integrating out massive string states (the *Wilsonian* action). The term $\ln(T_i + T_i^*)$ is due to explicit massless states. This is a general feature in supersymmetric quantum theory: it is the Wilsonian action, not the physical couplings, that has holomorphicity properties and satisfies nonrenormalization theorems. On the other hand, the physical couplings (16.4.38) are T-duality-invariant as one would expect.

Note on the other hand that the Wilsonian f is *not* T-duality-invariant, because it omits the term $\ln(T_i + T_i^*)$. This can be understood as follows. The various massless fields (including their fermionic components) transform nontrivially under T-duality due to their modular weights. This leads to an anomaly in the T-duality transformation, which is canceled by the explicit transformation of f. In fact, for orbifolds the moduli dependence of the full threshold correction $\tilde{\Delta}_a$ can be determined from holomorphicity and the cancellation of the T-duality anomaly. It has the same functional form as Δ_a but with coefficients given by sums over the modular weights.

Exercises

16.1 Find the massless spectrum of the $SO(32)$ heterotic string on the $\mathbf{Z}_3$ orbifold.

16.2 Find the massless spectrum of the $E_8 \times E_8$ heterotic string on the $\mathbf{Z}_4$ orbifold (16.3.33).

16.3 Find the massless spectrum of the six-dimensional $E_8 \times E_8$ heterotic string on the orbifold $T^4/\mathbf{Z}_2$,

$$X^m \to -X^m\,, \quad m = 6,7,8,9\,.$$

Determine the unbroken $d = 6$ supersymmetry and the supersymmetry multiplets of the massless states.

16.4 Repeat the previous exercise for the orbifold $T^4/\mathbf{Z}_3$,

$$Z^i \to \exp(2\pi i/3)Z^i\,, \quad i = 3,4\,.$$

If you do both this and the previous exercise, compare the spectra. These two orbifolds are special cases of the same *K3 surface*, to be discussed further in chapter 19.

16.5–16.7 Repeat the previous three exercises for the type IIA string.

16.8–16.10 Repeat the same three exercises for the type IIB string.

17

Calabi–Yau compactification

The study of compactification on smooth manifolds requires new, geometric, tools. A full introduction to this subject and its application to string theory would be a long book in itself. What we wish to do in this chapter is to present just the most important results, with almost all calculations and derivations omitted.

17.1 Conditions for $N = 1$ supersymmetry

We will assume four-dimensional Poincaré invariance. The metric is then of the form

$$G_{MN} = \begin{bmatrix} f(y)\eta_{\mu\nu} & 0 \\ 0 & G_{mn}(y) \end{bmatrix} . \tag{17.1.1}$$

We denote the noncompact coordinates by x^μ with $\mu, \nu = 0, \ldots, 3$ and the compact coordinates by y^m with $m, n = 4, \ldots, 9$. The indices M, N run over all coordinates, $0, \ldots, 9$. The other potentially nonvanishing fields are $\Phi(y)$, $\tilde{H}_{mnp}(y)$, and $F_{mn}(y)$.

It is convenient to focus from the start on backgrounds that leave some supersymmetry unbroken. The condition for this is that the variations of the Fermi fields are zero. This is discussed further in appendix B, in connection with eq. (B.2.25). For the $d = 10$, $N = 1$ supergravity of the heterotic string these variations are

$$\delta\psi_\mu = \nabla_\mu \varepsilon , \tag{17.1.2a}$$

$$\delta\psi_m = \left(\partial_m + \frac{1}{4}\Omega^-_{mnp}\Gamma^{np}\right)\varepsilon , \tag{17.1.2b}$$

$$\delta\chi = \left(\Gamma^m \partial_m \Phi - \frac{1}{12}\Gamma^{mnp}\tilde{H}_{mnp}\right)\varepsilon , \tag{17.1.2c}$$

$$\delta\lambda = F_{mn}\Gamma^{mn}\varepsilon . \tag{17.1.2d}$$

These are the variations of the gravitino, dilatino, and gaugino respectively. As in the corresponding nonlinear sigma model (12.3.30), the spin connection constructed from the metric appears in combination with the 3-form field strength,

$$\Omega^\pm_{MNP} = \omega_{MNP} \pm \frac{1}{2}H_{MNP} . \tag{17.1.3}$$

Under the decomposition $SO(9,1) \to SO(3,1) \times SO(6)$, the **16** decomposes as

$$\mathbf{16} \to (\mathbf{2},\mathbf{4}) + (\bar{\mathbf{2}},\bar{\mathbf{4}}) . \tag{17.1.4}$$

Thus a Majorana–Weyl **16** supersymmetry parameter can be written

$$\varepsilon(y) \to \varepsilon_{\alpha\beta}(y) + \varepsilon^*_{\alpha\beta}(y) , \tag{17.1.5}$$

where the indices on $\varepsilon_{\alpha\beta}$ transform respectively as $(\mathbf{2},\mathbf{4})$. If there is any unbroken supersymmetry, then by $SO(3,1)$ rotations we can generate further supersymmetries and so reach the form

$$\varepsilon_{\alpha\beta} = u_\alpha \zeta_\beta(y) \tag{17.1.6}$$

for an arbitrary Weyl spinor u. Each internal spinor $\zeta_\beta(y)$ for which $\delta(\text{fermions}) = 0$ thus gives one copy of the minimum $d = 4$ supersymmetry algebra.

The conditions that the variations (17.1.2) vanish for some spinor $\zeta_\beta(y)$ can be solved to obtain conditions on the background fields. Again, we quote the results without going through the calculations. Until the last section of this chapter we will make the additional assumption that the antisymmetric tensor field strength (often called the *torsion* in the literature) vanishes,

$$\tilde{H}_{mnp} = 0 . \tag{17.1.7}$$

From the vanishing of $\delta\chi$ one can then deduce that if there is any unbroken supersymmetry then the dilaton is constant,

$$\partial_m \Phi = 0 . \tag{17.1.8}$$

The vanishing of $\delta\psi_\mu$ next implies that

$$G_{\mu\nu} = \eta_{\mu\nu} , \tag{17.1.9}$$

forbidding a y-dependent scale factor. The vanishing of $\delta\psi_m$ then implies that

$$\nabla_m \zeta = 0 , \tag{17.1.10}$$

so that ζ is covariantly constant on the internal space. This is a strong condition. It implies, for example, that

$$[\nabla_m, \nabla_n]\zeta = \frac{1}{4}R_{mnpq}\Gamma^{pq}\zeta = 0 . \tag{17.1.11}$$

This means that the components Γ^{pq} that appear are not general $SO(6)$ rotations but must lie in a subgroup leaving one component of the spinor invariant. The subgroup with this property is $SU(3)$. In eq. (B.1.49) we show that under $SO(6) \to SU(3)$, the spinor decomposes $\mathbf{4} \to \mathbf{3} + \mathbf{1}$, so that if $R_{mnpq}\Gamma^{pq}$ is in this $SU(3)$ then there will be an invariant spinor.

The existence of a covariantly constant spinor ζ is thus the condition that the manifold have $SU(3)$ *holonomy*. In other words, under parallel transport around a closed loop, a spinor (or any other covariant quantity) comes back to itself not with an arbitrary rotation but with a rotation in $SU(3) \subset SO(6)$. This is the same as the condition for $N = 1$ supersymmetry in orbifolds. To see this, transport a spinor from any point to its image under the orbifold rotation: this is a closed loop on the orbifold. The orbifold is locally flat, but to compare the spinor to its original value we must rotate back. Thus the orbifold point group is the holonomy, and as we found in chapter 16, a point group in $SU(3)$ gives unbroken $d = 4$, $N = 1$ supersymmetry. Similarly, $SU(2)$ holonomy leaves a second spinor invariant and so gives an unbroken $d = 4$, $N = 2$ supersymmetry.

The final supersymmetric variation $\delta\lambda^a$ vanishes if $F^a_{mn}\Gamma^{mn}$ is also an $SU(3)$ rotation. Writing the indices on F_{mn} in terms of the complex indices transforming under $SU(3)$, this means that

$$F_{ij} = F_{\bar{i}\bar{j}} = 0\,, \quad G^{i\bar{j}}F_{i\bar{j}} = 0\,. \tag{17.1.12}$$

In addition we must impose the Bianchi identities on the various field strengths. In particular, for the torsion this is

$$d\tilde{H}_3 = \frac{\alpha'}{4}\left[\mathrm{tr}(R_2 \wedge R_2) - \mathrm{Tr}_v(F_2 \wedge F_2)\right]\,. \tag{17.1.13}$$

For vanishing $\tilde{H}$, this condition is quite strong, and the only solution seems to be to set R_2 and F_2 essentially equal. That is, consider $SO(6) \subset SO(16) \subset E_8$, and require the gauge connection to be equal to the spin connection ω_μ of the Lorentz $SO(6)$. This is referred to as embedding the spin connection in the gauge connection, generalizing the same idea in the orbifold. Recall that the corrections (17.1.13) were deduced in section 12.3 from anomalies on the world-sheet. When the spin connection is embedded in the gauge connection, six of the current algebra λ^A couple in the same way as the $\tilde{\psi}^m$. The relevant part of the world-sheet theory is then parity-invariant, accounting for the cancellation of anomalies.

With the spin connection embedded in the gauge connection, the conditions (17.1.12) for the vanishing of the gaugino variation follow from $SU(3)$ holonomy. The Bianchi identity for the field strength also follows from that for the curvature. It remains to consider the equations of motion. We might have begun with these, but it is easiest to save them for the end because at this point they are automatically satisfied. With vanishing

torsion and a constant dilaton the field equations reduce to

$$R_{mn} = 0\,, \quad \nabla^m F_{mn} = 0\,. \tag{17.1.14}$$

These can be shown to follow respectively from $SU(3)$ holonomy and the conditions (17.1.12).

We should remember that the field equations and supersymmetry variations in this section are only the leading terms in an expansion in derivatives, $\alpha'^{1/2}\partial_m$ being the dimensionless parameter. The conditions we have found are therefore correct when the length scale R_c of the compactified manifold is large compared to the string scale. However, we will see in section 17.5 that many of the conclusions have a much wider range of validity.

17.2 Calabi–Yau manifolds

To summarize, we found in the last section that under the assumption of vanishing torsion, the compactified dimensions must form a space of $SU(3)$ holonomy. In this section we present some of the relevant mathematics. Again, we give only definitions and results, without derivations. All manifolds in this section are assumed to be compact.

Real manifolds

We need to introduce the ideas of *cohomology* and *homology*. The exterior derivative d introduced in section B.4 is nilpotent, $d^2 = 0$. As with the BRST operator, this allows us to define a cohomology. A p-form ω_p is *closed* if $d\omega_p = 0$ and *exact* if $\omega_p = d\alpha_{p-1}$ for some $(p-1)$-form. A closed p-form can always be written locally in the form $d\alpha_{p-1}$, but not necessarily globally. Thus we define the pth *de Rham cohomology* of a manifold K,

$$H^p(K) = \frac{\text{closed } p\text{-forms on } K}{\text{exact } p\text{-forms on } K}\,. \tag{17.2.1}$$

The dimension of $H^p(K)$ is the *Betti number* b_p. The Betti numbers depend only on the topology of the space. In particular, the *Euler number* is

$$\chi(K) = \sum_{p=0}^{d}(-1)^p b_p\,. \tag{17.2.2}$$

The operator

$$\Delta_d = *d*d + d*d* = (d + *d*)^2 \tag{17.2.3}$$

is a second order differential on p-forms which reduces to the Laplacian in flat space. The Poincaré $*$ is defined in section B.4. A p-form is said to

be *harmonic* if $\Delta_d\omega = 0$. It can be shown that the harmonic p-forms are in one-to-one correspondence with the group $H^p(K)$: each equivalence class contains exactly one harmonic form. Using the Poincaré dual one can turn a harmonic p-form into a harmonic $(d-p)$-form. This is the Hodge $*$ map between $H^p(K)$ and $H^{d-p}(K)$, and implies

$$b_p = b_{d-p} \,. \tag{17.2.4}$$

For submanifolds of K one can define the boundary operator δ, which is also nilpotent. Rather than on a submanifold N itself it is useful to focus on the corresponding integral

$$\int_N \tag{17.2.5}$$

since these form a vector space: we can consider arbitrary real linear combinations, called *chains*.[1] We can then define *closed* and *exact* with respect to δ; a closed chain is a *cycle*. The *simplicial homology* for p-dimensional submanifolds (p-chains) is

$$H_p(K) = \frac{\text{closed } p\text{-chains in } K}{\text{exact } p\text{-chains in } K} \,. \tag{17.2.6}$$

That is, it consists of closed submanifolds that are not themselves boundaries.

There is a one-to-one correspondence between $H^p(K)$ and $H_{d-p}(K)$. For any p-form ω_p there is a $(d-p)$-cycle $N(\omega)$ with the property that

$$\int_K \omega_p \wedge \alpha_{d-p} = \int_{N(\omega)} \alpha_{d-p} \tag{17.2.7}$$

for all closed $(d-p)$-forms.

Complex manifolds

A complex manifold is an even-dimensional manifold, $d = 2n$, such that we can form n complex coordinates z^i and the transition functions

$$z'^i(z^j) \tag{17.2.8}$$

are holomorphic between all pairs of patches. Specifically, this is a *complex n-fold*. We have encountered this idea for $n = 1$ on the string world-sheet. Two complex manifolds are equivalent if there is a one-to-one holomorphic map between them. As we have seen in the case of Riemann surfaces, a manifold of given topology can have more than one inequivalent complex

[1] This will define *real* homology; by analogy one can define integer homology, complex homology, and so on.

structure. A *Hermitean metric* on a complex manifold is one for which

$$G_{ij} = G_{\bar{i}\bar{j}} = 0 \,. \tag{17.2.9}$$

On a complex manifold we can define (p,q)-forms as having p antisymmetric holomorphic indices and q antisymmetric antiholomorphic indices,

$$\omega_{i_1\cdots i_p\bar{j}_1\cdots\bar{j}_q} \,. \tag{17.2.10}$$

The relative order of the different types of index is not important and can always be taken as shown. We can similarly separate the exterior derivative, $d = \partial + \bar{\partial}$, where

$$\partial = dz^i\partial_i \,, \quad \bar{\partial} = d\bar{z}^{\bar{i}}\partial_{\bar{i}} \,. \tag{17.2.11}$$

Then ∂ and $\bar{\partial}$ take (p,q)-forms into $(p+1,q)$-forms and $(p,q+1)$-forms respectively. Each is nilpotent,

$$\partial^2 = \bar{\partial}^2 = 0 \,. \tag{17.2.12}$$

Thus we can define the *Dolbeault cohomology*

$$H^{p,q}_{\bar{\partial}}(K) = \frac{\bar{\partial}\text{-closed } (p,q)\text{-forms in } K}{\bar{\partial}\text{-exact } (p,q)\text{-forms in } K} \,. \tag{17.2.13}$$

The dimension of $H^{p,q}_{\bar{\partial}}(K)$ is the *Hodge number* $h^{p,q}$.

Using the inner product

$$\int d^nz d^n\bar{z}\,(G)^{1/2}G^{\bar{i}i'}\cdots G^{j\bar{j}'}\cdots(\omega_{i\cdots\bar{j}\cdots})^*\omega_{i'\cdots\bar{j}'\cdots} \,, \tag{17.2.14}$$

one defines the adjoints $\partial^\dagger$ and $\bar{\partial}^\dagger$ and the Laplacians

$$\Delta_\partial = \partial\partial^\dagger + \partial^\dagger\partial \,, \quad \Delta_{\bar{\partial}} = \bar{\partial}\bar{\partial}^\dagger + \bar{\partial}^\dagger\bar{\partial} \,. \tag{17.2.15}$$

Then the $\Delta_{\bar{\partial}}$-harmonic (p,q)-forms are in one-to-one correspondence with $H^{p,q}_{\bar{\partial}}(K)$.

Kähler manifolds

Kähler manifolds are complex manifolds with a Hermitean metric of a special form. The additional restriction can be stated in several ways. Define the *Kähler form*

$$J_{1,1} = iG_{i\bar{j}}dz^i d\bar{z}^{\bar{j}} \,. \tag{17.2.16}$$

One way to define a Kähler manifold is that the Kähler form is closed,

$$dJ_{1,1} = 0 \,. \tag{17.2.17}$$

A second is that parallel transport takes holomorphic indices only into holomorphic indices. In other words, the holonomy is in $U(n) \subset SO(2n)$.

A final equivalent statement is that the metric is locally of the form

$$G_{i\bar{\jmath}} = \frac{\partial}{\partial z^i}\frac{\partial}{\partial \bar{z}^{\bar{\jmath}}}K(z,\bar{z})\ . \tag{17.2.18}$$

The *Kähler potential* $K(z,\bar{z})$ need not be globally defined. The potential

$$K'(z,\bar{z}) = K(z,\bar{z}) + f(z) + f(z)^* \tag{17.2.19}$$

gives the same metric, and it may be necessary to take different potentials in different patches. We are now focusing on the spacetime geometry, but we have seen this same idea in *field space* in eq. (B.2.32).

For Kähler metrics the various Laplacians become identical,

$$\Delta_d = 2\Delta_{\bar{\partial}} = 2\Delta_{\partial}\ . \tag{17.2.20}$$

Then the cohomologies

$$H^{p,q}_{\bar{\partial}}(K) = H^{p,q}_{\partial}(K) \equiv H^{p,q}(K) \tag{17.2.21}$$

are the same. The Hodge and Betti numbers are therefore also related,

$$b_k = \sum_{p=0}^{k} h^{p,k-p}\ . \tag{17.2.22}$$

Complex conjugation gives

$$h^{p,q} = h^{q,p} \tag{17.2.23}$$

and the Hodge $*$ gives

$$h^{n-p,n-q} = h^{p,q}\ . \tag{17.2.24}$$

Since the Kähler form is closed it is in $H^{1,1}(K)$. Its equivalence class is known as the *Kähler class* and is always nontrivial. Taking a basis ω_A for $H^{1,1}(K)$, we can expand

$$J_{1,1} = \sum_A v^A \omega_{1,1\,A}\ , \tag{17.2.25}$$

and the real parameters v^A label the Kähler class.

Manifolds of $SU(3)$ holonomy

A manifold has $SU(3)$ holonomy if and only if it is Ricci-flat and Kähler. While there are many examples of Kähler manifolds, there are few explicit examples of Ricci-flat Kähler metrics. There is, however, an important existence theorem. For a Kähler manifold, only the mixed components $R_{i\bar{\jmath}}$ of the Ricci tensor are nonzero. Further, the *Ricci form*

$$\mathscr{R}_{1,1} = R_{i\bar{\jmath}}dz^i d\bar{z}^{\bar{\jmath}} \tag{17.2.26}$$

is closed, $d\mathscr{R}_{1,1} = 0$. It therefore defines an equivalence class in $H^{1,1}(K)$. With normalization $\mathscr{R}_{1,1}/2\pi$, this is known as the *first Chern class* c_1. Obviously this class is trivial for a Ricci-flat manifold. The hard theorem, conjectured by Calabi and proved by Yau, is that for any Kähler manifold with $c_1 = 0$ there exists a unique Ricci-flat metric with a given complex structure and Kähler class. A vanishing first Chern class, $c_1 = 0$, means that $\mathscr{R}_{1,1}$ is exact. A Kähler manifold with $c_1 = 0$ is known as a *Calabi–Yau manifold.*

Another theorem states that a Kähler manifold has $c_1 = 0$ if and only if there is a nowhere vanishing holomorphic $(3,0)$-form $\Omega_{3,0}$. The $(3,0)$-form is covariantly constant in the Ricci-flat metric. It further can be shown that

$$h^{p,0} = h^{3-p,0}\ . \tag{17.2.27}$$

For any complex manifold $h^{0,0} = 1$, corresponding to the constant function. Finally, for a Calabi–Yau manifold of exactly $SU(3)$ holonomy and not a subgroup, it can be shown that

$$b_1 = h^{1,0} = h^{0,1} = 0\ . \tag{17.2.28}$$

Using the various properties above, all the Hodge numbers of a Calabi–Yau 3-fold are fixed by just two independent numbers, $h^{1,1}$ and $h^{2,1}$. The full set of Hodge numbers is conventionally displayed as *Hodge diamond,*

$$\begin{array}{ccccccc} & & h^{3,3} & & 1 & & \\ & h^{3,2} & & h^{2,3} & & 0\ \ 0 & \\ h^{3,1} & & h^{2,2} & & h^{1,3} & 0\ \ h^{1,1}\ \ 0 & \\ h^{3,0} & & h^{2,1} & & h^{1,2} & & h^{0,3} \\ & h^{2,0} & & h^{1,1} & & h^{0,2} & \\ & & h^{1,0} & & h^{0,1} & & \\ & & & h^{0,0} & & & \end{array} = \begin{array}{ccccccc} & & 1 & & 1 & & \\ & 0 & & 0 & & 0\ \ 0 & \\ 0 & & h^{1,1} & & 0 & 1\ \ 1\ \ 1 & \\ 1 & & h^{2,1} & & h^{2,1} & & 1 \\ & 0 & & h^{1,1} & & 0 & \\ & & 0 & & 0 & & \\ & & & 1 & & & \end{array}\ . \tag{17.2.29}$$

In particular, the Euler number (17.2.2) is

$$\chi = 2(h^{1,1} - h^{2,1})\ . \tag{17.2.30}$$

Examples

An even-dimensional torus is a Calabi–Yau manifold but an uninteresting one: the holonomy is trivial. To break to $N = 1$ supersymmetry we need nontrivial $SU(3)$ holonomy. The $\mathbf{Z}_3$ orbifold of T^6 has this property but is not a manifold, having orbifold singularities. A smooth Calabi–Yau space can be produced by blowing up all the singularities, as follows. The *Eguchi–Hanson space* EH_3 has three complex coordinates w_i with metric

$$G_{i\bar{\jmath}} = \left(1 + \frac{\rho^6}{r^6}\right)^{1/3}\left[\delta_{i\bar{\jmath}} - \frac{\rho^6 w_i \bar{w}_{\bar{\jmath}}}{r^2(\rho^6 + r^6)}\right]\ , \tag{17.2.31}$$

where $r^2 = w_i \bar{w}_{\bar{i}}$ and ρ is a constant that sets the scale of the geometry. After the identification

$$w_i \cong \exp(2\pi i/3) w_i \,, \tag{17.2.32}$$

this becomes an everywhere smooth space which is asymptotically $R^6/\mathbf{Z}_3$. This is the same as the geometry around the $T^6/\mathbf{Z}_3$ orbifold fixed points. Each orbifold fixed point can be replaced by a small EH_3 to give a smooth Calabi–Yau space. The $(1,1)$-forms are the nine $dz_i d\bar{z}_{\bar{j}}$ and the 27 blow-up modes $\partial G_{i\bar{j}}/\partial\rho$ from varying the sizes of the EH_3s. There is only one complex structure, so

$$h^{1,1} = 36 \,, \quad h^{2,1} = 0 \,, \quad \chi = 72 \,. \tag{17.2.33}$$

A second construction starts with *complex projective space* CP^n, formed by taking $n+1$ complex coordinates and identifying

$$(z_1, z_2, \ldots, z_{n+1}) \cong (\lambda z_1, \lambda z_2, \ldots, \lambda z_{n+1}) \tag{17.2.34}$$

for any complex λ. The identification is important because it makes the space compact. The space CP^n is Kähler but not Calabi–Yau; many Calabi–Yau manifolds can be obtained from it as submanifolds. In particular, let G be a homogeneous polynomial in the z^i,

$$G(\lambda z_1, \ldots, \lambda z_{n+1}) = \lambda^k G(z_1, \ldots, z_{n+1}) \tag{17.2.35}$$

for some k. The submanifold of CP^n defined by

$$G(z_1, \ldots, z_{n+1}) = 0 \tag{17.2.36}$$

is a Kähler manifold of complex dimension $n-1$. It can be shown that this submanifold has vanishing c_1 for $k = n+1$, so that a quintic polynomial in CP^4, which is $(n,k) = (4,5)$, gives a good manifold for string compactification. This manifold can be shown to have

$$h^{1,1} = 1 \,, \quad h^{2,1} = 101 \,, \quad \chi = -200 \,. \tag{17.2.37}$$

The unique Kähler modulus is the overall scale of the manifold. The complex structure moduli correspond to the parameters in the polynomial G, which after taking into account linear coordinate redefinitions number $9!/(5!\cdot 4!) - 25 = 101$.

Obvious generalizations include starting with a product of CP^n spaces, requiring several polynomials to vanish, and using weighted projective spaces where coordinates scale by different powers of λ. One can also divide by a discrete symmetry. For example, a particular case of the quintic polynomial in CP^4,

$$z_1^5 + z_2^5 + z_3^5 + z_4^5 + z_5^5 = 0 \,, \tag{17.2.38}$$

has a $\mathbf{Z}_5 \times \mathbf{Z}_5$ symmetry which is freely acting, meaning that it has no fixed points. Since the Euler number χ can be written as an integral over the curvature, identifying by this $\mathbf{Z}_5 \times \mathbf{Z}_5$ reduces χ by a factor of 25, to $\chi = -8$. Identifying by a symmetry with fixed points produces a space with orbifold singularities. These can be blown up, but the Euler number is then not simply obtained by dividing by the order of the group, because of the curvature at the blow-ups.

Another example is the *Tian–Yau space*. This is formed from two copies of CP^3 with coordinates z_i and w_i by imposing three polynomial equations: one cubic in z, one cubic in w, and one linear in z and linear in w. This has $\chi = -18$, and there is a freely-acting $\mathbf{Z}_3$ symmetry which can reduce this to $\chi = -6$.

World-sheet supersymmetry

With the spin connection embedded in the gauge connection, the interacting part of the world-sheet theory is invariant under parity, which interchanges ψ^i with $\lambda^{(i-1)+}$ for $i = 2,3,4$. Since the heterotic theory has a $(0,1)$ superconformal symmetry, the parity symmetry implies that it is enlarged to $(1,1)$.

For any metric $G_{mn}(y)$, the superfield formalism of section 12.3 allows us to write a nonlinear sigma model having (1,1) supersymmetry. If in addition the metric is Kähler, then there is actually (2,2) supersymmetry. One way to see this is to observe that this is the condition for the mixed components $\omega_{ai}^{\bar{j}}$ and $\omega_{a\bar{i}}^{j}$ of the spin connection to vanish, and therefore for the world-sheet action (12.3.30) to be invariant under a $U(1)$ rotation of the complex fermions,

$$\tilde{\psi}^i \to \exp(i\theta)\tilde{\psi}^i \,. \tag{17.2.39}$$

The right-moving supercurrent then separates into two terms

$$iG_{\bar{i}j}\bar{\partial}X^{\bar{i}}\tilde{\psi}^j + iG_{i\bar{j}}\bar{\partial}X^i\tilde{\psi}^{\bar{j}} \,, \tag{17.2.40}$$

which have opposite charges under the $U(1)$ symmetry and so must be separately conserved. The left-moving supercurrent also separates. Another way to see the enlarged supersymmetry is by dimensional reduction of $d = 4$, $N = 1$ supersymmetry. As discussed in section B.2, this supersymmetry requires that the field space be Kähler; dimensional reduction takes the four generators of $d = 4$, $N = 1$ into $d = 2$ (2,2).

When the metric satisfies the stronger condition of $SU(n)$ holonomy, the sigma model is conformally invariant, the supersymmetries are extended to superconformal symmetries, and the $U(1)$ global symmetry is extended to left- and right-moving $U(1)$ current algebras.

For Calabi–Yau compactification, the interacting part of the world-sheet theory is a $(c,\tilde{c}) = (9,9)$ CFT. Since the spin connection is embedded in the gauge connection, the six interacting λ^A couple in the same way

as supersymmetric fermions ψ^m. Thus, as with the orbifold example, the world-sheet theory has (2,2) superconformal symmetry. In chapters 18 and 19 we will study this world-sheet symmetry systematically. We will see that a minimum of (0,2) supersymmetry on the world-sheet is necessary in order to have *spacetime* supersymmetry. We will also see that the extra left-moving supersymmetry of Calabi–Yau compactification is responsible for a great deal of special structure.

17.3 Massless spectrum

We now look at the spectrum of fluctuations around the background. We will use lower case a, g, b, and ϕ to distinguish the fluctuations from the background fields. The various wave operators separate into noncompact and internal pieces, for example

$$\nabla_M \nabla^M = \partial_\mu \partial^\mu + \nabla_m \nabla^m , \tag{17.3.1a}$$

$$\Gamma_M \nabla^M = \Gamma_\mu \partial^\mu + \Gamma_m \nabla^m . \tag{17.3.1b}$$

The solutions similarly separate into a sum over functions of x^μ times a complete set of functions of y^m. Massless fields in four dimensions arise from those modes of the ten-dimensional massless fields that are annihilated by the internal part of the wave operator.

We start with the ten-dimensional gauge field. The ten-dimensional index separates $M \to \mu, i, \bar{\imath}$. Similarly the adjoint decomposes under

$$E_8 \times E_8 \to SU(3) \times E_6 \times E_8 \tag{17.3.2}$$

into

$$a\colon\ (\mathbf{1},\mathbf{78},\mathbf{1}) + (\mathbf{1},\mathbf{1},\mathbf{248}) , \tag{17.3.3a}$$

$$ix\colon\ (\mathbf{3},\mathbf{27},\mathbf{1}) , \quad \bar{\imath}\bar{x}\colon\ (\bar{\mathbf{3}},\bar{\mathbf{27}},\mathbf{1}) , \quad i\bar{\jmath}\colon\ (\mathbf{8},\mathbf{1},\mathbf{1}) . \tag{17.3.3b}$$

That is, a denotes the adjoint of $E_6 \times E_8$, x the **27** of E_6 and i, j the **3** of $SU(3)$. We use the same index for the **3** of the gauge $SU(3)$ and the spacetime $SU(3)$ because their connections are the same.

We denote the various components of the gauge fluctuation as $a_{M,X}$ with X any of the gauge components (17.3.3). The massless modes of the form $a_{\mu,X}$ are the unbroken gauge fields in four dimensions. These arise from gauge symmetries that commute with the background fields. Since the latter are in $SU(3)$, the four-dimensional gauge symmetry is $E_6 \times E_8$, meaning $X = a$. In terms of the wave operator, the internal part acting on $a_{\mu,X}$ is the scalar Laplacian $\nabla^m \nabla_m$ with gauge-covariant derivative. It has zero modes only for fields that are neutral under the background gauge fields. Comparing with the $\mathbf{Z}_3$ orbifold, the low energy symmetry $SU(3)$ is absent. This is consistent with the analysis (16.4.29) of the blowing-up modes, which were seen to break the $SU(3)$ symmetry.

The field $a_{i,a}$ can be regarded as a $(1,0)$-form, with only the index i coupling to the background connection, and the relevant wave operator is in fact Δ_d. The number of zero modes is then $h^{1,0}$, which vanishes for $SU(3)$ holonomy.

The field $a_{i,jx}$ is not a $(2,0)$-form because the tangent space and gauge indices are not antisymmetrized. However, by using the metric and the antisymmetric three-form we can produce

$$a_{i\bar{l}\bar{m}x} = a_{i,jx} G^{j\bar{k}} \Omega_{\bar{k}\bar{l}\bar{m}} , \tag{17.3.4}$$

which is a $(1,2)$-form on the indices $i\bar{l}\bar{m}$. The relevant wave operator is again Δ_d so the number of zero modes is $h^{2,1}$. These fields are scalars in the **27** of E_6.

The field $a_{i,\bar{\jmath}\bar{x}}$ is a $(1,1)$-form and the relevant wave operator is again Δ_d. The number of zero modes is $h^{1,1}$. These fields are scalars in the $\bar{\mathbf{27}}$ of E_6.

The field $a_{i,j\bar{k}}$ cannot be written as a (p,q)-form and the number of massless modes is not given by a Hodge number. This field can be regarded as a 1-form (the index i) transforming as a generator of the Lorentz group (the indices $j\bar{k}$); the corresponding cohomology is denoted $H^1(\mathrm{End}\, T)$. Because these are neutral under E_6 they are less directly relevant to the low energy physics than the charged fields. We will discuss some of their physics in section 17.6.

The zero modes of $a_{\bar{\imath},X}$ are the conjugates of those of $a_{i,X}$.

The massless modes of the gaugino must be the same as those of $a_{M,X}$ by supersymmetry. This is related directly to the $SU(3)$ holonomy. Under $SO(9,1) \to SO(3,1) \times SU(3)$,

$$\mathbf{16} \to (\mathbf{2},\mathbf{1}) + (\mathbf{2},\mathbf{3}) + (\bar{\mathbf{2}},\bar{\mathbf{1}}) + (\bar{\mathbf{2}},\bar{\mathbf{3}}) . \tag{17.3.5}$$

The $(\mathbf{2},\mathbf{1})$ is neutral under the tangent space group and so couples in the same way as a^μ, providing the four-dimensional gauginos. The $(\mathbf{2},\mathbf{3})$ couples in the same way as a_i and so provides the fermionic partners of those scalars. Thus there are $h^{2,1}$ **27**s and $h^{1,1}$ $\bar{\mathbf{27}}$s in the **2** of $SO(3,1)$. The spectrum is chiral and the net number of generations minus antigenerations is

$$|h^{2,1} - h^{1,1}| = \frac{|\chi|}{2} . \tag{17.3.6}$$

This is 36 for the blown-up orbifold, just as for the singular orbifold, and 100 for the quintic in CP^4. However, dividing by $\mathbf{Z}_5 \times \mathbf{Z}_5$ reduces the latter number to a more reasonable 4, while the Tian–Yau space has a net

of 3 generations. The relation (17.3.6) can be understood from an index theorem for the Dirac equation.

Dividing by a discrete group is also useful for breaking the E_6 symmetry. We saw for the orbifold that this could be done by Wilson lines, gauge backgrounds that are locally trivial but give a net rotation around closed curves. The Calabi–Yau spaces produced by polynomial equations in projective spaces are simply connected, but dividing by a freely-acting group produces nontrivial closed curves running from a point to its image. Adding a Wilson line means that the string theory is twisted by the product of the spacetime symmetry and a gauge rotation W. The nontrivial curves produced in dividing by $\mathbf{Z}_n$ have the property that if traversed n times they become closed paths on the original (covering) space, which are all topologically trivial. Thus the Wilson line must also satisfy $W^n = 1$.

For a freely-acting group, adding Wilson lines does not change the net number of generations. However, the different quark and lepton multiplets of a given generation in general come from different **27**s of the untwisted theory. Thus, while the inheritance principle requires the Standard Model gauge couplings to satisfy E_6 relations, the Yukawa couplings of the quarks and leptons in general do not. This is good because the E_6 relations for the gauge couplings (which are the same as the $SU(5)$ relations) work rather well, while those for the Yukawa couplings are more mixed, with only the heaviest generation ratio m_b/m_τ working well. This may also help to account for the stability of the proton, as the Higgs couplings that give mass to the quarks and leptons are no longer related to couplings of color triplet scalars that might mediate baryon decay.

Now we consider the bosonic supergravity fields, g_{MN}, b_{MN}, and ϕ. The components with all indices noncompact, $g_{\mu\nu}$, $b_{\mu\nu}$, and ϕ, each have a single zero mode (the constant function) giving the corresponding field in four dimensions.

The components $g_{\mu i}$ and $b_{\mu i}$ are $(1,0)$-forms on the internal space and so have no zero modes because $h^{1,0} = 0$. In particular, massless modes of $g_{\mu i}$ would be Kaluza–Klein gauge bosons, which are in one-to-one correspondence with the continuous symmetries of the internal space. It can be shown that a Calabi–Yau manifold has no continuous symmetries.

The components g_{ij} correspond to changes in the complex structure, since a coordinate change would be needed to bring the metric back to Hermitean form. This field is symmetric and so not a (p,q)-form, but by the same trick as for $a_{i\bar{l}\bar{m}x}$ we can form

$$g_{i\bar{l}\bar{m}} = g_{ij}G^{j\bar{k}}\Omega_{\bar{k}\bar{l}\bar{m}} \,. \tag{17.3.7}$$

The wave operator is Δ_d and so the number of complex structure moduli is $h^{2,1}$. These are complex fields, with $g_{\bar{i}j}$ being the conjugate. The field b_{ij} is a $(2,0)$-form and has $h^{2,0} = 0$ zero modes.

The fluctuation $g_{i\bar{j}}$ is a $(1,1)$-form, and the wave operator is Δ_d. Thus it gives rise to $h^{1,1}$ real moduli. The field $b_{i\bar{j}}$ also is a $(1,1)$-form and gives $h^{1,1}$ real moduli. These combine to form $h^{1,1}$ complex fields.

The full massless spectrum is

- $d = 4$, $N = 1$ supergravity: $G_{\mu\nu}$ and the gravitino.
- The dilaton–axion chiral superfield S.
- Gauge bosons and gauginos in the adjoint of $E_6 \times E_8$.
- $h^{2,1}$ chiral superfields in the **27** of E_6.
- $h^{1,1}$ chiral superfields in the $\overline{\mathbf{27}}$ of E_6.
- $h^{2,1}$ chiral superfields for the complex structure moduli.
- $h^{1,1}$ chiral superfields for the Kähler moduli.
- Some number of E_6 singlets from $H^1(\text{End } T)$.

17.4 Low energy field theory

We would now like to deduce the effective four-dimensional action for the massless fields. We emphasize again that the actual calculations are omitted, but we will outline the method and the results. The general $d = 4$, $N = 1$ supersymmetric action depends on two holomorphic functions, the gauge kinetic term and the superpotential, and one general function, the Kähler potential. We will show in the next chapter that the gauge kinetic term is the same in all heterotic string compactifications, and so we need determine only the other two functions. In this section and the next we will ignore the E_6 singlet fields from $H^1(\text{End } T)$, setting their values to zero.

We consider the low energy effective field theory at string tree level. For now we assume the compactification radius to be large compared to the string length, so that we can restrict attention to the massless fields of the ten-dimensional theory and also ignore higher dimension terms in the effective action. This is the *field-theory approximation.* In the next section we consider corrections to this approximation.

Expand each ten-dimensional field in a complete set of eigenfunctions $f_m(y)$ of the appropriate wave operator on the internal space, schematically

$$\varphi(x,y) = \sum_m \phi_m(x) f_m(y) \,. \tag{17.4.1}$$

Insert this into the ten-dimensional action and integrate over the internal

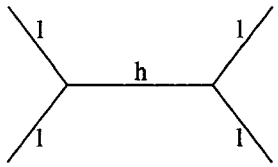


Fig. 17.1. Quartic interaction among light fields induced by integrating out a heavy field.

coordinates to obtain the four-dimensional Lagrangian density,

$$\mathscr{L}_4(\phi) = \int d^6y\, \mathscr{L}_{10}(\varphi)\,. \tag{17.4.2}$$

This still depends on all the functions $\phi_m(x)$, the infinite number of massive fields as well as the finite number of massless ones. Split $\varphi(x,y)$ into 'light' and 'heavy' parts,

$$\varphi = \varphi_{\mathrm{l}} + \varphi_{\mathrm{h}}, \tag{17.4.3}$$

according to whether f_m has a zero or nonzero eigenvalue under the internal wave operator. We want to integrate φ_{h} out so as to obtain an effective action for the finite number of four-dimensional fields in φ_{l}. The simplest approach would be to set $\varphi_{\mathrm{h}} = 0$ in $\mathscr{L}_4$, but this is not quite right. Since we are at string tree level we can treat the problem classically: what we must do is extremize the action with respect to φ_{h} with φ_{l} fixed. The result is the effective action for φ_{l}. As a schematic example, consider the following terms

$$m\varphi_{\mathrm{h}}^2 + g\varphi_{\mathrm{h}}\varphi_{\mathrm{l}}^2\,. \tag{17.4.4}$$

Setting φ_{h} to its extremum $-g\varphi_{\mathrm{l}}^2/2m$ leaves the effective interaction

$$-\frac{g^2}{4m}\varphi_{\mathrm{l}}^4 \tag{17.4.5}$$

for the light fields. Figure 17.1 shows the corresponding Feynman graph. This is known as a Kaluza–Klein correction to the low energy action. It is easy to see that these always involve at least four light fields. With an interaction $\varphi_{\mathrm{h}}\varphi_{\mathrm{l}}$ we could induce a quadratic or cubic term, but this is absent by definition. It is an off-diagonal mass term mixing the light and heavy fields, but the latter are defined to be eigenstates of zero mass. The terms that we will be interested in contain two or three light fields and so we can ignore the Kaluza–Klein corrections.

Let us first consider the fields associated with $(1,1)$-forms, beginning with the superpotential for the $\overline{\mathbf{27}}$s of E_6. We will focus on the renormalizable terms, which are at most cubic in the fields. The quadratic terms vanish because the $\overline{\mathbf{27}}$s are massless, and the linear terms vanish because their presence would imply that the background is not supersymmetric by eq. (B.2.25); actually both terms are forbidden by E_6 as well. Thus we are interested in terms that are precisely cubic. These are related to four-dimensional Yukawa couplings. The relevant expansions for φ_{l} are

$$a_{i,\bar{j}\bar{x}}(x,y) = \sum_A \phi^A{}_{\bar{x}}(x)\omega_{Ai\bar{j}}(y)\,, \tag{17.4.6a}$$

$$\lambda_{i,\bar{j}\bar{x}}(x,y) = \sum_A \lambda^A{}_{\bar{x}}(x)\omega_{Ai\bar{j}}(y)\,, \tag{17.4.6b}$$

where A runs over a complete set of nontrivial $(1,1)$-forms; henceforth summation convention is used for this index. A four-dimensional Weyl spinor index on λ is suppressed. Inserting these expansions into the action, the ten-dimensional term

$$\int d^6y\, \mathrm{Tr}_{\mathrm{v}}(\bar{\lambda}\Gamma^m[A_m,\lambda])\,, \tag{17.4.7}$$

becomes

$$d^{\bar{x}\bar{y}\bar{z}}\bar{\lambda}^A{}_{\bar{x}}\lambda^B{}_{\bar{y}}\phi^C{}_{\bar{z}}\int_K \omega_{1,1\,A}\wedge\omega_{1,1\,B}\wedge\omega_{1,1\,C}\,. \tag{17.4.8}$$

Here $d^{\bar{x}\bar{y}\bar{z}}$ is the E_6 invariant for $\overline{\mathbf{27}}\cdot\overline{\mathbf{27}}\cdot\overline{\mathbf{27}}$. The superpotential is then

$$W(\phi) = d^{\bar{x}\bar{y}\bar{z}}\phi^A{}_{\bar{x}}\phi^B{}_{\bar{y}}\phi^C{}_{\bar{z}}\int_K \omega_{1,1\,A}\wedge\omega_{1,1\,B}\wedge\omega_{1,1\,C}\,. \tag{17.4.9}$$

The wedge product of the internal wavefunctions is a $(3,3)$-form and so can be integrated over the internal space without using the metric.

This part of the superpotential is independent of all moduli, and so is topological. To make this explicit, we use the correspondence (17.2.7) between 2-forms and 4-cycles. Take a basic N_A of nontrivial 4-cycles, and let ω_A be the corresponding basis of 2-forms. Three 4-cycles will generically intersect in isolated points. We can therefore define the intersection number, the total number of intersections weighted by orientation; this is a topological invariant.[2] A standard result from topology relates the intersection number of N_A, N_B, and N_C to the integral of the wedge product:

$$\#(N_A,N_B,N_C) = \int_K \omega_{1,1\,A}\wedge\omega_{1,1\,B}\wedge\omega_{1,1\,C}\,. \tag{17.4.10}$$

Thus the superpotential is determined by these integers. Typically many of the intersection numbers vanish for topological reasons unrelated to

[2] If the cycles do not intersect only in isolated points, as is obviously the case if for example two are the same, one can make them do so by deforming them within the same homology class. This then defines the intersection number.

symmetry; this may be useful in understanding the stability of the proton and the rich texture of the Yukawa couplings in the Standard Model. For later reference we mention that there is a dual basis N^A of 2-cycles such that

$$\#(N^A, N_B) = \int_{N^A} \omega_{1,1\,B} = \delta^A{}_B \ . \tag{17.4.11}$$

For the (1,1) Kähler moduli the superpotential is zero. The static Calabi–Yau space solves the field equations for any value of the moduli, so the potential and therefore the superpotential for these vanishes.

Now we consider the Kähler potential, starting with the (1,1) moduli. These have the expansion

$$(g_{i\bar{\jmath}} + b_{i\bar{\jmath}})(x, y) = \sum_A T^A(x)\omega_{Ai\bar{\jmath}}(y) \ . \tag{17.4.12}$$

The four-dimensional kinetic term is obtained from the ten-dimensional kinetic term by inserting this expansion. The result is

$$G_{A\bar{B}} = \frac{1}{V}\int d^6y\,(\det G)^{1/2} G^{i\bar{k}} G^{l\bar{\jmath}} \omega_{Ai\bar{\jmath}} \omega^*_{Bk\bar{l}} \ . \tag{17.4.13}$$

The integral can be related to the one appearing in the superpotential by using the Kähler form $J_{1,1}$ defined in (17.2.16). Parameterize the Kähler moduli space by $h^{1,1}$ complex numbers T^A,

$$J_{1,1} + iB_{1,1} = T^A \omega_{1,1\,A} \ , \quad T^A = v^A + ib^A \ . \tag{17.4.14}$$

Then after some calculation,

$$G_{A\bar{B}} = -\frac{\partial^2}{\partial T^A \partial T^{B*}} \ln W(v) \ , \tag{17.4.15}$$

where $2v^A = T^A + T^{A*}$ and

$$W(v) = \#(N_A, N_B, N_C) v^A v^B v^C = \int_K J_{1,1} \wedge J_{1,1} \wedge J_{1,1} \ . \tag{17.4.16}$$

This is just the superpotential, evaluated at $\phi = v$; it is also equal to the volume of the Calabi–Yau space. The $N = 1$ spacetime supersymmetry requires that this metric be Kähler.[3] The expression (17.4.15) gives the metric on Kähler moduli space directly in Kähler form, with

$$K_1(T, T^*) = -\ln W(v) \ . \tag{17.4.17}$$

Thus the Kähler potential for the moduli is determined in terms of the superpotential W. This is a very special property, which we will see later to be a consequence of the (2,2) world-sheet supersymmetry. The Kähler potential depends on the Kähler moduli and so is not a topological invariant, but it is quasitopological in the sense that its dependence on these moduli is determined by topological data. Note that it is independent of the complex structure moduli, another consequence of the (2,2) world-sheet supersymmetry.

[3] Kähler, Kähler, everywhere. Note that in some places it is the geometry of the compactification that is referred to, while here it is the geometry of the low energy scalar field space.

The metric for the $\overline{27}$ kinetic terms is closely related,

$$G'_{A\bar{B}} = \exp[\kappa_4^2(K_2 - K_1)/3] G_{A\bar{B}} \tag{17.4.18}$$

with K_2 to be defined below. The Kähler potential for these fields is then $G'_{A\bar{B}}\phi^A\phi^{B*}$.

Similar results hold for the $(2,1)$-forms, though the precise statements and the derivations (which are again omitted) are somewhat more intricate. Expand

$$a_{i,jx}(x, y) = \frac{1}{2}\sum_a \chi^a{}_x(x)\omega_{ai\bar{k}\bar{l}}(y)\Omega_j^{\bar{k}\bar{l}}(y) \ , \tag{17.4.19a}$$

$$\lambda_{i,jx}(x, y) = \frac{1}{2}\sum_a \lambda^a{}_x(x)\omega_{ai\bar{k}\bar{l}}(y)\Omega_j^{\bar{k}\bar{l}}(y) \ , \tag{17.4.19b}$$

where a runs over the (1,2)-forms. In this case it is the kinetic term for the moduli that has a simple expression in terms of forms,

$$\begin{aligned} G_{a\bar{b}} &= -\frac{\int_K \omega_{1,2\,a} \wedge \omega^*_{1,2\,a}}{\int_K \Omega_{3,0} \wedge \Omega^*_{3,0}} \\ &= -\frac{\partial}{\partial X^a}\frac{\partial}{\partial X^{\bar{a}}} K_2(X, X^*) \ , \end{aligned} \tag{17.4.20}$$

with

$$K_2(X, X^*) = \ln\left(i\int_K \Omega_{3,0} \wedge \Omega^*_{3,0}\right) \ . \tag{17.4.21}$$

Here X^a are coordinates for the moduli space of complex structures, $a = 1, \ldots, h^{2,1}$, and $X^{\bar{a}} = X^{a*}$.

To relate the superpotential to the Kähler potential it is useful to take *special coordinates* on moduli space. The Betti number b_3 is $2h^{2,1} + 2$. One can always find a basis of 3-cycles

$$\{A^I, B_J\} \ , \quad I, J = 0, \ldots h^{2,1} \tag{17.4.22}$$

such that the intersection numbers are

$$\#(A^I, B_J) = \delta^I{}_J \ , \quad \#(A^I, A^J) = \#(B_I, B_J) = 0 \ . \tag{17.4.23}$$

The corresponding (1,2)-forms are $\alpha_{1,2\,I}$ and $\beta^J_{1,2}$. Thus we can define

$$Z^I = \int_{A^I} \Omega_{3,0} \ . \tag{17.4.24}$$

These $h^{2,1}+1$ complex numbers are one too many to serve as coordinates on the complex structure moduli space. However, there is no natural normalization for Ω, so we must identify

$$(Z^0, Z^1, \ldots, Z^n) \cong (\lambda Z^0, \lambda Z^1, \ldots, \lambda Z^n) \tag{17.4.25}$$

with $n = h^{2,1}$, and this projective space has the correct number of coordinates. The integrals

$$\mathcal{G}_I(Z) = \int_{B_I} \Omega_{3,0} \tag{17.4.26}$$

then cannot be independent variables; for given topology the $\mathcal{G}_I$ are known functions of the Z^J. These can be determined in terms of a single function $\mathcal{G}(Z)$,

$$\mathcal{G}_I = \frac{\partial \mathcal{G}}{\partial Z^I}\,, \quad \mathcal{G}(\lambda Z) = \lambda^2 \mathcal{G}(Z)\,. \tag{17.4.27}$$

The nonprojective coordinates are then $X^a = Z^a/Z^0$ for $a = 1, \ldots, h^{2,1}$.

The Kähler potential (17.4.21) for the complex structure moduli can be expressed in terms of $\mathcal{G}$,

$$K_2(Z, Z^*) = \ln \operatorname{Im}(Z^{I*} \partial_I \mathcal{G}(Z))\,. \tag{17.4.28}$$

So also can the superpotential,

$$W(Z, \chi) = \frac{\chi^a \chi^b \chi^c}{3!} \frac{\partial^3 \mathcal{G}(Z)}{\partial Z^a \partial Z^b \partial Z^c}\,. \tag{17.4.29}$$

The matter metric is again slightly different from that for the moduli,

$$G'_{a\bar{b}} = \exp[\kappa_4^2 (K_1 - K_2)/3] G_{a\bar{b}}\,. \tag{17.4.30}$$

The intersection numbers (17.4.23) are invariant under a symplectic change of basis,

$$\begin{bmatrix} A'^I \\ B'_J \end{bmatrix} = S \begin{bmatrix} A^I \\ B_J \end{bmatrix} \tag{17.4.31}$$

for $S \in Sp(h^{2,1}+1, \mathbf{Z})$. The new coordinates

$$\begin{bmatrix} Z'^I \\ \mathcal{G}'_J \end{bmatrix} = S \begin{bmatrix} Z^I \\ \mathcal{G}_J \end{bmatrix} \tag{17.4.32}$$

are then another set of special coordinates for the same moduli space.

To summarize, the low energy effective action is determined in terms of two holomorphic functions, $W(T)$ and $\mathcal{G}(Z)$. Each of these is determined in turn by the topology of the Calabi–Yau manifold and can be calculated by well-developed methods from analytic geometry. Notice that the actual Ricci-flat metric is never used — a good thing, as the explicit form is not known in any nontrivial example.

17.5 Higher corrections

Thus far we have considered only the leading term in an expansion in α'/R_c^2. We now consider the corrections, remaining in this chapter at the string tree level. The ten-dimensional action derived from string theory has an infinite series of higher derivative corrections, each derivative accompanied by $\alpha'^{1/2}$. These terms can be deduced from the momentum expansion of the tree-level scattering amplitudes. Alternatively, they can be obtained from the higher loop corrections to the world-sheet beta functions, where again the expansion parameter is α'/R_c^2. The supersymmetry transformations are given by a similar series.

The most immediate questions would seem to be whether the Calabi–Yau manifolds solve the full field equations, and whether they remain supersymmetric. Actually they do not. They continue to solve the field equations when the terms quadratic and cubic in the curvatures and field strengths are included in the action, but with the inclusion of the quartic terms (corresponding to four loops in the world-sheet sigma model) they in general do not. However, this is not really the right question. Rather, as in any perturbation theory, we need to know whether the solution can be corrected order-by-order so as to solve the field equations at each order. It is not trivial that this is possible — as in other forms of perturbation theory there is a danger of vanishing denominators — but it has been shown to be possible from an analysis of the detailed form of the corrections to the beta functions.

Remarkably, this same result obtained by a rather technical world-sheet argument can be obtained much more easily and usefully from an analysis of the spacetime effective action — a common theme in supersymmetric theories. First note that regardless of whether the Calabi–Yau space can be corrected to give an exact solution, we can still study the physics for nearby configurations by the method of the previous section. Expand the fields as background plus fluctuation, separate the fluctuations into light and heavy, and integrate out the heavy fields to obtain an effective action for the light fields. Corrections in the α'/R_c^2 expansion give additional terms in the low energy action. Now, an important point is that any mass scale appearing in these terms will be the compactification energy R_c^{-1} times a power of the small parameter α'/R_c^2. Thus there is still a clean separation between the light and heavy fields, and it makes sense to discuss the effective action for the former.

The final key point is that this low energy effective action must be supersymmetric. Because the full theory is supersymmetric, any breaking must be spontaneous rather than explicit. To see this another way, note that as $\alpha'/R_c^2 \to 0$ with R_c fixed, supersymmetry is restored and so the

gravitino becomes massless; near this limit it remains one of the light fields. However, the only consistent theory of light spin-$\frac{3}{2}$ particles is spontaneously broken supergravity.

We will see in the next chapter that the gauge kinetic term receives no corrections at string tree level to any order of α'/R_c^2. All corrections to the low energy effective action must then appear in the Kähler potential or the superpotential. We can now state the criterion for the corrections to spoil the solution: they must produce a correction to superpotential that depends only on the moduli, $\delta W(T,Z)$. In this case there will be a potential for the moduli, which at general points will not be stationary so that most or all of the previous static supersymmetric solutions are gone.

Now let us argue that this is impossible. Consider how a string amplitude depends on the moduli b^A for the $B_{i\bar{j}}$ background, eq. (17.4.14). This background enters into the string amplitude as

$$\frac{1}{2\pi\alpha'}\int_M B_{1,1} = \frac{n_A b^A}{2\pi\alpha'} . \tag{17.5.1}$$

Since the background $B_{1,1}$ form is closed, the integral depends only on the topology of the embedding of the world-sheet M in spacetime. The embedding is equivalent to a sum $n_A N^A$ of generators of $H_2(K)$, and so the integral follows as in eq. (17.4.11). Now, world-sheet perturbation theory is an expansion around the configuration $X^\mu(\sigma) =$ constant, which is topologically trivial. To all orders of perturbation theory, $n_A = 0$ and the amplitudes are independent of b^A. There are thus $h_{1,1}$ symmetries

$$T^A \to T^A + i\epsilon^A . \tag{17.5.2}$$

Since the superpotential must be holomorphic in T^A, this implies that it is actually independent of T^A.

To obtain a nonrenormalization theorem, let us write

$$T^A = c^A T \tag{17.5.3}$$

with the c^A fixed complex numbers, and focus on the dependence on T. Varying T at fixed c^A rescales $G_{i\bar{j}}$ and so scales the size of K while holding its shape fixed. Thus the world-sheet perturbation expansion parameter is T^{-1}. Since the superpotential is holomorphic in T, it can receive no corrections in world-sheet perturbation theory.

Thus the terms that might destabilize the vacuum and break supersymmetry cannot be generated, and so the Calabi–Yau solution can be perturbatively corrected to all orders to give a static supersymmetric background. A potential could also be generated by a Fayet–Iliopoulos term in the more general case that the gauge group includes a $U(1)$ factor. A separate argument excludes this; we postpone the discussion to section 18.7. It also follows that the superpotential for the matter fields, though calculated above by means specific to the $\alpha'/R_c^2 \to 0$ limit, is exact to all orders in α'/R_c^2.

Ordinarily the Kähler potential does not have similar nonrenormalization properties because it is not holomorphic, and so could have an arbitrary dependence on $T + T^*$. However, the presence of (2,2) superconformal symmetry on the world-sheet puts strong additional constraints on the theory. The action obtained in the field-theory approximation of the previous section had two notable properties. First, there was no superpotential for the moduli. Second, the full low energy action was determined by two holomorphic functions, one depending only on the Kähler moduli and the other only on the complex structure moduli. We will see in chapter 19 that these properties actually follow from (2,2) world-sheet supersymmetry and so are exact properties of the string tree-level action in Calabi–Yau compactification. The nonrenormalization of the superpotential then implies the same for the Kähler potential, and the full effective action found in the field-theory approximation is exact to all orders in α'/R_c^2, except for one term to be discussed in chapter 19.

For future reference, note that the Kähler potential (17.4.16) for the overall scale T is

$$-3\ln(T + T^*) , \tag{17.5.4}$$

up to instanton corrections that are exponentially small in T.

Instanton corrections

The discussion above does not exclude the possibility of corrections that are nonperturbative on the string world-sheet. Indeed, these do break the shift symmetries (17.5.2). Consider a *world-sheet instanton*, meaning a topologically nontrivial embedding of the world-sheet in spacetime: the string world-sheet wraps around some noncontractible surface in spacetime. The n_A defined in eq. (17.5.1) are then nonzero and the amplitude depends on b^A, breaking the $T^A \to T^A + i\epsilon^A$ symmetry.

To see whether these can affect the superpotential, compare the Polyakov action

$$\frac{1}{2\pi\alpha'}\int d^2z\, G_{i\bar{j}}(\partial_z Z^i \partial_{\bar{z}} Z^{\bar{j}} + \partial_z Z^{\bar{i}} \partial_{\bar{z}} Z^j) \tag{17.5.5}$$

to

$$\frac{1}{2\pi\alpha'}\int_M J_{1,1} = \frac{1}{2\pi\alpha'}\int d^2z\, G_{i\bar{j}}(\partial_z Z^i \partial_{\bar{z}} Z^{\bar{j}} - \partial_z Z^{\bar{i}} \partial_{\bar{z}} Z^j) = \frac{n_A v^A}{2\pi\alpha'} . \tag{17.5.6}$$

The two terms in the action (17.5.5) are nonnegative, so the action is bounded below by $|n_A v^A|/2\pi\alpha'$. When this bound is attained, then either $\partial_{\bar{z}} Z^i$ or $\partial_{\bar{z}} Z^{\bar{\imath}}$ vanishes, and the embedding of the world-sheet in spacetime is a *holomorphic instanton*. In this case, the Polyakov action and the coupling to B combine to give the path integral factor

$$\exp(-n_A T^A/2\pi\alpha') \tag{17.5.7}$$

or its conjugate. This is holomorphic in T^A and so can appear in the superpotential: holomorphic instantons can, and do, correct the superpotential. In particular they correct the cubic terms in $\overline{\mathbf{27}}^3$. By (2,2) supersymmetry they also then correct the metric for the (1,1)-forms. However, study of the detailed form of the instanton amplitudes, in particular the fermion zero modes, shows that they cannot generate a superpotential for the moduli fields alone and so do not destabilize the solution. Again, this will be understood later as a consequence of (2,2) symmetry: no superpotential for the moduli can be generated.

Instantons cannot correct the metric for the (1,2)-forms. Instanton corrections depend as in eq. (17.5.7) on the (1,1) modulus T, and the metric for the (1,2)-forms cannot depend on (1,1) moduli (another consequence of (2,2) supersymmetry). They cannot then correct the $\overline{\mathbf{27}}^3$ superpotential either. The low energy action for the (1,2)-forms, though obtained in the field theory limit, is exact at string tree level. The low energy action for the (1,1)-forms receives instanton corrections.

17.6 Generalizations

Let us now consider the E_6 singlets from $H^1(\mathrm{End}\, T)$. In particular, are there flat directions for these fields? Also, are there flat directions for the charged fields in the **27**s and $\overline{\mathbf{27}}$s? In each case the massless fields originate from the compact components of the ten-dimensional gauge field, so flat directions would correspond to varying the gauge field away from the 'spin connection = gauge connection' form assumed so far. The Bianchi identity (17.1.13) then implies that generically the torsion $\tilde{H}$ must be nonvanishing, so these flat directions take us outside the vanishing-torsion ansatz with which we began. Also, with the spin connection unrelated to the gauge connection there are in general no longer any left-moving supersymmetries on the world-sheet, and the world-sheet supersymmetry is reduced to (0, 2). We will thus refer to the fields parameterizing these potential flat directions as (0,2) moduli.

The analysis of the general solution is somewhat more intricate than for vanishing torsion, but again there are existence theorems to the effect that under appropriate topological conditions solutions exist in the field theory limit. The nonrenormalization theorem above still applies, so that these remain solutions to all orders of world-sheet perturbation theory. Nonperturbatively, instantons in (0, 2) backgrounds have fewer fermion zero modes and there is no general argument forbidding a superpotential for the (0,2) moduli. Initially it was believed that a superpotential would generically appear and destabilize most (0,2) vacua. However, it is now known that many of the (0,2) directions are exactly flat, so that the typical (2,2) moduli space is embedded in a larger moduli space of (0,2) theories. There are likely also moduli spaces of (0,2) theories that are not connected to any (2,2) theories.

The understanding of (0,2) theories is much less complete than for (2,2) theories, and the analysis of them is more intricate. We will therefore not discuss them in any detail, though some of the methods to be developed in chapter 19 for (2,2) theories are also useful in the (0,2) case.

We would like to mention briefly some phenomenological features of the (0,2) vacua. We have emphasized that the (2,2) theories look much like a grand unified Standard Model, with an E_6 gauge group and matter in the **27**. Under the $SU(3)\times SU(2)\times U(1)$ subgroup of E_6, the **27** contains 15 states with chiral gauge couplings, having the precise quantum numbers of a generation of quarks and leptons, and 12 with parity-symmetric couplings. The latter can have $(SU(3)\times SU(2)\times U(1))$-invariant mass terms and so can be much more massive than the weak scale. Indeed, these extra states in the **27** can mediate baryon decay, so they must be much heavier than the weak scale. Other arguments based on the running of the gauge couplings and the lightness of the Standard Model neutrinos also suggest that the extra states are quite massive. In addition to the extra states within each **27**, typical (2,2) theories have both **27**s and $\overline{\mathbf{27}}$s, which from the low energy point of view correspond to generations with left- and right-handed weak interactions. Although it is possible that some right-handed 'mirror generations' exist near the weak scale, this seems unlikely for a number of reasons. Fortunately the gauge symmetry allows a **27** and a $\overline{\mathbf{27}}$ to pair up and become massive.

Although these various masses are allowed by the low energy gauge symmetry, we need a specific mechanism for generating them. As long as we stay within the (2,2) theories, even adding Wilson lines to break the E_6 symmetry, the general properties of these theories guarantee that the quantum numbers of the low energy fields add up to complete multiplets of E_6. However, along the (0,2) directions the extra states can become massive. In addition to the $\mathbf{27}^3$ and $\overline{\mathbf{27}}^3$ terms already discussed, the lowest order superpotential contains terms $\mathbf{1}^3$ and $\mathbf{1}\cdot\mathbf{27}\cdot\overline{\mathbf{27}}$, and together these have the potential to generate all the needed masses.

The (0,2) moduli from the **27**s and $\overline{\mathbf{27}}$s may be useful for another related reason. In most examples, such as those discussed in section 17.2,

the freely-acting discrete symmetry is Abelian, for example $\mathbf{Z}_3$. The Wilson lines must have the same algebra as the space group, so they commute and can be taken to lie in a $U(1)^6$ subalgebra of E_6. Since the low energy group is that part of E_6 that commutes with the Wilson lines, it contains at least this $U(1)^6$ and so has rank 6, as compared to the rank 4 of the Standard Model: the closest we can come in this way to the Standard Model is $SU(3) \times SU(2) \times U(1)^3$. The additional $U(1)$s might be broken somewhat above the weak scale, but again there are problems; in particular the extra 12 states in the **27** are chiral under the additional $U(1)$s so this prevents them from becoming very massive. One way to break these symmetries is to twist by a *non-Abelian* discrete group, but another is to give expectation values to (0,2) moduli from the **27** and $\overline{\mathbf{27}}$. Clearly this breaks some of the E_6 symmetry, and in fact it necessarily reduces the rank of the gauge group. The group theory in section 11.4 shows that the **27** contains two singlets of $SU(3) \times SU(2) \times U(1)$. If both of these have expectation values they break E_6 to the minimal grand unified group $SU(5)$, and combined with Wilson line breaking this can give the Standard Model gauge group.

Thus it is an attractive possibility that our vacuum is given by turning on some of the (0,2) moduli of a Calabi–Yau compactification. We should emphasize, however, that if one considers the large set of (0,2) theories that can be constructed by asymmetric orbifolds or free fermions, only a small subset of these have a close resemblance to the Standard Model.

No exercises

The nature of this chapter, all results and no derivations, does not lend itself to exercises. The reader who wishes to learn more should consult the references.

18
Physics in four dimensions

We have now studied two kinds of four-dimensional string theory, based on orbifolds and on Calabi–Yau manifolds. We saw that the low energy physics of the weakly coupled heterotic string resembles a unified version of the Standard Model rather well. In this chapter we present general results, valid for any compactification. In most of this chapter we are concerned with weakly coupled heterotic string theories, but at various points we will discuss how the results are affected by the new understanding of strongly coupled strings.

18.1 Continuous and discrete symmetries

An important result holding in all string theories is that *there are no continuous global symmetries; any continuous symmetries must be gauged.* We start with the bosonic string. Associated with any symmetry will be a world-sheet charge

$$Q = \frac{1}{2\pi i} \oint (dz\, j_z - d\bar{z}\, j_{\bar{z}}) \,. \tag{18.1.1}$$

This is to be a symmetry of the physical spectrum and so it must be conformally invariant. Thus j_z transforms as a $(1,0)$ tensor and $j_{\bar{z}}$ as a $(0,1)$ tensor. We can then form the two vertex operators

$$j_z \bar{\partial} X^\mu e^{ik\cdot X} \,, \quad \partial X^\mu j_{\bar{z}} e^{ik\cdot X} \,. \tag{18.1.2}$$

These create massless vectors coupling to the left- and right-moving parts of the charge Q. Thus the left- and right-moving parts of Q each give rise to a spacetime gauge symmetry. If Q is carried only by fields moving in one direction, then only one of the currents and only one of the vertex operators is nonvanishing. Turning the construction around, any local symmetry in spacetime gives rise to a global symmetry on the world-sheet.

For type I or II strings the same argument holds immediately if we use superspace, writing

$$Q = \frac{1}{2\pi i}\oint (dz d\theta\, J - d\bar{z} d\bar{\theta}\, \tilde{J}) \,. \tag{18.1.3}$$

Superconformal invariance requires that J be a $(\frac{1}{2},0)$ tensor superfield and $\tilde{J}$ a $(0,\frac{1}{2})$ tensor superfield. Combined with $\tilde{\psi}^\mu$ or ψ^μ respectively, these give gauge boson vertex operators, so again this is a gauge symmetry in spacetime. The same is true for the heterotic string, using the bosonic argument on one side and the supersymmetric argument on the other.

The absence of continuous global symmetries has often been imposed as an aesthetic criterion by model builders in field theory, and we see that it is realized in string theory. There is a slight loophole in the argument, which we will discuss later in the section.

We have seen in the examples from earlier chapters that string theories generally have discrete symmetries at special points in moduli space. It is harder to generalize about whether these are local or global symmetries because the difference is subtle for a discrete symmetry: there is no associated gauge boson in the local case. The meaning of a discrete local symmetry was discussed in section 8.5 in the context of the field theory on the world-sheet. The simplest way to verify that a discrete symmetry is local is to find a point in moduli space where it is enlarged to a continuous gauge symmetry. For example, this is the case for the T-duality of the bosonic and heterotic strings. To see what this would mean, consider a spacetime with x^8 and x^9 periodic, with the radius R_8 a function of x^9. Then $R_8(x^9)$ need not be strictly periodic; rather, it could also be that

$$R_8(2\pi R_9) = \alpha'/R_8(0) \,. \tag{18.1.4}$$

This is the essence of a discrete gauge symmetry: that on nontrivial loops fields need be periodic only up to a gauge transformation. Since T-duality is embedded in the larger U-duality of the type II theory, the latter must be a gauge symmetry as well. Thus we could have a similar aperiodicity in the IIB string coupling, for example:

$$\Phi(2\pi R_9) = -\Phi(0) \,, \quad g(2\pi R_9) = 1/g(0) \,. \tag{18.1.5}$$

It is not clear that this is true of all discrete symmetries in string theory, but it seems quite likely.

P, C, T, and all that

We would like to discuss briefly the breaking of the discrete spacetime symmetries P, C, and T in string theory.

Parity symmetry P is invariance under reflection of any one coordinate, say $X^3 \to -X^3$. It is not a good symmetry of the Standard Model, being violated by the gauge interactions. Classifying particles moving in the 1-direction by the helicity $\Sigma^{23} = s_1$, the helicity $+\frac{1}{2}$ states form some gauge representation r_+, and the helicity $-\frac{1}{2}$ states some representation r_-. Parity takes the helicity $s_1 \to -s_1$, and so is a good symmetry only if $r_+ = r_-$. In the Standard Model it appears (barring the discovery of new massive states with the opposite gauge couplings) that $r_+ \neq r_-$: the gauge couplings are *chiral*.

Let us consider the situation in string theory, starting with the ten-dimensional heterotic string. In ten dimensions states are labeled by their $SO(8)$ representation. Parity again reverses the spinor representations **8** and **8′**, and is a good symmetry only if the corresponding gauge representations are the same, $r = r'$. For the heterotic string, r is the adjoint representation while r' is empty, so the gauge couplings are chiral and there is no parity symmetry. To see how this arises, note that the heterotic string action and world-sheet supercurrent (or BRST charge) are invariant if we combine the reflection $X^3 \to -X^3$ with $\psi^3 \to -\psi^3$. However, this also flips the sign of $\exp(\pi i \tilde{F})$ in the R sector, and so it is not a symmetry of the theory because the GSO projection restricts the spectrum to $\exp(\pi i \tilde{F}) = +1$.

Although the ten-dimensional spectrum is chiral, compactification to four dimensions can produce a nonchiral spectrum. This is true of toroidal compactification, for example, as one sees from the discussion in section 11.6. The point is that the theory is invariant under simultaneous reflection of one spacetime and one internal coordinate, say X^3 and X^9, as well as their partners ψ^3 and ψ^9. This is a symmetry of the action, supercurrent, and GSO projection, and so of the full theory. From the ten-dimensional point of view, it is a rotation by π in the (3,9) plane, but from the four-dimensional point of view it is a reflection of the 3-axis, combined with an internal action which gives negative intrinsic parity to the 9-oscillators. This symmetry reverses the momenta $k^9_{R,L}$, which are the charges under the corresponding Kaluza–Klein gauge symmetries, while leaving the other internal momenta invariant. Strictly speaking, it is therefore not a pure parity operation (which by the usual definition leaves gauge charges invariant) or a CP transformation (which inverts all charges), but something in between.

In the $\mathbf{Z}_3$ orbifold example, the spectrum was found to be chiral. The orbifold twist removes all parity symmetries. Notice that simultaneous reflection of $X^{3,5,7,9}$, which takes $Z^i \leftrightarrow Z^{\bar{\imath}}$, satisfies $Pr = r^2 P$ and so commutes with the twist projection. However, to extend this action to the various spinor fields requires that P reflect $\psi^{3,5,7,9}$ and $\lambda^{2,4,6}$ as well. This

acts on an odd number of the λ fermions and so does not commute with the current algebra GSO projection. The combined effect of the orbifold twist and the ψ and λ GSO projections removes all parity symmetry and leaves a chiral spectrum. Chiral gauge couplings arise in many other kinds of string compactification.

There is one interesting general remark. The chirality of the spectrum can be expressed in terms of a mathematical object known as an *index*. Separate $\exp(\pi i\tilde{F})$ into a spacetime part and an internal part, $\tilde{F} = \tilde{F}_4 + \tilde{F}_K$. For massless fermions moving in the 1-direction, $2s_1 = -i\exp(\pi i\tilde{F}_4)$, which in turn is equal to $i\exp(\pi i\tilde{F}_K)$ due to the GSO projection. For massless R sector states the internal part is annihilated by G_0, so the net chirality (number of helicity $+\frac{1}{2}$ states minus helicity $-\frac{1}{2}$ states) in a given irreducible representation r is

$$N_{+\frac{1}{2},r} - N_{-\frac{1}{2},r} = \mathrm{Tr}_{r,\ker(G_0)}[i\exp(\pi i\tilde{F}_K)] \,, \tag{18.1.6}$$

the trace running over all states in the internal CFT which are in the representation r and are annihilated by G_0. One can now drop the last restriction on the trace,

$$N_{+\frac{1}{2},r} - N_{-\frac{1}{2},r} = \mathrm{Tr}_r[i\exp(\pi i\tilde{F}_K)] \,. \tag{18.1.7}$$

The point is that any state $|\psi\rangle$ with a positive eigenvalue v under G_0^2 is always paired with a state $G_0|\psi\rangle$ of opposite $\exp(\pi i\tilde{F}_K)$, so these states make no net contribution to the trace. The state $G_0|\psi\rangle$ cannot vanish because $G_0G_0|\psi\rangle = v|\psi\rangle$.

Such a trace is known as an index: this can be defined whenever one has a Hermitean operator G_0 anticommuting with a unitary operator $\exp(\pi i\tilde{F}_K)$. The index has the important property that it is invariant under continuous changes of the CFT. Under such a change, the eigenvalues v of G_0^2 change continuously, but the trace of $\exp(\pi i\tilde{F}_K)$ at $v = 0$ remains invariant because states can only move away from $v = 0$ in pairs with opposite $\exp(\pi i\tilde{F}_K)$. This invariance can also be understood from the spacetime point of view: a continuous change in the background fields can give mass to some previously massive states, but to make a massive representation one must combine states of opposite helicity.[1] Using this invariance, the index may often be calculated by deforming to a convenient limit. There is one subtlety that comes up in some examples: the index may change in certain limits due to states running off to infinity in field space.

Charge conjugation C leaves spacetime invariant but conjugates the gauge generators. In the Standard Model this is again broken by the

[1] This is one of those statements that, surprisingly, need no longer hold at strong coupling. We will discuss this further in sections 19.7 and 19.8.

gauge couplings of the fermions. For C invariance to hold, the fermion representations must satisfy $r_+ = \bar{r}_+$ and $r_- = \bar{r}_-$. CPT invariance, to be discussed below, implies that $r_+ = \bar{r}_-$ so that chiral gauge couplings violate C as well as P. Thus the orbifold example also violates C.

The combination CP takes $r_+ \to \bar{r}_-$ and so is automatically a symmetry of the gauge couplings as a consequence of CPT. In the Standard Model Lagrangian, CP is broken by phases in the fermion–Higgs Yukawa couplings. In the $\mathbf{Z}_3$ orbifold example, the transformation that reverses $X^{3,5,7,9}$, $\psi^{3,5,7,9}$, and all of the λ^I for I odd is a symmetry of the action, the BRST charge, and all projections. From the point of view of the four-dimensional theory this is CP, because the action on the λ^I changes the sign of all the diagonal generators, which is charge conjugation. The $\mathbf{Z}_3$ orbifold is thus CP-invariant. However, recall that there were many moduli. These included the flat metric background $G_{i\bar{j}}dZ^i d\bar{Z}^{\bar{j}}$. The operation CP takes $G_{i\bar{j}} \to G_{\bar{i}j}$. Reality of the metric requires $G_{i\bar{j}}$ to be Hermitean, while CP requires it to be real. The generic Hermitean $G_{i\bar{j}}$ is not real, so CP is broken almost everywhere in moduli space. One must also consider other possible CP operations, such as adding discrete rotations of some of the Z^i, or permutations of the Z^i, to the transformation. These will be symmetries at special points in moduli space, but are again broken generically. This is also true for most other string compactifications: there will be CP-invariant vacua, but some of the many moduli will be CP-odd so that CP-invariance is spontaneously broken at generic points.

It is interesting to note that CP, like the discrete symmetries discussed earlier, is a gauge symmetry. The operation described above can be thought of as rotations by π in the (3,5) and (7,9) planes, combined with a gauge rotation. These are all part of the local symmetry of the ten-dimensional theory, though this is partly spontaneously broken by the compactification.

In local, Lorentz-invariant, quantum field theory the combination CPT is always an exact symmetry. It is easy to show that CPT is a symmetry of string perturbation theory, using essentially the same argument as is used to prove the CPT theorem in field theory. Consider the operation θ that reverses $X^{0,3}$ and $\psi^{0,3}$. If we continue to Euclidean time this is just a rotation by π in the (iX^0, X^3) plane and so is obviously a symmetry. The analytic continuation is well behaved because $X^{0,3}$ and $\psi^{0,3}$ are free fields. Clearly θ includes parity and time-reversal. To see that it also implies charge conjugation, recall that a vertex operator $\mathscr{V}$ with $k^0 < 0$ creates a string in the initial state, while a vertex operator with $k^0 > 0$ destroys a string in the final state. If $\mathscr{V}$ carries some charge q it creates a string of charge q. The operation θ does not act on the charges, so $\theta \cdot \mathscr{V}$ also has charge q and so destroys a string of charge $-q$. Thus, θ takes a string in the in-state to the C-, P-, and T-reversed string in the out-state.

To make this slightly more formal, recall from section 9.1 that the

S-matrix is given schematically by

$$\langle \alpha, \text{out}|\beta, \text{in}\rangle = \left\langle \overline{\mathscr{V}}_\alpha \mathscr{V}_\beta \right\rangle , \tag{18.1.8}$$

where to be concise we have only indicated one vertex operator in each of the initial and final states. Then by acting with θ this becomes

$$\langle \alpha, \text{out}|\beta, \text{in}\rangle = \left\langle \theta\cdot\overline{\mathscr{V}}_\alpha \theta\cdot\mathscr{V}_\beta \right\rangle = \langle \theta\bar{\beta}, \text{out}|\theta\bar{\alpha}, \text{in}\rangle . \tag{18.1.9}$$

The CPT operation is antiunitary,

$$\langle CPT\cdot\beta, \text{out}|CPT\cdot\alpha, \text{in}\rangle = \langle \alpha, \text{out}|\beta, \text{in}\rangle , \tag{18.1.10}$$

so we see that CPT is θ combined with the conjugation of the vertex operator.

This argument is formulated in string perturbation theory. Elsewhere we have encountered results that hold to all orders of perturbation theory but are spoiled by nonperturbative effects. Without a nonperturbative formulation of string theory we cannot directly extend the CPT theorem, but we can 'prove' it by the strategy that we have used elsewhere: assert that the low energy physics of string theory is governed by quantum field theory, and then cite the CPT theorem from the latter. Still, there may be surprises; we can hope that when string theory is better understood it will make some distinctive non-field-theoretic prediction for observable physics.

The spin-statistics theorem is often discussed alongside the CPT theorem. The discussion in section 10.6 for free boson theories is easily generalized. Consider a basis of Hermitean (1,1) operators $\mathscr{A}_i$ with definite Σ^{01} eigenvalue s_0 and $\beta\gamma$ ghost number q. Now consider the OPE of such an operator with itself. In any unitary CFT, a simple positivity argument shows that the leading term in the OPE of a Hermitean operator with itself is the unit operator. Then

$$\mathscr{A}_i(z,\bar{z})\mathscr{A}_i(0,0) \sim (z\bar{z})^{-2}\bar{z}^{2(q+q^2-s_0^2)}\exp(2q\tilde{\phi} + 2is_0\tilde{H}_0) , \tag{18.1.11}$$

where the z- and $\bar{z}$-dependence follows from the weight $\tilde{h} = 2(q+q^2-s_0^2)$ of the exponential. For NS states, with integer spacetime spin, s_0 and q are integers, while for R states, with half-integer spacetime spin, they are half-integer. It follows that the operator product (18.1.11) is symmetric in the NS sector and antisymmetric in the R sector. The spacetime spin is thus correlated with world-sheet statistics, and the spacetime spin-statistics theorem then follows as in section 10.6. Again this is a rather narrow and technical way to establish this result.

The strong CP problem

In the Standard Model action CP violation can occur in two places, the fermion–Higgs Yukawa couplings and the theta terms

$$S_\theta = \frac{\theta}{8\pi^2}\int \text{Tr}(F_2 \wedge F_2) . \tag{18.1.12}$$

This is θ times the instanton number, the trace normalized to the **n** of $SU(n)$. For the weak $SU(2)$ and $U(1)$ gauge interactions the fluctuations of the gauge field are small and the effect of S_θ is negligible, but for the strongly coupled $SU(3)$ gauge field the nontrivial topological sectors make significant contributions. The result is CP violation proportional to θ in the strong interactions. The limits on the neutron electron dipole moment imply that

$$|\theta| < 10^{-9} . \tag{18.1.13}$$

The CP-violating phases in the fermion–Higgs couplings are known from kaon physics not to be much less than unity. Understanding the small value of θ is the *strong CP problem.*

One proposed solution, *Peccei–Quinn (PQ) symmetry,* is automatically incorporated in string theory. In eq. (16.4.13) we found the coupling

$$\frac{1}{2g_4^2}\int aF_2^a \wedge F_2^a . \tag{18.1.14}$$

Aside from this term, the action is invariant under

$$a \to a + \epsilon , \tag{18.1.15}$$

known as PQ symmetry. The field a, which would be massless if the symmetry (18.1.15) were exact, is the axion. The axion and the θ-parameter appear only in the combination $\theta + 8\pi^2 a/g_4^2$, so θ has no physical effect: it can be absorbed in a redefinition of a. The effective physical value θ_{eff} is $\theta + 8\pi^2\langle a\rangle/g_4^2$. The strong interaction produces a potential for a, which is minimized precisely at $\theta_{\text{eff}} = 0$ because at this point the various contributions to the path integral add coherently. The weak interactions induce a nonzero value, but this is acceptably small.

The axion a is known as the *model-independent axion* because the coupling (18.1.14) is present in every four-dimensional string theory: the amplitude with one $B_{\mu\nu}$ vertex operator and two gauge vertex operators does not depend on the compactification. Unfortunately, the model-independent axion may not solve the strong CP problem. There are likely to be several non-Abelian gauge groups below the string scale. Low energy string theories typically have hidden gauge groups larger than $SU(3)$, and the corresponding strong interaction scales are $\Lambda_{\text{hidden}} > \Lambda_{\text{QCD}}$. We will see later in the chapter that this is a likely source of supersymmetry breaking.

The model-independent axion couples to all gauge fields. The gauge group with the largest scale Λ gives the largest contribution, so that the axion sets the θ-parameter for *that* gauge group approximately to zero. In a CP-violating theory, the θ-parameters for the different gauge groups will in general differ, so that $\theta_{\rm QCD}$ remains large. Nonperturbative effects at the string scale may also contribute to the axion potential.

Another difficulty is cosmological. The axion a, being closely related to the graviton and dilaton, couples with gravitational strength κ. In other words, the *axion decay constant* is close to the Planck length. A decay constant this small leads to an energy density in the axion field today that is too large; it takes a rather nonstandard cosmology to evade this bound.

Both problems might be evaded if there were additional axions with appropriate decay constants. In Calabi–Yau compactifications there are shift symmetries (17.5.2) of the $B_{i\bar{j}}$ background, $T^A \to T^A + i\epsilon^A$. Further, the threshold corrections discussed in section 16.4 induce the coupling (18.1.14) to the gauge fields. However, these are only approximate PQ symmetries, because world-sheet instantons generate interactions proportional to

$$\exp(-n_A T^A/2\pi\alpha') = \exp[-n_A(v^A + ib^A)/2\pi\alpha'] \,. \tag{18.1.16}$$

These spoil the PQ symmetries and generate masses for the axions b^A. There is some suppression if $v^A/2\pi\alpha'$ is large, and possibly additional suppression from light fermion masses, which appear in relating the instanton amplitudes to the actual axion mass. However, the suppression must be very large, so that the axion mass from this source is well below the QCD scale, if this is to solve the strong CP problem.

In the type I and II theories the scalars from the R–R sector are also potential axions. As discussed in section 12.1, their amplitudes vanish at zero momentum, implying a symmetry $C \to C + \epsilon$ for each such scalar. In addition they can have the necessary couplings to gauge fields. They receive mass from D-instanton effects.

In summary we have potentially three kinds of axion — model-independent, $B_{i\bar{j}}$, and R–R — which receive mass from three kinds of instanton — field theory, world-sheet, and Dirichlet. Not surprisingly, one can show that these are related by various string dualities. It may be that in some regions of parameter space the axions are light enough to solve the strong CP problem. There may also be additional approximate PQ symmetries from light fermions coupling to some of the strong groups. Or it may be that the solution to the strong CP problem lies in another direction, depending on details of the origin of CP violation.

Incidentally, these PQ symmetries are continuous global symmetries, seemingly violating the result obtained earlier. The loophole is that the world-sheet charge Q *vanishes* in each case — strings do not carry any of the PQ charges. We know this for the R–R charges; for the others it follows because the axion vertex operator at zero momentum is a total derivative. However, since in each case these are not really symmetries, being violated by the various instanton effects, the general conclusion about continuous global symmetries evidently still holds.

The arguments thus far are based on our understanding of perturbative string theory, but it is likely that the conclusion also holds at strong coupling. If a symmetry is exact at large g, it remains a symmetry as g is taken into the perturbative regime, since this is just a particular point in field space. At weak coupling it can then take one of two forms. It could be visible in string perturbation theory, meaning that it holds at each order of perturbation theory; it is then covered by the above discussion. Or, it could hold only in the full theory; the duality symmetries are of this type, but these are all discrete symmetries.

18.2 Gauge symmetries

Gauge and gravitational couplings

In sections 12.3 and 12.4 we obtained the relation between the gauge and gravitational couplings of the heterotic string in ten dimensions:

$$g_{10}^2 = \frac{4\kappa_{10}^2}{\alpha'} \,. \tag{18.2.1}$$

If we compactify, then by the usual dimensional reduction

$$g_4^2 = g_{10}^2/V \,, \quad \kappa_4^2 = \kappa_{10}^2/V \,, \tag{18.2.2}$$

with V the compactification volume. The relation between the parameters in the four-dimensional action is then the same,

$$g_4^2 = \frac{4\kappa_4^2}{\alpha'} \,. \tag{18.2.3}$$

Also, the actual physical values of the couplings depend on the dilaton as[2] e^{Φ_4}, but this enters in the same way on each side so that

$$g_{\rm YM}^2 = \frac{4\kappa^2}{\alpha'} \,. \tag{18.2.4}$$

This derivation is valid only in the field-theory limit, but with one generalization it holds for any four-dimensional string theory. For gauge bosons

[2] When $\langle\Phi_4\rangle \neq 0$, the rescaling (16.4.6) changes the background value of the metric. To study the physics in a given background, as we are doing in this chapter, one should instead rescale

$$G'_{\mu\nu\,\rm Einstein} = \exp[-2(\Phi_4 - \langle\Phi_4\rangle)]G_{\mu\nu} \,,$$

and the coefficient of the gravitational action is then the physical coupling $\kappa = \exp(\langle\Phi_4\rangle)\kappa_4$.

with polarizations and momenta in the four noncompact directions, the explicit calculation (12.4.13) of the three-gauge-boson amplitude involves only the four-dimensional and current algebra fields and so is independent of the rest of the theory. The only free parameter is the parameter $\hat{k}$ from the current algebra, which appeared in the three-gauge-boson amplitude as $\hat{k}^{-1/2}$. Thus the general result is

$$g_{\rm YM}^2 = \frac{4\kappa^2}{\hat{k}\alpha'} \ . \tag{18.2.5}$$

For completeness[3] let us recall that $\hat{k}$ is the coefficient of $z^{-2}\delta^{ab}$ in the $j^a j^b$ OPE, and that the gauge field Lagrangian density is defined to be

$$-\frac{1}{4g_{\rm YM}^2} F^a_{\mu\nu} F^{a\mu\nu} \ . \tag{18.2.6}$$

The parameter $\hat{k}$ differs from the quantized level of the current algebra through the convention for the normalization of the gauge generators, which can be parameterized in terms of the length-squared of a long root, $\psi^2 = 2\hat{k}/k$. The common current algebra convention is $\psi^2 = 2$ so that $\hat{k} = k$. The common particle physics convention is that the inner product for $SO(n)$ groups is the trace in the vector representation, and the inner product for $SU(n)$ groups is twice the trace in the fundamental representation. Both of these give $\psi^2 = 1$ so that $\hat{k} = \frac{1}{2}k$. We should emphasize that it is the quantized level k that matters physically — for example, it determines the allowed gauge representations — but that when we deal with expressions that require a normalization of the generators (like the gauge action) it is generally the parameter $\hat{k}$ that appears.

It is interesting to consider the corresponding relation in open string theory. The ten-dimensional coupling was obtained in eq. (13.3.31),

$$\frac{g_{\rm YM}^2}{\kappa} = 2(2\pi)^{7/2}\alpha' \quad \text{(type I, } d = 10) \ . \tag{18.2.7}$$

Under compactification this becomes

$$\frac{g_{\rm YM}^2}{\kappa} = \frac{2(2\pi)^{7/2}\alpha'}{V^{1/2}} \quad \text{(type I, } d < 10) \ . \tag{18.2.8}$$

Unlike the closed string relation, this depends on the compactification volume.

[3] We feel compelled to be precise about the factors of 2, but most readers will want to skip such digressions as this paragraph.

Gauge quantum numbers

For a gauge group based on a current algebra of level k, only certain representations can be carried by the massless states. The total left-moving weight h of the matter part of any vertex operator is unity. Since the energy-momentum tensor is additive,

$$T_B = T_B^{\rm s} + T_B' \ , \tag{18.2.9}$$

the contribution of the current algebra to h is at most unity. This leaves two possibilities. Either the current algebra state is a primary field with $h \leq 1$, or it is a descendant of the form

$$j^a_{-1} \cdot 1 = j^a \ . \tag{18.2.10}$$

Let us consider the latter case first. The current j^a has $h = 1$, so for bosons the remainder of the matter vertex operator has weight $(0, \frac{1}{2})$. One possibility is ψ^μ, which just gives the gauge boson states. There could also be $(0, \frac{1}{2})$ fields from the internal CFT, but we will see later in the section that this is inconsistent with having any chiral gauge interactions. For fermions the remainder of the matter vertex operator would have weight $(0, \frac{5}{8})$. This combines with the $\beta\gamma$ ghost vertex operator $e^{-\vec{\phi}/2}$ to give a $(0,1)$ current. This is a spacetime spinor, and so is the world-sheet current associated with a spacetime supersymmetry. Thus there are massless fermions of this type only if the theory is supersymmetric, in which case they are the gauginos.

For massless states based on current algebra primaries, the restriction (11.5.43) limits the representations that may appear. For $SU(2)$ at $k = 1$ only the **1** and **2** are allowed, while for $SU(3)$ at $k = 1$ only the **1**, **3**, and $\bar{\mathbf{3}}$ are allowed.

In the Standard Model, there are several notable patterns in the gauge quantum numbers of the quarks and leptons: replication of generations, chirality, quantization of the electric charge, and absence of large ('exotic') representations of $SU(2)$ and $SU(3)$. We have seen in the orbifold and Calabi–Yau examples that multiple generations arise frequently in four-dimensional string theories. This is an attractive feature of higher-dimensional theories in general. The generations arise from massless excitations that differ in the compact dimensions but have the same spacetime quantum numbers. Chirality was discussed in section 18.1, and quantization of electric charge will be discussed in section 18.4. Finally, the absence of exotics, the fact that only the **1** and **2** of $SU(2)$ and the **1**, **3**, and $\bar{\mathbf{3}}$ of $SU(3)$ are found, is 'explained' by string theory if we assume that these gauge symmetries arise from $k = 1$ current algebras. Also, the only scalar in the Standard Model is the $SU(2)$ doublet Higgs scalar, and from tests of this model it is known that no more than

$O(1\%)$ of the $SU(2) \times U(1)$ breaking can come from larger representations.

Unfortunately, this is not a firm prediction of string theory. While the simplest four-dimensional string theories have $k = 1$, there is still an enormous number of tree-level string vacua with higher level current algebras. Also, as discussed in section 16.3, $k = 1$ is impossible if a grand unified group remains below the string scale. For $SU(5)$ only the representations **1**, **5**, $\bar{\mathbf{5}}$, **10**, and $\overline{\mathbf{10}}$ are allowed, for $SO(10)$ only **1**, **16**, $\overline{\mathbf{16}}$, and **10**, and for E_6 only **1**, **27**, and $\overline{\mathbf{27}}$. In each case this includes the representations carried by the quarks, leptons, and the Higgs scalar that breaks the electroweak symmetry, but not the representations needed to break the unified group to $SU(3) \times SU(2) \times U(1)$. The latter are allowed for levels $k \geq 2$. We will return to this point in the next section.

Right-moving gauge symmetries

Thus far we have considered gauge symmetries carried by the left-moving degrees of freedom of the heterotic string. For these the conformal invariance leads to a current algebra. For gauge symmetries carried by the right-movers, the superconformal algebra plus gauge symmetry give rise to a *superconformal current algebra (SCCA)*. The matter part of the gauge boson vertex operator in the -1 picture is

$$\partial X^\mu \tilde{\psi}^a e^{ik\cdot X} \tag{18.2.11}$$

with $\tilde{\psi}^a$ a weight $(0, \frac{1}{2})$ superconformal tensor field. Then

$$\tilde{G}_{-1/2} \cdot \tilde{\psi}^a = \tilde{\jmath}^a \tag{18.2.12}$$

is a $(0,1)$ field. It is nontrivial because

$$\tilde{G}_{1/2} \cdot \tilde{\jmath}^a = 2\tilde{L}_0 \cdot \tilde{\psi}^a = \tilde{\psi}^a . \tag{18.2.13}$$

Also, $\tilde{\jmath}^a$ is a conformal tensor, annihilated by $\tilde{L}_n$ for $n > 0$, though not a superconformal tensor. The $\tilde{\jmath}^a$ thus form a right-moving current algebra.

We take the current algebra to be based on a simple group g at level k, and for simplicity use the current algebra normalization (which is no problem, because we are about to see that these gauge symmetries will never appear in particle physics!). Using the Jacobi identity we can fill in

the rest of the operator products,

$$\tilde{\psi}^a(\bar{z})\tilde{\psi}^b(0) \sim \frac{k\delta^{ab}}{\bar{z}} , \tag{18.2.14a}$$

$$\tilde{\jmath}^a(\bar{z})\tilde{\psi}^b(0) \sim \frac{if^{abc}}{\bar{z}}\tilde{\psi}^c(0) , \tag{18.2.14b}$$

$$\tilde{T}_F(\bar{z})\tilde{\psi}^a(0) \sim \frac{1}{\bar{z}}\tilde{\jmath}^a(0) , \tag{18.2.14c}$$

$$\tilde{\jmath}^a(\bar{z})\tilde{\jmath}^b(0) \sim \frac{k\delta^{ab}}{\bar{z}^2} + \frac{if^{abc}}{\bar{z}}\tilde{\jmath}^c(0) , \tag{18.2.14d}$$

$$\tilde{T}_F(\bar{z})\tilde{\jmath}^a(0) \sim \frac{1}{\bar{z}^2}\tilde{\psi}^a(0) + \frac{1}{\bar{z}}\bar{\partial}\tilde{\psi}^a(0) . \tag{18.2.14e}$$

In particular, the $\tilde{\psi}^a$ are free right-moving fields with a nonstandard normalization.

We can now carry out a generalization of the Sugawara construction. The $\tilde{\jmath}\tilde{\psi}$ product implies that if we define

$$\tilde{\jmath}^a = \tilde{\jmath}^a_\psi + \tilde{\jmath}'^a , \tag{18.2.15}$$

where

$$\tilde{\jmath}^a_\psi = -\frac{i}{2k} f^{abc}\tilde{\psi}^b\tilde{\psi}^c , \tag{18.2.16}$$

then $\tilde{\jmath}'^a$ is nonsingular with respect to the $\tilde{\psi}^a$. It follows that there are actually two current algebras. One is built out of the $\tilde{\psi}^a$ and has current $\tilde{\jmath}^a_\psi$ and level $k_\psi = h(g)$. The other commutes with the $\tilde{\psi}^a$ and has current $\tilde{\jmath}'^a$ and level $k' = k - k_\psi$. We see that $k \geq h(g)$, with equality if and only if $\tilde{\jmath}'^a$ is trivial.

As in the Sugawara construction we can separate $\tilde{T}_F$,

$$\tilde{T}_F = \tilde{T}^{\rm s}_F + \tilde{T}''_F , \tag{18.2.17}$$

where

$$\tilde{T}^{\rm s}_F = -\frac{i}{6k^2} f^{abc}\tilde{\psi}^a\tilde{\psi}^b\tilde{\psi}^c + \frac{1}{k}\tilde{\psi}^a\tilde{\jmath}'^a \tag{18.2.18}$$

and $\tilde{T}''_F$ is nonsingular with respect to $\tilde{\psi}^a$ and $\tilde{\jmath}'^a$. Further,

$$\tilde{T}_B = \tilde{T}^\psi_B + \tilde{T}'_B + \tilde{T}''_B , \tag{18.2.19}$$

with

$$\tilde{T}^\psi_B = -\frac{1}{2k}\tilde{\psi}^a\bar{\partial}\tilde{\psi}^a , \tag{18.2.20a}$$

$$\tilde{T}'_B = \frac{1}{2(k' + h(g))} :\tilde{\jmath}'\tilde{\jmath}': . \tag{18.2.20b}$$

The remainders $\tilde{T}''_F$ and $\tilde{T}''_B$ are nonsingular with respect to both $\tilde{\psi}^a$ and $\tilde{\jmath}'^a$. The CFT thus separates into three pieces, with central charges

$$\tilde{c}^{\psi} = \frac{\dim(g)}{2}\,, \quad \tilde{c}' = \frac{k'\dim(g)}{k'+h(g)}\,, \quad \tilde{c}'' = \tilde{c} - \tilde{c}^{\psi} - \tilde{c}'\,. \tag{18.2.21}$$

The SCFT separates into only two pieces, because $\tilde{\psi}^a$ and $\tilde{\jmath}'^a$ are coupled in the supercurrent. In particular, the central charge for the $\tilde{\psi}\tilde{\jmath}'$ SCFT is

$$\tilde{c}^{\psi} + \tilde{c}' = \frac{(3k'+h(g))\dim(g)}{2(k'+h(g))}\,. \tag{18.2.22}$$

This lies in the range

$$\frac{\dim(g)}{2} \le \tilde{c}^{\psi} + \tilde{c}' \le \frac{3\dim(g)}{2}\,. \tag{18.2.23}$$

The lower bound is reached only when $\tilde{\jmath}'^a$ vanishes, and the upper only for an Abelian algebra.

For an Abelian SCCA, the non-Abelian terms in the OPE (18.2.14) vanish. In particular, $\tilde{\jmath}_{\psi}$ vanishes and $k = k'$, so a nontrivial theory requires that $k' \neq 0$. We can then normalize the currents to set $k = k' = 1$. Writing the current as the derivative of a free boson, $\tilde{\jmath} = i\tilde{\partial}H$, gives

$$\tilde{T}^{\rm s}_F = i\tilde{\psi}\tilde{\partial}H\,, \quad \tilde{T}^{\psi}_B + \tilde{T}'_B = -\frac{1}{2}\tilde{\psi}\tilde{\partial}\tilde{\psi} - \frac{1}{2}\tilde{\partial}H\tilde{\partial}H\,. \tag{18.2.24}$$

If there is a right-moving gauge symmetry below the string scale the gauge boson vertex operator must be periodic, and so the fermionic currents $\tilde{\psi}^a$ must always have the same periodicity as the supercurrent $\tilde{T}_F$. This defines an *untwisted* SCCA.

One can derive strong results restricting the relevance of right-moving gauge symmetries to physics. In the $(1,0)$ heterotic string,

1. If there are any massless fermions, then there are no non-Abelian SCCAs.

2. All massless fermions are neutral under any Abelian SCCA gauge symmetries.

3. If any fermions have chiral gauge couplings, then there are no SCCAs.

The first two results are sufficient to imply that the Standard Model $SU(3) \times SU(2) \times U(1)$ gauge symmetries must come from the left-moving gauge symmetries in heterotic string theory. If, as it appears, the $SU(3) \times SU(2) \times U(1)$ gauge couplings are chiral, then there are no right-moving gauge symmetries at all.

To show these, consider the vertex operator for any massless spin-$\frac{1}{2}$ state, whose matter part is

$$S_\alpha \mathscr{V}_K e^{ik_\mu X^\mu}\,. \tag{18.2.25}$$

Here S_α is a spin field for the four noncompact dimensions, leaving a weight $(1, \frac{3}{8})$ operator $\mathscr{V}_K$ from the internal theory. The Ramond generator $\tilde{G}_0$ is Hermitean, implying that

$$\tilde{G}_0^2 = \tilde{L}_0 - \frac{\tilde{c}}{24} \ge 0 \tag{18.2.26}$$

in any unitary SCFT. The internal theory here has central charge 9, and so the internal part $\mathscr{V}_K$ of any massless spin-$\frac{1}{2}$ state saturates the inequality. Incidentally, this also implies that there can never be fermionic tachyons. Further, if the internal theory decomposes into a sum of SCFTs, $G_0 = \sum_i G_0^i$, then the same argument requires that

$$\tilde{L}_0^i = \frac{\tilde{c}^i}{24} \tag{18.2.27}$$

within *each* SCFT.

Now suppose that one of these SCFTs is a non-Abelian SCCA. In the R sector the $\tilde{\psi}^a$ and $\tilde{\jmath}^a$ are periodic. Then $\tilde{L}_0^{\psi} + \tilde{L}'_0$ is bounded below by the zero-point energy $\frac{1}{16}\dim(g)$ of the $\tilde{\psi}^a$, and

$$\tilde{L}_0^{\psi} + \tilde{L}'_0 - \frac{\tilde{c}^{\psi} + \tilde{c}'}{24} \ge \frac{h(g)\dim(g)}{24(k'+h(g))} > 0\,. \tag{18.2.28}$$

This is strictly positive for all states, so massless fermions are impossible and the first result is established. For an Abelian SCCA, the same form holds with $k' = 1$ and $h(g) = 0$, so equality is possible. However, the term $\frac{1}{2}j_0 j_0$ in $\tilde{L}'_0$ makes an additional positive contribution unless the charge j_0 is zero for the state, establishing the second result.

The equivalence (18.2.24) means that a $U(1)$ SCCA algebra has the same world-sheet action as a flat dimension. Further, as noted above, for an SCCA associated with a gauge interaction the periodicity of the fermionic current ψ is the same as that of the ψ^μ. Then if there is a $U(1)$ SCCA the massless R sector ground states will be the same as those of a *five*-dimensional theory. The $SO(4,1)$ spinor representation **4** decomposes into one four-dimensional representation of each chirality, $\mathbf{2} + \bar{\mathbf{2}}$, so the massless states come in pairs of opposite chirality. In other words, the $SO(4,1)$ spin $\psi_0^\mu \psi_0$ commutes with the GSO projection and (in the massless sector) with the superconformal generators, and so takes massless physical states into massless physical states of the opposite four-dimensional chirality. This establishes the third result, and shows that heterotic string vacua with right-moving gauge symmetries are not relevant to the Standard Model.

Gauge symmetries of type II strings

Now let us consider the possibility of getting the Standard Model from the type II string. Here, both sides are supersymmetric, so the vertex operators of gauge bosons are of one of the two forms

$$\psi^a\tilde{\psi}^\mu e^{ik\cdot X}\ , \quad \psi^\mu\tilde{\psi}^a e^{ik\cdot X}\ , \tag{18.2.29}$$

where ψ^a is associated with a left-moving SCCA and $\tilde{\psi}^a$ with a right-moving SCCA. For example, one could take the internal theory to consist of 18 right-moving and 18 left-moving fermions with trilinear supercurrents (16.1.29). This leads to gauge algebra $g_R \times g_L$ with g_R and g_L each of dimension 18. This can then be broken to the Standard Model by twists. This seems much more economical than the heterotic string, where the dimension of the gauge group can be much larger. However, we will see that the Standard Model does not quite fit into the type II string theory.

The same analysis as used in the heterotic string shows that only one of the two types of gauge boson (18.2.29) may exist. If there are chiral fermions in the R–NS sector there can be no left-moving SCCA, and if there are chiral fermions in the NS–R sector there can be no right-moving SCCA. In order to have both chiral fermions and gauge symmetries, the fermions must all come from one sector, say R–NS, and the gauge symmetries all from right-moving SCCAs.

Now let us see that this does not leave room for the Standard Model. To be precise, it is impossible to have an $SU(3) \times SU(2) \times U(1)$ gauge symmetry with massless $SU(3)$ triplet and $SU(2)$ doublet fermions. The internal part of any massless state has weight $\tilde{h} = \frac{1}{2}$. This restricts the current algebra part to be either a primary state of the SCCA, annihilated by all the $\tilde{\psi}^a_r$ and $\tilde{J}'^a_n$ for $r, n > 0$, or of the form $\tilde{\psi}^a_{-1/2}|1\rangle$. The latter is a gaugino, in the adjoint representation, so the triplets and doublets must be primary states instead. By the same argument as in the conformal case, the allowed representations for the primary states are restricted according to the level k' of the current $\tilde{J}'^a$ of the SCCA, so that $k' \geq 1$ in both the $SU(2)$ and $SU(3)$ factors in order to have doublets and triplets respectively. Noting that the central charge (18.2.22) increases with k', the total central charge of the SCCAs is

$$\tilde{c} \geq \frac{8}{2} + \tilde{c}^{SU(3),1} + \frac{3}{2} + \tilde{c}^{SU(2),1} + \tilde{c}^{U(1)} = 4 + 2 + \frac{3}{2} + 1 + \frac{3}{2} = 10\ . \tag{18.2.30}$$

This exceeds the total $\tilde{c} = 9$ of the internal theory, so there is a contradiction.

This is an elegant argument, using only the world-sheet symmetries. However, progress in string duality has made its limitations clearer. Since all string theories are connected by dualities, we would expect that nonperturbatively a spectrum that can be obtained in one string theory can be obtained in any other. The most obvious limitation of the argument is that it applies only to vacua without D-branes, because the latter would have additional open string states. One might also wonder whether some or all of the Standard Model states can originate not as strings but as D-branes. As long as string perturbation theory is valid then all D-branes and other nonperturbative states should have masses that diverge as $g \to 0$, so that string perturbation theory gives a complete account of the physics at any fixed energy. However, we will see in the next chapter that D-branes can become massless at some points in moduli space, and that this is associated with a breakdown of string perturbation theory.

18.3 Mass scales

There are a number of important mass scales in string theory:

1. The *gravitational* scale $m_{\rm grav} = \kappa^{-1} = 2.4 \times 10^{18}$ GeV, at which quantum gravitational effects become important; this is somewhat more useful than the Planck mass, which is a factor of $(8\pi)^{1/2}$ greater.
2. The *electroweak* scale $m_{\rm ew}$, the scale of $SU(2) \times U(1)$ breaking, $O(10^2)$ GeV.
3. The *string* scale $m_{\rm s} = \alpha'^{-1/2}$, the mass scale of excited string states.
4. The *compactification* scale $m_{\rm c} = R_{\rm c}^{-1}$, the characteristic mass of states with momentum in the compact directions.
5. The *grand unification* scale $m_{\rm GUT}$, at which the $SU(3) \times SU(2) \times U(1)$ interactions are united in a simple group.
6. The *superpartner* scale $m_{\rm sp}$, the mass scale of the superpartners of the Standard Model particles.

In this section we consider relations among these scales. Of course, there may be additional scales. The unification of the gauge group may take place in several steps, and there may be other intermediate scales at which new degrees of freedom appear. Also, these scales may not all be relevant. For example, when the internal CFT is a sigma model on a manifold large compared to the string scale, the idea of compactification applies. There are states with masses-squared of order $m_{\rm c}^2 \ll m_{\rm s}^2$, states which would be massless in the noncompact theory and which have internal momenta of order $m_{\rm c}$. However, as $m_{\rm c}$ increases to $m_{\rm s}$ these states become indistinguishable from the various 'stringy' states, and compactification

is not so meaningful. The internal CFT may have several equivalent descriptions as a quantum field theory, with 'internal excitations' and 'stringy states' interchanging roles. Similar remarks apply to the grand unification and supersymmetry scales.

For most of the discussion we will assume explicitly that the string theory is weakly coupled, and that the Standard Model gauge couplings remain perturbative up to the string scale. In this case it is possible to make some fairly strong statements. As we know from chapter 14, strong coupling opens up many new dynamical possibilities. The consequences for physics in four dimensions have not been fully explored; we will make a few comments at the end of the section.

The relation between the string and gravitational scales follows from the relation (18.2.5) between the couplings,

$$\frac{m_s}{m_{grav}} = g_{YM}(k/2)^{1/2} . \tag{18.3.1}$$

The quantities on the right are not too far from unity, so the string and gravitational scales are comparable. In the minimal supersymmetric model to be discussed below, the coupling g_{YM} at high energy is of order 0.7; for $k = 1$ this gives $m_s \approx 1.2 \times 10^{18}$ GeV . This result is shown graphically in figure 18.1: plotted as a function of energy E are the four-dimensional gauge coupling $\alpha_{YM} = g_{YM}^2/4\pi$ and the corresponding dimensionless gravitational coupling $\kappa^2 E^2$. The scale where these meet is the expected scale of unification of the gravitational and gauge interactions, the string scale.

Now consider the compactification scale. Suppose that there are k dimensions compactified at some scale $m_c \ll m_s$. Between the scales m_c and m_s, physics is described by a $(4 + k)$-dimensional field theory, in which a gauge coupling α_{4+k} has dimension m^{-k} and the gravitational coupling G_{4+k} has dimension m^{-k-2}. The behaviors of the dimensionless couplings $\alpha_{4+k}E^k$ and $G_{4+k}E^{k+2}$ are indicated in figure 18.1 by dashed lines. The gauge coupling rises rapidly from its four-dimensional value α_{YM}. Our assumption that the coupling remains weak up to the string scale then implies that the latter is not far above the compactification scale (in this section 'scale' always refers to energy, rather than the reciprocal length). Also, it presumably does not make sense for the compactification scale to be greater than the string scale, as illustrated by T-duality for toroidal compactification. Thus the string, gravitational, and compactification scales are reasonably close to one another. In open string theory, the quantitative relation (18.2.8) between the scales is different, but the reader can show that with the weak-coupling assumption these three scales are again close to one another.

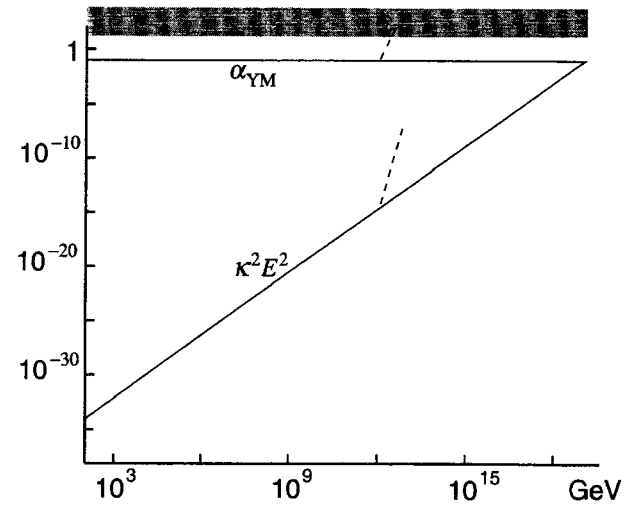


Fig. 18.1. The dimensionless gauge and gravitational couplings as a function of energy. On the scale of this graph we neglect the differences between gauge couplings and the running of these couplings. The dashed curves illustrate the effect of a compactification scale below the Planck scale, at 10^{12} GeV in this example (the slopes correspond to all six compact dimensions being at this same scale, and are reduced if there are fewer). The shaded region indicates the breakdown of perturbation theory.

Next consider the unification scale. First let us review $SU(5)$ unification of the Standard Model. The Standard Model gauge group $SU(3)\times SU(2)\times U(1)$ can be embedded in the **5** representation of $SU(5)$, with $SU(3)$ being the upper 3×3 block, $SU(2)$ the lower 2×2 block, and $U(1)$ hypercharge the diagonal element

$$\frac{Y}{2} = \mathrm{diag}\left(-\frac{1}{3}, -\frac{1}{3}, -\frac{1}{3}, \frac{1}{2}, \frac{1}{2}\right) . \tag{18.3.2}$$

The $SU(n)$ generators for the fundamental representation **n** are conventionally normalized $\mathrm{Tr}(t^a t^b) = \frac{1}{2}\delta^{ab}$. This is also true for $U(1)$ if we define $t^{U(1)} = (\frac{3}{5})^{1/2}\frac{1}{2}Y$, in which case $SU(5)$ symmetry implies

$$g_3 = g_2 = g_1 = g_{SU(5)} \tag{18.3.3}$$

for the $SU(3) \times SU(2) \times U(1)$ couplings. The hypercharge coupling g' is defined by

$$\tfrac{1}{2}g'Y = g_{U(1)}t^{U(1)} \Rightarrow g' = (3/5)^{1/2}g_1 . \tag{18.3.4}$$

The $SU(5)$ prediction is then

$$(5/3)^{1/2}g' = g_2 = g_3 \,. \tag{18.3.5}$$

The weak mixing angle θ_w is defined by $\sin^2\theta_w = g'^2/(g_2^2 + g'^2)$. Before taking into account radiative corrections, the $SU(5)$ prediction is $\sin^2\theta_w = \frac{3}{8}$. The same holds for standard $SO(10)$ and E_6 unification, because $SU(5)$ is just embedded in these.

For the purposes of the present section we will assume that the same relation (18.3.5) holds in string theory; in the next we will discuss the circumstances under which this is true. In both string theory and grand unified field theory, this tree-level relation receives substantial renormalization group corrections below the scale of $SU(5)$ breaking. To one-loop order, the couplings depend on energy as

$$\mu\frac{\partial}{\partial\mu}g_i = \frac{b_i}{16\pi^2}g_i^3 \,. \tag{18.3.6}$$

This integrates to

$$\alpha_i^{-1}(\mu) = \alpha_i^{-1}(m_{GUT}) + \frac{b_i}{4\pi}\ln(m_{GUT}^2/\mu^2) \,, \tag{18.3.7}$$

where $\alpha_i = g_i^2/4\pi$. For a non-Abelian group the constant b_i is

$$b_i = -\frac{11}{3}T_g + \frac{1}{3}\sum_{\substack{\text{complex}\\\text{scalars}}}T_r + \frac{2}{3}\sum_{\substack{\text{Weyl}\\\text{fermions}}}T_r \,, \tag{18.3.8}$$

where $\mathrm{Tr}(t_r^a t_r^b) = T_r\delta^{ab}$ and $T_g = T_{r=\text{adjoint}}$. For a $U(1)$ group the result is the same with $T_g = 0$ and T_r replaced by q^2.

The couplings at the weak interaction scale M_Z are $\alpha_1^{-1} \approx 59$, $\alpha_2^{-1} \approx 30$, and $\alpha_3^{-1} \approx 9$. Extrapolating the couplings $\alpha_i(\mu)$ as in eq. (18.3.7), $SU(5)$ unification makes the prediction (18.3.3) that at some scale m_{GUT} they become equal. This is often expressed as a prediction for $\sin^2\theta_w(m_Z)$: use $\alpha_1^{-1}(m_Z)$ and $\alpha_3^{-1}(m_Z)$ to solve for m_{GUT} and α_{GUT}, and then extrapolate downwards to obtain a prediction for $\alpha_2^{-1}(m_Z)$. The prediction depends on the spectrum of the theory through the beta function (18.3.8).[4] For the minimal $SU(5)$ unification of the Standard Model,

$$\sin^2\theta_w(m_Z) = 0.212 \pm 0.003 \,. \tag{18.3.9}$$

For the minimal supersymmetric Standard Model, which consists of the Standard Model plus a second Higgs doublet plus the supersymmetric partners of these,

$$\sin^2\theta_w(m_Z) = 0.234 \pm 0.003 \,. \tag{18.3.10}$$

The experimental value is

$$\sin^2\theta_w(m_Z) = 0.2313 \pm 0.0003 \,. \tag{18.3.11}$$

The minimal nonsupersymmetric model is clearly ruled out. On the other hand, the agreement between the minimal supersymmetric $SU(5)$ prediction and the actual value is striking, considering that *a priori* $\sin^2\theta_w(m_Z)$ could have been anywhere between 0 and 1. The agreement between the supersymmetric prediction and the actual value means that the three gauge couplings meet, with

$$m_{GUT} = 10^{16.1\pm0.3}\ \text{GeV} \,, \quad \alpha_{GUT}^{-1} \approx 25 \,. \tag{18.3.12}$$

In the nonsupersymmetric case, the disagreement with $\sin^2\theta_w(m_Z)$ implies that the three couplings do not meet at a single energy, but meet pairwise at three energies ranging from 10^{13} GeV to 10^{17} GeV.

To a first approximation, the unification scale (18.3.12) is fairly close to the string scale and so to the compactification and gravitational scales. This is also necessary for the stability of the proton. The running of the couplings is shown pictorially in figure 18.2. We should note that a direct comparison of the string and unification scales is not appropriate at the level of accuracy of the extrapolation (18.3.12). Rather, we should compare the measured couplings to a full one-loop string calculation: this is just the calculation (16.4.32). Ignoring for now the threshold correction, this relation is of the form (18.3.7) with the string unification scale (16.4.36)

$$m_{SU} = k^{1/2}g_{YM} \times 5.27 \times 10^{17}\,\text{GeV} \to 3.8 \times 10^{17}\,\text{GeV} \,. \tag{18.3.13}$$

We have inserted the relation (18.3.1) between the gauge and gravitational scales and then carried out the numerical evaluation using the unified coupling (18.3.12) and assuming $k = 1$. The resulting discrepancy between the string unification scale and the value in minimal SUSY unification is a factor of 30. This is larger than the experimental uncertainty, but small compared to the fifteen orders of magnitude difference between the electroweak scale and the string scale. This suggests that the unification and string scales are actually one and the same, so that not just the three gauge couplings but also the gravitational coupling meet at a single point; the apparent difference between the unification and string scales would then be due to some small additional correction.

Before discussing what such a correction might be, let us consider the consequences if the two scales actually are separated. This means that there is a range $m_{GUT} < E < m_s$ in which physics is described by a grand unified *field* theory, with $SU(3) \times SU(2) \times U(1)$ contained in $SU(5)$ or another

[4] The experiment and theory are sufficiently precise that one must take into account the two-loop beta function, threshold effects at the weak and unified scales, and other radiative corrections to the weak interaction.

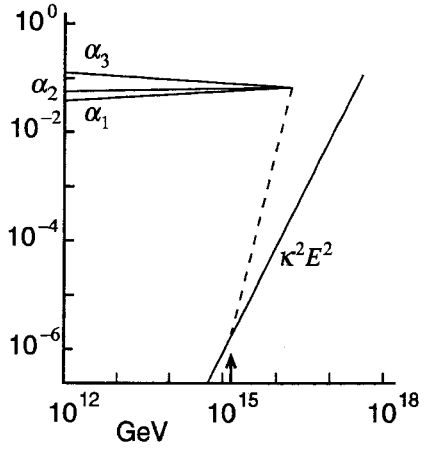


Fig. 18.2. The unification of the gauge couplings in the minimal supersymmetric unified model, and the near-miss of the gravitational coupling. The dashed line shows the potential effect of an extra dimension of the form $S_1/\mathbf{Z}_2$ at the scale indicated by the arrow.

simple group. This theory is presumably four-dimensional, because even a factor of 30 difference between the string and compactification scales is difficult to accommodate. The unified group must then be broken to $SU(3)\times SU(2)\times U(1)$ by the usual Higgs mechanism. As we have discussed in the previous section, this is not possible if the underlying current algebra is level one, because a Higgs scalar in the necessary representation cannot be lighter than the string scale. There do exist higher level string models in which such a separation of scales is possible.

An intermediate possibility is partial unification, embedding $SU(3) \times SU(2) \times U(1)$ in one of

$$SU(5)' \times U(1) \subset SO(10)\,, \qquad (18.3.14a)$$

$$SU(4) \times SU(2)_L \times SU(2)_R \subset SO(10)\,, \qquad (18.3.14b)$$

$$SU(3)_C \times SU(3)_L \times SU(3)_R \subset E_6\,. \qquad (18.3.14c)$$

The group $SU(5)'\times U(1)$ is known as *flipped* $SU(5)$. Color $SU(3)$ and weak $SU(2)$ are embedded in $SU(5)$ in the usual way, but hypercharge is a linear combination of a generator from $SU(5)$ and the $U(1)$ generator. String models based on flipped $SU(5)$ have been studied in some detail. The group $SU(4)\times SU(2)_L\times SU(2)_R$ is known as *Pati–Salam unification.* Color $SU(3)$ is in the $SU(4)$ factor, weak $SU(2)$ is $SU(2)_L$, and hypercharge is a linear

combination of a generator from $SU(4)$ and a generator from $SU(2)_R$. In the $SU(3)^3$ group, sometimes called *trinification,* color is $SU(3)_C$, weak $SU(2)$ is in $SU(3)_L$, and hypercharge is a linear combination of generators from $SU(3)_L$ and $SU(3)_R$. When G is one of these partially unified groups and is embedded in a simple group as indicated in eq. (18.3.14), then the Standard Model group within G has the same embedding as in simple unification. The tree-level prediction for $\sin^2\theta_w(m_Z)$ is therefore again $\frac{3}{8}$, but the running of the couplings will of course be different between m_{GUT} and m_s. These partially unified groups can all be broken to the Standard Model by Higgs fields that are allowed at level one.

Now let us consider the corrections that might eliminate the difference between m_{GUT} and m_{SU}. The quoted uncertainties in the grand unified predictions come primarily from the uncertainty in the measured value of α_3, and in the supersymmetric case from the unknown masses of the superpartners. There is a far greater uncertainty implicit in the assumption that the spectrum below the unification scale is minimal. Adding a few extra light fields, either at the electroweak scale or at an intermediate scale, can change the running by an amount sufficient to bring the unification scale up to the string scale.

There is also a threshold correction due to loops of string-mass fields. This is a function of the moduli, as in the orbifold example (16.4.38),

$$\Delta_a = c_a - \sum_i \frac{b_a^i|G^i|}{|G|} \ln\left[(T_i+T_i^*)|\eta(T_i)|^4(U_i+U_i^*)|\eta(U_i)|^4\right]\,. \qquad (18.3.15)$$

Although this correction reflects a sum over the infinite set of string states, its numerical value is rather small for values of the moduli of order 1. It can become large if the moduli become large. For example,

$$\Delta_a \approx \sum_i \frac{b_a^i|G^i|}{|G|}\frac{\pi(T_i+T_i^*)}{6} \qquad (18.3.16)$$

for large T_i, from the asymptotics of the eta function. For large enough T_i, in those models where the correction has the correct sign, this can account for the apparent difference between the string and unification scales.

Finally, in more complicated string models the tree-level predictions may be different and so also the predicted unification scale. We will discuss this somewhat in the next section.

All of these modifications have the drawback that a change large enough to raise the unification scale to the string scale will generically change the prediction for $\sin^2\theta_w$ by an amount greater than the experimental and theoretical uncertainty, so that the excellent agreement is partly accidental. Since the gauge couplings already meet, it would be simple and economical to leave them unchanged and instead change the energy dependence of

the gravitational coupling so that it meets the other three. However, this seems impossible, since the 'running' of the gravitational coupling $\kappa^2 E^2$ is just dimensional analysis: the gravitational interaction is essentially classical below the string scale and quantum effects do not affect its energy dependence.

This is one point where the new dynamical ideas arising from strongly coupled string theory can make a difference. One way to change the dimensional analysis is to change the dimension! It does not help to have a low compactification scale of the ordinary sort: as shown in figure 18.1, all the couplings increase more rapidly but they do not meet any sooner. Consider, however, the *strongly coupled* $E_8 \times E_8$ heterotic string compactified on a Calabi–Yau space K. From the discussion in chapter 14, this is the eleven-dimensional M-theory compactified on a product space

$$K \times \frac{S_1}{\mathbf{Z}_2} . \tag{18.3.17}$$

The scales of the two factors are independent; let us suppose that the space $S_1/\mathbf{Z}_2$ is larger, so that its mass scale R_{10}^{-1} lies below the unification scale. The point is that the gauge and matter fields live on the boundary of this space, which remains four-dimensional, while the gravitational field lives in the five-dimensional bulk. The effect is as shown in figure 18.2: the gauge couplings evolve as in four dimensions, while the gravitational coupling has a kink. For an appropriate value of R_{10}, all four couplings meet at a point.

With the only data points being the low energy values of the gauge couplings, there is no way to distinguish between these various alternatives. If in fact supersymmetry is found at particle accelerators, then measurement of the superpartner masses will allow similar renormalization group extrapolations and may enable us to unravel the 'fine structure' at the string scale.

This brings us to the next scale, which is $m_{\rm sp}$. The lower limits on the various charged and strongly interacting superpartners are of order 10^2 GeV. If supersymmetry is the solution to the hierarchy problem, the cancellation of the quantum corrections to the Higgs mass requires that the splitting between the Standard Model particles and their superpartners be not much larger than this,

$$10^2 \text{ GeV} \lesssim m_{\rm sp} \lesssim 10^3 \text{ GeV} . \tag{18.3.18}$$

Of all the new phenomena associated with string theory, supersymmetry is the one that is likely to be directly accessible to particle accelerators.

Finally, we should ask why the supersymmetry and electroweak scales lie so far below the others; we will discuss this briefly in section 18.8.

18.4 More on unification

In this section we collect a number of additional results on the relation between string theory and grand unification.

The first issue is the condition under which the grand unified relation $g_1 = g_2 = g_3$ holds in string theory at tree level. This is obviously the case in theories where a unified group remains unbroken below the string scale. It is also true if, as in the orbifold and Calabi–Yau cases, a unified group is broken at the string or compactification scale by twists. Although there is no scale at which the world looks like a four-dimensional grand unified theory, the inheritance principle guarantees that the equality of the tree-level couplings persists after the twist.

More generally one can make some statements just from current algebra arguments. The current algebra relation (18.2.5) between the gravitational coupling and any single gauge coupling implies that for the $SU(2)$ and $SU(3)$ gauge couplings

$$\frac{\alpha_2}{\alpha_3} = \frac{\hat{k}_3}{\hat{k}_2} = \frac{k_3}{k_2} . \tag{18.4.1}$$

Thus the grand unified prediction $\alpha_2 = \alpha_3$ holds whenever the levels of the $SU(3)$ and $SU(2)$ current algebras are equal. In any case one expects that the levels are small integers, models with large levels having complicated spectra, so that if the levels are not equal their ratio differs substantially from unity. Since the unification scale can be determined from any pair of couplings, this implies a large change in the unification scale, spoiling the near-equality between the unification and string scales. Thus it is likely that, whatever the levels of the $SU(2)$ and $SU(3)$ current algebras, they are equal.

For the $U(1)$ coupling there is no similar statement, because there is no level to give an absolute normalization to the current. One general result concerns the common situation that there is a continuous moduli space of vacua, all with an unbroken $U(1)$ symmetry: if there are chiral fermions, then at tree level the coupling g_1, and so also $\sin^2\theta_{\rm w}$, is the same for all the connected vacua. To see this, write the $U(1)$ current algebra in terms of a left-moving boson $H(z)$. Let us consider how H might appear in the vertex operator for the modulus that interpolates between the vacua. The $U(1)$ is assumed to be unbroken for all vacua, so the vertex operator must be invariant under $H \to H + \epsilon$ — it can only contain derivatives of H. Dimensionally, the only operator that can then appear in a massless vertex operator is ∂H, and the whole matter vertex operator must be

$$\partial H \tilde{\psi} e^{ik\cdot X} \tag{18.4.2}$$

for some $(0, \frac{1}{2})$ superconformal tensor $\tilde{\psi}$. However, we know from sec-

tion 18.2 that such tensors are inconsistent with chirality, so H cannot appear in the vertex operator at all. Expectation values of the $U(1)$ current are then independent of the modulus, and therefore so is the gauge coupling.

A related issue is the quantization of electric charge. An isolated fractional multiple of the electron charge has never been seen in nature. The Standard Model has fractionally charged quarks, of course, but these are confined in hadrons of integer charge. It is therefore useful to work with

$$Q' = Q_{\rm EM} + \frac{T}{3}\,, \tag{18.4.3}$$

where the triality T, defined mod 3, is $+1$ for an $SU(3)$ **3** and -1 for a $\bar{\mathbf{3}}$. One can take T to be the $SU(3)$ generator which is diag$(1,1,-2)$ in the **3** representation. Quarks are confined in states with $T = 0$ mod 3, so for all isolated states $Q' = Q_{\rm EM}$ mod 1. The charge Q' has been defined so as to be an integer for all Standard Model fields, so it follows that $Q_{\rm EM}$ is an integer for all isolated states.

Now consider this issue in string theory, starting with some special cases. If there is an $SU(5)$ gauge group below the string scale, there can be no isolated fractional charges. In the $SU(5)$ **5**, the charge

$$Q' = Q_{\rm EM} + \frac{1}{3}T = \frac{1}{2}Y + I_3 + \frac{1}{3}T \tag{18.4.4}$$

is

$$\begin{aligned}&\mathrm{diag}\left(-\frac{1}{3},-\frac{1}{3},-\frac{1}{3},\frac{1}{2},\frac{1}{2}\right) + \mathrm{diag}\left(0,0,0,\frac{1}{2},-\frac{1}{2}\right) + \mathrm{diag}\left(\frac{1}{3},\frac{1}{3},-\frac{2}{3},0,0\right)\\ &\qquad = \mathrm{diag}(0,0,-1,1,0)\,.\end{aligned} \tag{18.4.5}$$

Since Q' is an integer for all states in the **5** and all representations can be obtained as tensor products of **5**s, Q' is an integer for all states and so $Q_{\rm EM}$ is an integer for all isolated states.

Now consider the case in which there is a level one $SU(5)$ current algebra at the string scale, broken by twists to $SU(3)\times SU(2)\times U(1)$. Let us represent this current algebra by free fermions $\lambda^{K\pm}$ for $K = 4,\ldots,8$, with $SU(3)$ acting on $K = 4,5,6$ and $SU(2)$ acting on $K = 7,8$ (the numbering is kept consistent with the orbifold and Calabi–Yau chapters). The current corresponding to Q' is thus

$$j' = \lambda^{6-}\lambda^{6+} - \lambda^{7-}\lambda^{7+} = i\partial(H_7 - H_6)\,. \tag{18.4.6}$$

In a sector with boundary conditions

$$\lambda^{K+}(\sigma_1 + 2\pi) = \exp(2\pi i v_K)\lambda^{K+}(\sigma_1)\,, \tag{18.4.7}$$

the bosonized vertex operator

$$\exp\left[i\sum_K (1/2 - v_K)H_K\right] \tag{18.4.8}$$

has charge

$$Q' = v_6 - v_7\,. \tag{18.4.9}$$

Thus there will be isolated fractional charges if there are twisted sectors with $v_6 \neq v_7$. In fact there *must* be such sectors. Consider the gauge boson associated with the current $\lambda^{6+}\lambda^{7-}$. This carries the $SU(3)\times SU(2)$ representation $(\mathbf{3},\mathbf{2})$ and is one of the $SU(5)$ bosons that is removed by the twists that break the $SU(5)$ symmetry. One of the twists must therefore have $\exp[2\pi i(v_6 - v_7)] \neq 1$, and the corresponding twisted sector has fractional Q'.

The lightest fractionally charged particle must be stable due to charge conservation. The number of fractional charges in ordinary matter is known to be less than 10^{-20} per nucleon. If fractionally charged particles of mass m were in thermal equilibrium in the early universe at temperatures $T > m$, it is estimated that annihilation would only reduce their present abundance to approximately 10^{-9} per nucleon. Whether this is a problem depends critically on the masses of the fractionally charged states, whether all are near the string scale or whether some are near the weak scale. If all the fractional charges are superheavy then the situation is very similar to that with magnetic monopoles in grand unified theories. Diluting the density of relic monopoles was one of the original motivations for inflationary cosmology; this would also sufficiently dilute the fractional charges. It may also be the case that the universe was never hot enough to produce string-scale states thermally. Fractionally charged particles with masses near the weak scale are a potentially severe problem, unless they are charged under a new strongly coupled gauge symmetry and so confined.

In Calabi–Yau compactification the fractionally charged states are superheavy. The twist that breaks $SU(5)$ is accompanied by a freely-acting spacetime symmetry, so that any string in the twisted sector of the gauge group will be stretched in spacetime. In orbifold compactifications there can be massless fractionally charged states from the twisted sectors, but the Calabi–Yau result suggests that superheavy masses are more generic.

Let us mention a generalization of the previous result. If the $SU(3)$ and $SU(2)$ gauge symmetries are at level one, and the tree-level value of $\sin^2\theta_{\rm w}$ is the $SU(5)$ value $\frac{3}{8}$, and $SU(5)$ is broken to $SU(3)\times SU(2)\times U(1)$, then there are states of fractional Q'. To see this, write the $SU(3)\times SU(2)\times U(1)$

current algebra in terms of free bosons, the diagonal currents being[5]

$$j^3_{SU(3)} = \frac{i}{2}\partial(H^4 - H^5)\,, \tag{18.4.10a}$$

$$j^8_{SU(3)} = \frac{i}{2\times 3^{1/2}}\partial(H^4 + H^5 - 2H^6)\,, \tag{18.4.10b}$$

$$j^3_{SU(2)} = \frac{i}{2}\partial(H^7 - H^8)\,, \tag{18.4.10c}$$

$$j_{Y/2} = \frac{i}{6}\partial[-2(H^4 + H^5 + H^6) + 3(H^7 + H^8)]\,. \tag{18.4.10d}$$

The current $j_{Y/2}$ is normalized so that the z^{-2} term in the $j_{Y/2}j_{Y/2}$ operator product is $\frac{5}{3}$ times that of the non-Abelian currents, giving the tree-level value $\sin^2\theta_w = \frac{3}{8}$. Then

$$j' = j_{Y/2} + j^3_{SU(2)} + \frac{2}{3^{1/2}}j^8_{SU(3)} = i\partial(H^7 - H^6) \tag{18.4.11}$$

just as above, and $Q' = k^7 - k^6$. If Q' were an integer for all states, then the $(1,0)$ operator

$$\exp[i(H^6 - H^7)] \tag{18.4.12}$$

would have single-valued OPEs with respect to all vertex operators. However, this would mean that the current algebra is larger than the assumed $SU(3)\times SU(2)\times U(1)$; in fact, closure of the OPE gives a full $SU(5)$ algebra and gauge group. So under the assumptions given there must be fractional charges. This is more general than the earlier result, the assumption of a twisted $SU(5)$ current algebra having been replaced by a weaker assumption about the weak mixing angle.

There are various further generalizations. By an extension of the above argument it can be shown that if the current algebras are level one, and there are no states of fractional Q', and $SU(5)$ is broken, then the tree-level $\sin^2\theta_w$ must take one of the values $\frac{3}{20}, \frac{3}{32}, \frac{3}{44}, \ldots$. To make these values consistent with experiment takes a very nonstandard running of the couplings, suggesting that either the current algebras are higher level or that supermassive fractional charges should be expected to exist. One can also obtain constraints on higher level models, but they are less restrictive. We mention in passing that at higher levels we cannot use the same free-boson representation of the current algebras. Rather, simple currents, defined below eq. (15.3.19), play the role that exponentials of free fields play in the level one case.

[5] Only four free bosons are needed to represent the current algebra — the linear combination $H^4+H^5+H^6+H^7+H^8$ does not appear. The notation is chosen to correspond to the bosonization of the earlier free Fermi representation.

If unconfined fractional charges do exist, electric charge is quantized in a unit e/n smaller than the electron charge. The Dirac quantization condition implies that any magnetic monopole must have a magnetic charge which is an integer multiple of $2\pi n/e$. Various classical monopole solutions exist in string theories, and one expects that the minimum value allowed by the Dirac quantization is attained. Discovery of a monopole with charge $2\pi/e$ would imply the nonexistence of fractional charges, and so have implications for string theory through the above theorems.

The final issues are proton decay and neutrino masses. The details here are rather model-dependent, but we will outline some of the general issues. Two of the successes of the Standard Model are that it explains the stability of the proton and the lightness of the neutrinos. The most general renormalizable action with the fields and gauge symmetries of the Standard Model has no terms that violate baryon number B. This is termed an *accidental symmetry,* meaning that the long life of the proton is indirectly implied by the gauge symmetries. The allowed $\Delta B \neq 0$ terms of lowest dimension are some four-fermion interactions. These will be induced in grand unified theories by exchange of heavy gauge (X) bosons. The operators have dimension 6, so the amplitude goes as M_X^{-2}, and an estimate of the resulting proton lifetime is

$$\tau_P \approx \left(\frac{M_X}{10^{15}\text{ GeV}}\right)^4 \times 10^{31\pm1}\text{ years.} \tag{18.4.13}$$

The experimental bound is of order 10^{32} years, so this is an interesting rate although very sensitive to the unification scale. Similarly, a mass for the Weyl neutrinos would violate lepton number, and L is another accidental symmetry of the Standard Model.

In supersymmetric theories there are gauge-invariant dimension 3, 4, and 5 operators that violate B and/or L. These are the superpotential terms

$$\begin{aligned}&\mu_1 H_1 L\\ &+ \eta_1 U^c D^c D^c \;+\; \eta_2 QLD^c \;+\; \eta_3 LLE^c\\ &+ \frac{\lambda_1}{M}QQQL \;+\; \frac{\lambda_2}{M}U^cU^cD^cE^c \;+\; \frac{\lambda_3}{M}LLH_2H_2\,.\end{aligned} \tag{18.4.14}$$

Here Q, U^c, D^c, L, and E^c are chiral superfields, containing respectively the left-handed quark doublet, anti-up quark, anti-down quark, lepton doublet, and the positron; H_1 and H_2 are chiral superfields containing the two Higgs scalars needed in the supersymmetric Standard Model. Gauge and generation indices are omitted. The dimension 3 term in the first line would generate a neutrino mass and so it must be that $\mu_1 \leq 10^{-3}$ GeV, which is small compared to the weak scale and minuscule compared to the unification scale. The terms in the second line are of

dimension 4, unsuppressed by heavy mass scales, and their dimensionless coefficients must be very small. For example, the first two terms together can induce proton decay, so $\eta_1\eta_2 \leq 10^{-24}$. The terms in the third line are of dimension 5, suppressed by one power of mass; the proton decay limit $\lambda_{1,2}/M \leq 10^{-25}$ GeV^{-1} requires a combination of heavy scales and small coefficients, while the lightness of the neutrino implies that $\lambda_3/M \leq 10^{-13}$ GeV^{-1}. Thus any supersymmetric theory needs discrete symmetries to eliminate almost completely the dimension 3 and 4 terms and at least to suppress the dimension 5 terms unless they are not proportional to small Yukawa couplings. Several groups have argued that the necessary symmetries exist in various classes of string vacua. In many examples these seem to be associated with an additional $U(1)$ gauge interaction broken in the TeV energy range.

There is at least one respect in which string theories, or at least higher-dimensional theories, may have an advantage over other supersymmetric unified theories. The $SU(2)$ doublet Higgs scalar that breaks the weak interaction must have a mass of order the electroweak scale, while its color triplet GUT partners can mediate proton decay and so must have masses near the unification scale. It is possible to arrange the necessary mass matrix for these states without fine tuning, but the models in general seem rather contrived. String theory provides another solution. When an $SU(5)$ current algebra symmetry is broken by twists, the low energy states do not in general fit into complete multiplets of the unified symmetry: some of the states are simply projected away. This is true somewhat more generally for any higher-dimensional gauge theory compactified to $d = 4$ with the gauge symmetry broken at the compactification scale by Wilson lines. In these cases one keeps certain attractive features, such as the unification of the gauge interactions and the prediction of mixing angle, but the undesired Higgs triplet need not be present.

18.5 Conditions for spacetime supersymmetry

Consider any four-dimensional string theory with $N = 1$ spacetime supersymmetry. We will show that there must be a right-moving $N = 2$ world-sheet superconformal symmetry, generalizing the results found in the orbifold and Calabi–Yau examples.

The current for spacetime supersymmetry is

$$\tilde{\jmath}_\alpha = e^{-\tilde{\phi}/2}\tilde{S}_\alpha\tilde{\Sigma}\,, \quad \tilde{\jmath}_{\dot\alpha} = e^{-\tilde{\phi}/2}\tilde{S}_{\dot\alpha}\tilde{\bar{\Sigma}}\,. \tag{18.5.1}$$

We have separated the four-dimensional spin field into its $\mathbf{2}$ and $\bar{\mathbf{2}}$ components, denoted respectively by undotted and dotted indices. The four-dimensional spin fields have opposite values of $\exp(\pi i\tilde{F})$, so the internal parts $\tilde{\Sigma}$ and $\tilde{\bar{\Sigma}}$ must also have opposite values by the GSO projection. These are the vertex operators for the ground states of the compact CFT. They must each be of weight $(0, \frac{3}{8})$ in order that the total currents have weight $(0, 1)$. As shown in section 18.2, this is the minimum weight for a field in this sector, and so $\tilde{G}_0$ annihilates both $\tilde{\Sigma}$ and $\tilde{\bar{\Sigma}}$.

The single-valuedness of the OPEs of $\tilde{\jmath}_\alpha$ and $\tilde{\jmath}_{\dot\alpha}$ implies that

$$\tilde{\Sigma}(\bar{z})\tilde{\bar{\Sigma}}(0) = \bar{z}^{-3/4}\cdot\text{single-valued}\,, \tag{18.5.2a}$$

$$\tilde{\Sigma}(\bar{z})\tilde{\Sigma}(0) = \bar{z}^{3/4}\cdot\text{single-valued}\,, \tag{18.5.2b}$$

in order to cancel the branch cuts from the other factors. By unitarity, the coefficient of the unit operator in the OPE

$$\tilde{\Sigma}(\bar{z})\tilde{\bar{\Sigma}}(0) = \bar{z}^{-3/4}\left(1 + \frac{\bar{z}}{2}\tilde{\jmath} + \ldots\right) \tag{18.5.3}$$

cannot vanish, and so can be normalized to 1 as shown. The point of the following argument will be to show that the second term is also nonvanishing, so that there is an additional conserved current $\tilde{\jmath}$.

The OPE of supersymmetry currents is

$$\tilde{\jmath}_\alpha(\bar{z})\tilde{\jmath}_{\dot\beta}(0) \sim \frac{1}{2^{1/2}\bar{z}}(C\Gamma^\mu)_{\alpha\dot\beta}e^{-\tilde{\phi}}\tilde{\psi}_\mu(0)\,. \tag{18.5.4}$$

As required by the supersymmetry algebra, the residue on the right-hand side is the spacetime momentum current; this is in the -1 picture $e^{-\tilde{\phi}}\tilde{\psi}^\mu$ just as in the ten-dimensional equation (12.4.18). It also follows from the supersymmetry algebra that the OPE $\tilde{\jmath}_\alpha\tilde{\jmath}_\beta$ of two undotted currents is nonsingular, implying that

$$\tilde{\Sigma}(\bar{z})\tilde{\Sigma}(0) = O(\bar{z}^{3/4})\,. \tag{18.5.5}$$

The four-point function is then

$$\left\langle \tilde{\Sigma}(\bar{z}_1)\tilde{\bar{\Sigma}}(\bar{z}_2)\tilde{\Sigma}(\bar{z}_3)\tilde{\bar{\Sigma}}(\bar{z}_4)\right\rangle = \left(\frac{\bar{z}_{13}\bar{z}_{24}}{\bar{z}_{12}\bar{z}_{14}\bar{z}_{23}\bar{z}_{34}}\right)^{3/4} f(\bar{z}_1,\bar{z}_2,\bar{z}_3,\bar{z}_4)\,, \tag{18.5.6}$$

where the OPEs as various points become coincident imply that f is a holomorphic function of its arguments. The $\bar{z}^{-3/4}$ behavior as any of the $(0, \frac{3}{8})$ fields is taken to infinity then implies that f is bounded at infinity and so a constant. Taking the limit of the four-point function as $\bar{z}_{12} \to 0$, the term of order $\bar{z}_{12}^{-3/4}$ implies that $f = 1$. The term of order $\bar{z}_{12}^{1/4}$ then implies

$$\left\langle \tilde{\jmath}(\bar{z}_2)\tilde{\Sigma}(\bar{z}_3)\tilde{\bar{\Sigma}}(\bar{z}_4)\right\rangle = \frac{3\bar{z}_{34}^{1/4}}{2\bar{z}_{23}\bar{z}_{24}}\,, \tag{18.5.7}$$

so that in particular $\tilde{\jmath}$ is nonzero. The further limits $\bar{z}_{23} \to 0$, $\bar{z}_{24} \to 0$, and

$\bar{z}_{34} \to 0$ then reveal that

$$\tilde{\jmath}(\bar{z})\tilde{\Sigma}(0) \sim \frac{3}{2\bar{z}}\tilde{\Sigma}(0)\,, \tag{18.5.8a}$$

$$\tilde{\jmath}(\bar{z})\tilde{\bar{\Sigma}}(0) \sim -\frac{3}{2\bar{z}}\tilde{\bar{\Sigma}}(0)\,, \tag{18.5.8b}$$

$$\tilde{\jmath}(\bar{z})\tilde{\jmath}(0) \sim \frac{3}{\bar{z}^2}\,. \tag{18.5.8c}$$

As in the discussion of bosonization, the $\tilde{\jmath}\tilde{\jmath}$ OPE implies that the expectation values of the current can be written in terms of those of a right-moving boson $\tilde{H}$,

$$\tilde{\jmath}(\bar{z}) = 3^{1/2} i\bar{\partial}\tilde{H}(\bar{z})\,. \tag{18.5.9}$$

The energy-momentum tensor separates into one piece constructed from the current and another commuting with it,

$$\tilde{T}_B = -\frac{1}{2}\bar{\partial}\tilde{H}\bar{\partial}\tilde{H} + \tilde{T}'_B\,. \tag{18.5.10}$$

The $\tilde{\jmath}\tilde{\Sigma}$ OPE implies that

$$\tilde{\Sigma} = \exp(3^{1/2} i\tilde{H}/2)\tilde{\Sigma}'\,, \tag{18.5.11}$$

with $\tilde{\Sigma}'$ commuting with the current. The weight of the exponential is $(0, \frac{3}{8})$, the same as that of $\tilde{\Sigma}$ itself, so $\tilde{\Sigma}'$ is of weight (0,0) and must be the identity. Thus the R ground state operators are functions only of the free field,

$$\tilde{\Sigma} = \exp(3^{1/2} i\tilde{H}/2)\,, \quad \tilde{\bar{\Sigma}} = \exp(-3^{1/2} i\tilde{H}/2)\,. \tag{18.5.12}$$

Now consider the supercurrent T_F of the compact CFT. Since $\tilde{\Sigma}$ and $\tilde{\bar{\Sigma}}$ are primary fields in the R sector and are annihilated by $\tilde{G}_0$, we have

$$\tilde{T}_F(\bar{z})\tilde{\Sigma}(0) = O(\bar{z}^{-1/2})\,, \quad \tilde{T}_F(\bar{z})\tilde{\bar{\Sigma}}(0) = O(\bar{z}^{-1/2})\,. \tag{18.5.13}$$

Using the explicit form (18.5.12), this implies

$$\tilde{T}_F = \tilde{T}_F^+ + \tilde{T}_F^-\,, \tag{18.5.14a}$$

$$\tilde{T}_F^+ \propto \exp(i\tilde{H}/3^{1/2})\,, \quad \tilde{T}_F^- \propto \exp(-i\tilde{H}/3^{1/2})\,. \tag{18.5.14b}$$

In other words,

$$\tilde{\jmath}(\bar{z})\tilde{T}_F^+(0) \sim \frac{1}{\bar{z}}\tilde{T}_F^+(0)\,, \quad \tilde{\jmath}(\bar{z})\tilde{T}_F^-(0) \sim -\frac{1}{\bar{z}}\tilde{T}_F^-(0)\,. \tag{18.5.15}$$

Applying the Jacobi identity, one obtains the full (0,2) superconformal OPE (11.1.4).

To summarize, the existence of $N = 1$ supersymmetry in *spacetime* implies the existence of $N = 2$ right-moving superconformal symmetry on the *world-sheet*. That is, there is at least (0,2) superconformal symmetry.

The various components of the spacetime supersymmetry current are now known explicitly in terms of free scalar fields; for example

$$\tilde{\jmath}_{\frac{1}{2}\frac{1}{2}} = \exp\left[\tfrac{1}{2}(-\tilde{\phi} + i\tilde{H}_0 + i\tilde{H}_1 + 3^{1/2} i\tilde{H})\right]\,. \tag{18.5.16}$$

Single-valuedness of this current with any vertex operator thus implies that all states have integer charge under

$$\tilde{\jmath}_{\mathrm{GSO}} = \frac{1}{2}\bar{\partial}(-\tilde{\phi} + i\tilde{H}_0 + i\tilde{H}_1 + 3^{1/2} i\tilde{H})\,. \tag{18.5.17}$$

This integer charge condition is the generalization of the GSO projection.

The converse holds as well: if the (0,1) world-sheet supersymmetry of the heterotic string is actually embedded in a (0,2) or larger algebra, and if all states carry integer charge under the current $\tilde{J}$, then the theory has spacetime supersymmetry. The argument is simple: if there is an $N = 2$ right-moving supersymmetry, then by bosonizing the current $\tilde{J}$ we can construct the operator (18.5.16). This is a (0,1) field, a world-sheet current. By the integer charge assumption it is local with respect to all the vertex operators, and so has a well-defined action on the physical states. It is a spacetime spinor and so corresponds to a spacetime supersymmetry. Lorentz and CPT invariance generate the remaining components of the supersymmetry current (18.5.1). Combining these currents with ∂X^μ gives the gravitino vertex operators, so the supersymmetry is local.

The same argument can be applied to extended spacetime supersymmetry. The analysis is a bit longer and is left to the references, but we summarize the results. If there is $N = 2$ spacetime supersymmetry in the heterotic string, then the right-moving internal CFT separates into two pieces. The first, with $\tilde{c} = 3$, is a specific (0,2) superconformal theory: two free scalars and two free fermions forming the standard (0,2) superfield discussed in section 11.1. The second, with $\tilde{c} = 6$, must have (0,4) supersymmetry but is otherwise arbitrary. If there is $N = 4$ spacetime supersymmetry, then the right-moving internal CFT consists precisely of six free scalars and six free fermions — in other words, it is a toroidal theory.

18.6 Low energy actions

In section 16.4 we obtained the low energy effective action for the $\mathbf{Z}_3$ orbifold. Several important features of that action actually hold at string tree level for all four-dimensional string theories with $N = 1$ supersymmetry:

1. The Kähler potential is $-\kappa^{-2}\ln(S + S^*)$ plus terms independent of S.

2. The superpotential is independent of S.

3. The nonminimal gauge kinetic term is

$$f_{ab} = \frac{2\hat{k}_a \delta_{ab}}{g_4^2} S \ . \tag{18.6.1}$$

Such general results are not surprising from a world-sheet point of view. The vertex operators for Φ_4 and a involve only the noncompact free fields X^μ and $\tilde{\psi}^\mu$, which are independent of the compactification. The gauge boson vertex operators involve only these fields and the $(1,0)$ gauge currents, which again are universal up to the coefficient $\hat{k}$.

Rather than a detailed world-sheet derivation, it is very instructive to give a derivation based on the spacetime effective action. The introduction (16.4.12) of the axion field depends only on the four-dimensional fields and so is always valid. Under a shift $a \to a + \epsilon$ the action changes only by a term proportional to

$$\int F_2 \wedge F_2 \ . \tag{18.6.2}$$

This is a topological invariant and vanishes in perturbation theory. In perturbation theory there is then a PQ symmetry

$$S \to S + i\epsilon \ . \tag{18.6.3}$$

Second, there is a scale invariance: under

$$S \to tS \ , \quad G_{\mu\nu 4\mathrm{E}} \to tG_{\mu\nu 4\mathrm{E}} \ , \tag{18.6.4}$$

with the other bosonic fields invariant, the action changes by

$$\boldsymbol{S} \to t\boldsymbol{S} \ . \tag{18.6.5}$$

This is just the statement that a constant dilaton only appears in the world-sheet action multiplying the world-sheet Euler number. The scaling (18.6.4) of the metric arises because the Einstein metric differs from the string metric by a function of the dilaton.

The PQ symmetry requires that the Kähler potential depend only on $S + S^*$. In the kinetic term for S, the metric contributes a scaling t and so this term must be homogeneous in S; this determines the form given above for the Kähler potential.[6] In the gauge kinetic term, the metric contributes no net t-dependence so f_{ab} must scale as t; by holomorphicity it must be proportional to S. The PQ symmetry then requires that it depend on no other fields, in order that the variation ϵ multiply the topological term (18.6.2). The dependence on $\hat{k}_a$ was obtained in section 18.2. It is

[6] Scale invariance seems to allow an additional term $(C + C^*)\ln(S + S^*)$, where C is any other superfield. To rule this out we appeal to the world-sheet argument that an off-diagonal metric $G_{C\bar{S}}$ is impossible because the CFT factorizes.

often conventional to choose the additive normalization of the dilaton and the multiplicative normalization of the axion to eliminate g_4,

$$f_{ab} = \delta_{ab} \frac{S}{8\pi^2} \ . \tag{18.6.6}$$

The physical value of the coupling is then

$$\frac{g_{\mathrm{YM}}^2}{8\pi^2} = \frac{1}{\mathrm{Re}\langle S \rangle} \ . \tag{18.6.7}$$

PQ invariance and the holomorphicity of the superpotential together require that the superpotential be independent of S. This is precisely consistent with the scaling of the action. To see this consider the term

$$\int d^4x (-G_{4\mathrm{E}})^{1/2} \exp(\kappa^2 K) \left(K^{\bar{i}j} W_{\bar{i}}^* W_{;j} - 3\kappa^2 W^* W \right) \tag{18.6.8}$$

in the potential (B.2.29). There is a scale-dependence t^2 from the metric and t^{-1} from $\exp(\kappa^2 K)$, and so the action has the correct scaling if the superpotential is scale-invariant.

One of the great strengths of this kind of argument is that it gives information to all orders of perturbation theory, and even nonperturbatively. An L-loop term in the effective action will scale as

$$\boldsymbol{S}_L \to t^{1-L} \boldsymbol{S}_L \ . \tag{18.6.9}$$

It follows from consideration of the potential again that an L-loop term in the superpotential scales as t^{-L}. PQ invariance requires $(S + S^*)^{-L}$ while holomorphicity requires S^{-L}, so only tree level is allowed, $L = 0$. This is an easy demonstration of one of the most important nonrenormalization theorems. The original proof in field theory involved detailed graphical manipulations; a parallel argument can be constructed in string perturbation theory using contour arguments. This nonrenormalization theorem has many important consequences. For example, particle masses or Yukawa couplings that vanish at tree level also vanish to all orders in perturbation theory (except in certain cases where D-terms are renormalized, as discussed in the next section).

For the gauge kinetic term f an L-loop contribution will scale as t^{1-L}. Again it must be holomorphic and PQ-invariant, allowing only $L = 1$, or $L = 0$ with the precise field dependence S. Thus, aside from this tree-level term f receives only one-loop corrections.[7] With $N = 1$ supersymmetry there are no such constraints on the Kähler potential because it need not

[7] Such statements are often rather subtle in that one must be precise about what is not being renormalized. The discussion in section 16.4 of the physical coupling versus the Wilsonian action illustrates some of the issues.

be holomorphic. An L-loop term $(S+S^*)^{-L}$ times any function of the other fields is allowed.

The PQ symmetry is broken by nonperturbative effects because the integral of $F_2 \wedge F_2$ is nonzero for a topologically nontrivial instanton field. The superpotential and gauge kinetic terms can then receive corrections, which can often be determined exactly. We will see an example of a nonperturbative superpotential below.

One final point: there is a useful general result about the metric for the space of scalar fields. Suppose we have a compactification with some moduli ϕ^i, which we take to be real. The world-sheet Lagrangian density $\mathscr{L}_{\rm ws}$ is a function of the ϕ^i. One result of the analysis of string perturbation theory in chapter 9 was that the Zamolodchikov metric $\langle\!\langle\ |\ \rangle\!\rangle$, which is the two-point function on the sphere, determines the normalization of the vertex operators. In other words, the inner product of the string states created by ϕ^i and ϕ^j is

$$\mathscr{G}_{ij} = \left\langle\!\!\left\langle \frac{\partial \mathscr{L}_{\rm ws}}{\partial \phi^i} \middle| \frac{\partial \mathscr{L}_{\rm ws}}{\partial \phi^j} \right\rangle\!\!\right\rangle . \tag{18.6.10}$$

This implies that the kinetic term for these fields is

$$\frac{1}{2}\mathscr{G}_{ij}\partial_\mu \phi^i \partial^\mu \phi^j . \tag{18.6.11}$$

Thus the Zamolodchikov metric *is* the metric on moduli space. This result does not depend on having world-sheet supersymmetry, although in this case we have the additional information that the manifold is complex and Kähler.

18.7 Supersymmetry breaking in perturbation theory

Supersymmetry breaking at tree level

Now we would like to consider the spontaneous breaking of supersymmetry, with particular attention to the fact that the supersymmetry breaking scale is far below the string scale. The first question is whether it is possible to find examples having this property at string tree level. In fact it seems to be essentially impossible to do so.

Here is an example which illustrates the main issue. Consider the heterotic string on a simple cubic torus, $X^m \cong X^m + 2\pi R_m$ for $m = 4, \ldots, 9$, except that the translation in the 7-direction is accompanied by a $\pi/2$ rotation in the (8,9) plane. In other words, the (7,8,9)-directions form a cube with opposite faces identified, with a $\pi/2$ twist between one pair of opposite faces. This fits in the general category of orbifold models. However, the space is nonsingular because the combined rotation and translation has no fixed points. The rotation

$$(\phi_2, \phi_3, \phi_4) = (0, 0, \tfrac{1}{2}\pi) \tag{18.7.1}$$

is not in $SU(3)$ and so all the supersymmetries are broken. However, there is a limit, $R_7 \to \infty$, where the identification in the 7-direction becomes irrelevant and supersymmetry is restored. More explicitly, the effect of the twist is that $p_7 R_7$ for any state is shifted from integer values by an amount proportional to the spin s_4, thus splitting the boson and fermion masses. This is the *Scherk–Schwarz mechanism.* The mass-squared splittings are of order R_7^{-2} and so go to zero as the 7-direction decompactifies. The obvious problem with this is that the supersymmetry breaking scale is tied to the compactification scale, which is inconsistent with the discussion in section 18.3. This linking of the supersymmetry breaking and compactification scales appears to be a generic problem with tree-level supersymmetry breaking. We could avoid it in the above example by taking instead the angle $\phi_4 \to 0$; however, crystallographic considerations limit ϕ_4 to a finite set of discrete values. Note that a twist acting on $\tilde{\psi}^m$ without acting on X^m would be a symmetry of the CFT for any values of ϕ_4, but would not commute with $\tilde{T}_F$ and so would render the theory inconsistent.

There is a theorem that greatly restricts the possibilities for a large ratio of scales at tree level. The simplest way to obtain such a ratio would be to start with a supersymmetric vacuum and turn on a modulus that breaks the supersymmetry. Vacua in the neighborhood of the supersymmetric point would then have arbitrarily small breaking. However, this situation is not possible. If there is a continuous family of string vacua with vanishing cosmological constant, then either all members of the family are spacetime supersymmetric, or none is. We will give both a world-sheet and a spacetime demonstration of this.

On the world-sheet, we know that the supersymmetric point has (0,2) supersymmetry with a quantized $U(1)$ charge. As we move away from this point either the supersymmetry must be broken to (0,1), which in particular implies that the $U(1)$ in the (0,2) algebra is broken, or we must shift the quantization of the charge. To obtain either effect the vertex operator for the modulus must depend on the boson $\tilde{H}$. It can be shown that this is impossible; the argument makes rather detailed use of the (0,2) world-sheet algebra so we defer it to the next chapter.

For the spacetime argument, let us denote the modulus as t, with $t = 0$ the supersymmetric point. The condition that the potential (B.2.29) be flat is

$$(\partial_t \partial_{\bar{t}} K)^{-1} |\partial_t W + \kappa^2 \partial_t K W|^2 = 3\kappa^2 |W|^2 . \tag{18.7.2}$$

We assume that the modulus is neutral so that the D-term potential

vanishes, but the argument can be extended to the case that it is not. Physically, the metric $\partial_t \partial_{\bar{t}} K$ must be nonvanishing and nonsingular. As a differential equation for W, the condition (18.7.2) then implies that if W vanishes for any t then it vanishes for all t, as claimed. This shows that a continuous family of string vacua with zero cosmological constant cannot include both supersymmetric and nonsupersymmetric states in any theory with $N = 1$ supergravity, independent of string theory.

The Scherk–Schwarz mechanism gives arbitrarily small supersymmetry breaking, but the supersymmetric point $R_7 = \infty$ is at infinite distance. This evades the theorem but it is also what makes this example uninteresting. One could try to evade the theorem with a small discrete rather than continuous parameter. For example, the Sugawara $SU(2)$ theories have $c = 3 - 6/(k+2)$ with k an integer, and so cluster arbitrarily closely to $c = 3$ as $k \to \infty$. However, all attempts based on free, solvable, or smooth compactifications have run into the decompactification problem.

Supersymmetry breaking in the loop expansion

The conditions for unbroken supersymmetry are

$$W(\phi) = \partial_i W(\phi) = D^a(\phi, \phi^*) = 0 \,. \tag{18.7.3}$$

Now let us suppose that these conditions are satisfied at tree level and ask whether loop corrections can lead to them being violated. We know that the superpotential does not receive loop corrections, so the first two conditions will continue to hold to all orders. For non-Abelian D-terms, the vanishing of the D^a is implied by the gauge symmetry, so the key issue is the $U(1)$ D-terms.

The D-term potential is

$$V = \mathrm{Re}[(S/8\pi^2) + f_1(T)]\frac{D^2}{2} \tag{18.7.4a}$$

$$D = \frac{1}{\mathrm{Re}[(S/8\pi^2) + f_1(T)]}\left(2\xi - i\kappa^2 K_{,i}\frac{\delta\phi^i}{\delta\lambda}\right) . \tag{18.7.4b}$$

Here $\delta\phi^i/\delta\lambda$ is the $U(1)$ variation of the given scalar ϕ^i. We have used what we know about the gauge kinetic term — the threshold correction f_1 is included for completeness, but it is subleading and makes no difference in the following discussion. The scaling property (18.6.9) (which includes the scaling of the $(-G)^{1/2}$ in the action) implies that an L-loop contribution to the potential scales as t^{-L-1} and therefore as

$$S^{-L-1} \,. \tag{18.7.5}$$

Consider first the possibility of a nonzero Fayet–Iliopoulos term ξ being generated in perturbation theory. Expanding in powers of $1/S$, the leading term in the potential is of order $\xi^2/\mathrm{Re}(S)$. This is a *tree-level* effect, and so by assumption is absent.

Now consider the effect of gauging the PQ symmetry associated with S,

$$\delta S = iq\delta\lambda \,. \tag{18.7.6}$$

With the known form of the Kähler potential for S, the leading potential is

$$V \propto \frac{q^2}{(S+S^*)^3} \,. \tag{18.7.7}$$

This is a two-loop effect, so D itself is a one-loop effect. To see the significance of the variation (18.7.6), consider the effect on the PQ coupling

$$\delta\frac{1}{4\pi^2}\int \mathrm{Im}(S) F_2^a \wedge F_2^a = \frac{q\delta\lambda}{4\pi^2}\int F_2^a \wedge F_2^a \,. \tag{18.7.8}$$

This is not gauge-invariant but has just the right form to cancel against a one-loop anomaly in the gauge transformation, if the low energy fermion spectrum produces one. In fact, many compactifications do have anomalous spectra, and the anomaly is canceled by the variation (18.7.8) in a four-dimensional version of the Green–Schwarz mechanism. This is accompanied by cancellation of a gravitational anomaly. The induced D-term is proportional to $\mathrm{Tr}(Q)$, the total $U(1)$ charge of all massless left-handed fermions.

Thus $D \neq 0$ precisely if $\mathrm{Tr}(Q) \neq 0$, and then the supersymmetry of the original configuration is broken by a one-loop effect. The important question is whether the system can relax to a nearby supersymmetric configuration. The full D-term, including the other charged fields, is

$$D = \frac{q}{(S+S^*)} + \sum_{\phi^i \neq S} q_i \phi^{i*} \phi^i \tag{18.7.9}$$

and the potential is proportional to the square of this. If we can give the various ϕ^i small expectation values, of order $(S+S^*)^{-1/2}$, such that the D-term is set to zero while preserving $W = \partial_i W = 0$, then there is a supersymmetric minimum near the original configuration. In fact, in the known examples this is the case. Notice that while supersymmetry is restored, the new vacuum is qualitatively different from the original one. In particular, the $U(1)$ gauge symmetry is now broken by the expectation value of e^S, and the gauge boson is massive. Being a one-loop effect, the gauge boson mass-squared is of order $g^2/8\pi^2$ times the string scale. Thus the one-loop D-term produces a modest hierarchy of scales; this might be useful, for example, in accounting for the pattern of quark and lepton masses. Other massless particles may also become massive due to the shift in the ϕ^i. These are effects that cannot occur with only F-terms in the potential.

It is also interesting to consider the case that the PQ-like symmetry associated with the (1,1) moduli T^A is gauged,

$$\delta T^A = iq^A \delta\lambda \,. \tag{18.7.10}$$

To leading order in S the potential is then

$$V = \frac{(q^A \partial_A K)^2}{(S+S^*)} \,. \tag{18.7.11}$$

This is a tree-level effect. We are assuming that we have a supersymmetric tree-level solution, which is still possible on the submanifold of moduli space where $q^A \partial_A K = 0$. The would-be moduli orthogonal to this submanifold are all massive. There is a natural origin for the gauge transformation (18.7.10). The imaginary part of T^A is the integral of B_2 over the 2-cycle N^A. In the heterotic string the gauge variation of B_2 is proportional to $\mathrm{Tr}(\delta\lambda F_2)$, so if the $U(1)$ field strength has an expectation value there is a transformation

$$\delta T^A \propto i\delta\lambda \int_{N^A} F_2 \,. \tag{18.7.12}$$

This is automatically absent for Calabi–Yau compactification, because the integral of the flux measures the first Chern class. This is also another example of the difficulty of breaking supersymmetry by a small amount at tree level. It might seem that we could break the supersymmetry of the $q^A \partial_A K \neq 0$ vacua slightly by making F_2 small, but the integral of F_2 over any 2-cycle must satisfy a Dirac quantization condition. By a generalization of the monopole argument, $q_i \int F_2$ must be a multiple of 2π, where F_2 is proportional to any $U(1)$ generator of $E_8 \times E_8$, and q_i runs over the $U(1)$ charges of all heterotic string states.

18.8 Supersymmetry beyond perturbation theory

An example

In the previous section, we saw that a vacuum that is supersymmetric at tree level usually remains supersymmetric to all orders of perturbation theory. Remarkably, it is known that in most tree-level $N = 1$ vacua the supersymmetry is broken spontaneously by nonperturbative effects. Our understanding of nonperturbative string theory is still limited, but below the string scale we can work in the effective quantum field theory. In fact, there is a reasonably coherent understanding of nonperturbative breaking of supersymmetry in field theory, and the low energy theories emerging from the string theory are typically of the type in which this breaking occurs. This subject is quite involved; there are several symmetry-breaking mechanisms (gaugino condensation, instantons, composite goldstinos), and

a variety of techniques are needed to unravel the physics. Fortunately, we can get a good idea of the issues by focusing on the simplest mechanism, gluino condensation, in the simplest $N = 1$ vacua.

Consider any (2,2) compactification, with the visible E_6 possibly broken by Wilson lines. The hidden E_8 generally has a large negative beta function

$$\beta_8 = \frac{b_8}{16\pi^2} g_{E_8}^3 \,, \quad -b_8 \gg 1 \,. \tag{18.8.1}$$

The running coupling is

$$g_{E_8}^2(\mu) = \frac{8\pi^2}{\mathrm{Re}(S) + b_8 \ln(m_s/\mu)} \tag{18.8.2}$$

(for the present discussion we are not concerned about the small numerical difference between m_s and m_{SU}), and so becomes strong at a scale

$$\Lambda_8 = m_s \exp[-\mathrm{Re}(S)/|b_8|] \,. \tag{18.8.3}$$

This is below the string scale but above the scale where any of the visible sector groups become strong. Just as with quarks in QCD, the strong attraction causes the gauginos to condense,

$$|\langle(\bar\lambda\lambda)_{\mathrm{hidden}}\rangle| \approx \Lambda_8^3 \,. \tag{18.8.4}$$

Here and below '$\approx$' means up to numerical coefficients. As in QCD this condensate breaks a chiral symmetry, but in the pure supersymmetric gauge theory (gauge fields and gauginos only) it is known *not* to break supersymmetry.

In string theory at tree level the fields of the hidden E_8 couple to precisely one other light superfield, namely S. We have discussed the coupling of the dilaton and the axion to the field strength, but in addition supersymmetry requires a coupling between the auxiliary field and the gauginos

$$\kappa F_S (\bar\lambda\lambda)_{\mathrm{hidden}} \,. \tag{18.8.5}$$

At scales below Λ_8 this looks like an effective interaction

$$\kappa F_S \langle(\bar\lambda\lambda)_{\mathrm{hidden}}\rangle \approx F_S \kappa m_{SU}^3 \exp(-3S/|b_8|) \,. \tag{18.8.6}$$

From the general $N = 1$ action (B.2.16) this implies an effective superpotential[8]

$$W \approx \kappa m_{SU}^3 \exp(-3S/|b_8|) \,. \tag{18.8.7}$$

This superpotential is nonperturbative, vanishing at large S faster than any power of $1/S$. This is an example of the violation of a perturbative

[8] This must be holomorphic in S, whereas the scale Λ_8 depends on $\mathrm{Re}(S)$. The point is that the phase of the condensate depends on the axion in just such a way as to account for the difference.

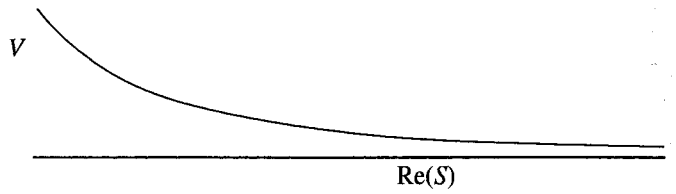


Fig. 18.3. The potential in a simple model of gluino condensation, as a function of the dilaton with other moduli held fixed.

nonrenormalization theorem by nonperturbative effects. This superpotential is not PQ-invariant, which is consistent with the earlier discussion.

What is more, this superpotential breaks supersymmetry. At tree level and to all orders of perturbation theory, the vacuum is supersymmetric for any value of S. Nonperturbatively,

$$F_S = \frac{\partial W}{\partial S} \approx \kappa m_{\rm SU}^3 \exp(-3S/|b_8|) \tag{18.8.8}$$

is nonzero, which is the criterion (B.2.25) for the breaking of supersymmetry. This simple model is not satisfactory because the potential is roughly

$$V \approx \kappa^2 m_{\rm SU}^6 (S+S^*)^k \exp[-3(S+S^*)/|b_8|] \,. \tag{18.8.9}$$

The power of $S+S^*$ comes from the Kähler potential for S and from the two-loop beta function. At small coupling (large S), where the calculation is valid, the potential has the qualitative form shown in figure 18.3 and there is no stable vacuum. Rather, the system rolls down the potential toward the point $\mathrm{Re}(S)=\infty$, where the theory is free and supersymmetric.

We will consider the problem of stabilizing the dilaton shortly, but for now let us see what happens if we assume that some higher correction, additional gauge group, or other modification gives rise to a stable supersymmetry-breaking vacuum at a point where S has roughly the value $8\pi^2/g_{\rm YM}^2 \approx 100$ found in simple grand unified models. The number 100 seems large, but noting that $|b_8| = 90$ this is actually the typical scale for the S-dependence.

Having broken supersymmetry, the next question is how this affects the masses of the ordinary quarks, leptons, gauge bosons, and their superpartners. The only tree-level coupling of the supersymmetry breaking field S to these fields is again through a gauge kinetic term, that of the Standard Model gauge fields. Thus F_S has a coupling of the same form as (18.8.5) but to the ordinary gauginos. Inserting the expectation value for F_S gives

a gaugino mass term,

$$\kappa\langle F_S\rangle\bar\lambda\lambda \approx \kappa^2 m_{\rm SU}^3 \exp(-3\langle S\rangle/|b_8|)\bar\lambda\lambda \,. \tag{18.8.10}$$

The mass is

$$m_\lambda \approx \kappa^2 m_{\rm SU}^3 \exp(-3\langle S\rangle/|b_8|) \approx \exp(-3\langle S\rangle/|b_8|) \times 10^{18}\ \text{GeV} \,. \tag{18.8.11}$$

To solve the Higgs naturalness problem the masses of the Standard Model superpartners must be of order 10^3 GeV or less. For the values $S \approx 100$ and $|b_8| = 90$ of this simple model this is not the case, but because these parameters appear in the exponent a modest ratio of parameters $S/|b| \approx 12$ would produce the observed large ratio of mass scales.

Once masses are generated for the Standard Model gauginos, loop corrections will give mass to the scalar partners of quarks and leptons. There is a simple reason why the (yet unseen) superpartners receive masses in this way while the quarks, leptons and gauge bosons do not: the latter masses are all forbidden by gauge invariance. Another feature to be understood is the negative mass-squared of the Higgs scalar, needed to break $SU(2)\times U(1)$, while the quark and lepton scalars must have positive masses-squared to avoid breaking baryon and lepton number. Again there is a simple general explanation, namely the one-loop correction to the Higgs potential coming from a top quark loop; the large top quark mass is just what is needed for this to work. The mass scale of the superpartners then determines the weak interaction scale.

The enormous ratio

$$\frac{m_{\rm ew}}{m_{\rm grav}} \approx 10^{-16} \tag{18.8.12}$$

thus arises ultimately from an exponent of order 10 in Λ_8, eq. (18.8.3). The renormalization group has this effect of amplifying modestly small couplings into large hierarchies. Thus, assuming the necessary stable vacuum, the enormous ratio of the weak and gravitational scales could emerge from a theory that has no free parameters.

We should point out that there is a distinction between the mass scale $m_{\rm sp}$ of the Standard Model superpartners and the scale $m_{\rm SUSY} = F_S^{1/2}$ of the supersymmetry-breaking expectation value. In fact,

$$m_{\rm SUSY}^2 \approx m_{\rm sp} m_{\rm grav} \,, \tag{18.8.13}$$

or

$$m_{\rm sp} \approx \kappa F_S \,. \tag{18.8.14}$$

This relation has a simple interpretation: the splittings in the Standard Model are given by the magnitude of the supersymmetry-breaking expectation value times the strength of the coupling between the Standard

Model and the supersymmetry breaking. There has also been much consideration of field theory models in which the two sectors couple more strongly, through gauge interactions, and $m_{\rm SUSY}$ is correspondingly lower. Such models could arise in string theory, in (0,2) vacua.

The form of supersymmetry breaking in this particular model, from $\langle F_S \rangle$, is known as *dilaton-mediated* supersymmetry breaking. Because the couplings of the dilaton are model-independent, the resulting pattern of superpartner masses is rather simple. In particular, the induced masses for the squarks and sleptons are to good accuracy the same for all three generations. This is important to account for the suppression of radiative corrections to rare decays (flavor changing neutral currents). More generally, radiative and other corrections can lead to a less universal pattern. Also, we have neglected all moduli other than the dilaton, but we will see below a simple model in which it is one of the Calabi–Yau moduli whose auxiliary field breaks supersymmetry.

The massless dilaton appears in the tree-level spectrum of every string theory, but not in nature: it would mediate a long-range scalar force of roughly gravitational strength. Measurements of the gravitational force at laboratory and greater scales restrict any force with a range greater than a few millimeters (corresponding to a mass of order 10^{-4} eV) to be several orders of magnitude *weaker* than gravity, ruling out a massless dilaton. We see from the present model that supersymmetry breaking can, and generically will, generate a potential for the dilaton. In this case there is no stable minimum, but the second derivative of the potential gives an indication of the typical mass

$$m_\Phi \approx m_{\rm sp} \,. \tag{18.8.15}$$

The superpotential (18.8.7) does not depend on any moduli other than S. This is because the scale Λ_8 is determined by the initial value of the gauge coupling, which at tree level depends only on S. We know that the one-loop correction to the gauge coupling depends on the other moduli, and this in turn induces a dependence in the superpotential. Thus if there is a stable minimum in the potential, generically all moduli will be massive.

Cosmological questions are outside our scope, but we note in passing that there is a potential cosmological problem with the moduli, in that their current energy density must not greatly exceed the critical density for closure of the universe. Typically the range of masses 10^{-7} GeV $< m < 10^4$ GeV is problematic. Below this, the mass is sufficiently small not to present a problem; above it, the decay rate of the particles is sufficiently great. Masses at either end of the range give interesting possibilities for dark matter.

Let us give an optimistic summary. Start with the simplest heterotic string vacuum with $N = 1$ supersymmetry, namely a (2,2) orbifold or

Calabi–Yau compactification. The result is a theory very much like the picture one obtains by starting from the Standard Model and trying to account for its patterns: gauge group E_6, chiral matter in the **27** representation, and a hidden sector that breaks supersymmetry (modulo the stabilization problem) and produces a realistic spectrum of superpartner masses. Of course, things may not work out so simply in detail; we know that the set of string vacua is vast, and we do not know any dynamical reason why these simple vacua should be preferred.

Another example

It is interesting to consider the following model,

$$K = -\ln(S+S^*) - 3\ln(T+T^*) \,, \tag{18.8.16a}$$

$$W = -w + \kappa m_{\rm SU}^3 \exp(-3S/|b_8|) \,. \tag{18.8.16b}$$

The Kähler potential for T is based on the large-radius limit of Calabi–Yau compactification. Inclusion of a constant $-w$ in the tree-level superpotential is consistent with the scaling and PQ transformations. After some cancellation, the potential is proportional to a square,

$$V = \frac{(S+S^*)|W_{;S}|^2}{(T+T^*)^3} \,, \tag{18.8.17a}$$

$$W_{;S} = \frac{w}{S+S^*} - \kappa m_{\rm SU}^3 \exp(-3S/|b_8|)\left(\frac{3}{|b_8|} + \frac{1}{S+S^*}\right) \,. \tag{18.8.17b}$$

When $W_{;S} = 0$ the potential is minimized, and the value at the minimum is zero. Nevertheless supersymmetry is broken, as

$$W_{;T} = -\frac{3W}{T+T^*} \neq 0 \,. \tag{18.8.18}$$

This is intriguing: supersymmetry is broken nonperturbatively yet the vacuum energy is still zero. Also, the field T is undetermined, so there is a degenerate family of vacua with arbitrary supersymmetry-breaking scale $W_{;T}$. This is known as a *no-scale model*. The special properties of the potential depend on the detailed form of the Kähler potential and the superpotential, in particular the factor of 3 in the former and the fact that the latter is independent of T. Higher order effects will spoil this. For example, as we have noted above, threshold corrections will introduce a T-dependence into the superpotential.

Discussion

Since $S \propto g^{-2}$, the superpotential (18.8.7) is of order $\exp[-O(1/g^2)]$, which is characteristic of nonperturbative effects in field theory. It is not invariant under the PQ symmetry $S \to S + i\epsilon$ but transforms in a simple

way. This can be related to the breaking of PQ invariance by instantons, but the argument is rather indirect and we will not pursue it.

It is interesting to consider at this point the order $\exp[-O(1/g)]$ stringy nonperturbative corrections deduced from the large order behavior of string perturbation theory. For the type II string we were able to relate these to D-instantons, but there is no analogous amplitude in the heterotic string. In the type II theory the D-instanton gives rise to an effect that does not occur in any order of perturbation theory, the nonconservation of the integrated R–R 1-form field strength. In the heterotic string it is unlikely that the stringy nonperturbative effects violate the perturbative nonrenormalization theorems. They would give rise to effects proportional to one of the forms

$$\exp(CS^{1/2}), \quad \exp[C(S+S^*)^{1/2}] \tag{18.8.19}$$

with C a constant. The first form is holomorphic and the second is PQ-invariant. Corrections to the superpotential would have to be of the first form, but these have a complicated PQ transformation which is probably not allowed. In particular, it is believed that a discrete subgroup of the PQ symmetry is unbroken by anomalies; this would forbid the form $\exp(CS^{1/2})$. The nonperturbative effects could then only modify the Kähler potential, but this in any case receives corrections at all orders of perturbation theory.

Now we return to the stabilization of the dilaton. One possibility is that there are two competing strong gauge groups. In this case the dilaton potential can have a minimum, which for appropriate choices of the groups can be at the weak coupling $S \approx 100$ which is suggested by grand unification and needed for a large hierarchy. Another possibility is that a weak-coupling minimum can be produced by including the stringy nonperturbative corrections to the Kähler potential. It may seem odd that these corrections can be important at weak coupling, but it has been suggested that for the modestly small but not infinitesimal couplings of interest, the stringy nonperturbative effects can dominate the perturbative corrections. There may also be minima at very strong coupling, where the dual M-theory picture is more useful, or at couplings of order 1 which are close to neither limit.

Another idea would be that the potential really is as in figure 18.3 and that the dilaton is time-dependent, rolling toward large S. However, a brief calculation shows that these solutions cannot describe our universe: given the age of the universe, the supersymmetry breaking and gauge couplings would be far too small.

However, it is impossible to separate the stabilization of the dilaton from the cosmological constant problem. A generic potential on field space will have some number of local minima, but there is no reason

that the value of the potential at any of the minima should vanish, either exactly or to the enormous accuracy required by the upper limit on the cosmological constant. So while the dilaton is stabilized, the metric is still 'unstable,' expanding exponentially, and the vacuum is not acceptable. The cosmological constant problem afflicts any theory of gravity, not just string theory. However, since predictive power in string theory is completely dependent on understanding the dynamics of the vacuum, any detailed discussion of the determination of the vacuum is likely to be premature until we understand why the cosmological constant is so small.

In any event, our current understanding would suggest that string theory has many stable vacua. Supersymmetry guarantees that the various moduli spaces with $N = 2$ and greater supersymmetry are exact solutions. In addition there are likely moduli spaces with $N = 1$ supersymmetry but no strong gauge groups and no breaking of supersymmetry. In addition there may be a number of isolated minima of approximate $N = 1$ supersymmetry, which are the ones we seek. There are also some string states of negative energy density. These are known to exist from one-loop calculations in nonsupersymmetric vacua with vanishing tree-level cosmological constant. The reader might worry that any vacuum with zero energy density will then be unstable. However, gravitational effects can completely forbid tunneling from a state of zero energy density to a state of negative energy density if the barrier between the two is sufficiently high. The conditions for this to occur are met rather generally in supersymmetric theories.

If there are many stable vacua, which of these the universe finds itself in would be a cosmological question, depending on the initial conditions, and the answer might be probabilistic rather than deterministic. This does not imply a lack of predictive power. Assuming that we eventually understand the dynamics well enough to determine the minima, there will likely be very few with such general features of the Standard Model as three generations. The key point is that because supersymmetry breaking leaves only isolated minima, there are no effective free parameters: the moduli are all determined by the dynamics.

This rather prosaic extrapolation is likely to be modified by new dynamical ideas. In particular, whatever principle is responsible for the suppression of the cosmological constant may radically change the rules of the game.

Exercises

18.1 Calculate the tree-level string amplitude with a model-independent axion and two gauge bosons.

18.2 Show from the explicit form of the string amplitudes that no scalar other than the model-independent axion has a tree-level coupling to $F_2 \wedge F_2$.

18.3 Derive the conditions cited at the end of section 18.6 for a heterotic string theory to have $N = 2$ and $N = 4$ spacetime supersymmetry.

18.4 Calculate the Zamolodchikov metric for two untwisted moduli of the $\mathbf{Z}_3$ orbifold and compare with the result obtained in chapter 16 by dimensional reduction.

18.5 Work out the one-loop vacuum amplitude for the twisted theory described at the beginning of section 18.7.

18.6 For the $SO(32)$ heterotic string on the $\mathbf{Z}_3$ orbifold, show that the gauge and mixed gauge–gravitational anomalies are nonzero. Show that they can be canceled by giving the superfield S the gauge transformation (18.7.6). Show that the resulting potential has supersymmetric minima.

18.7 If we integrate out the auxiliary field F_S, the couplings (18.8.5) lead to a tree-level interaction of four gauge fermions. Find this interaction using string perturbation theory. Note that it is independent of the compactification.

19
Advanced topics

In this final chapter we develop a number of intertwined ideas, concerning the perturbative and nonperturbative dynamics of the heterotic and type II theories. A common thread running through much of the chapter is world-sheet $N = 2$ superconformal symmetry, and we begin by developing this algebra in more detail. We then consider type II strings on Calabi–Yau and other (2,2) SCFTs, and heterotic strings on general (2,2) SCFTs. We next study string theories based on (2,2) minimal models, which leads us also to mirror symmetry. From there we move to some of the most interesting recent discoveries, phase transitions involving a change of topology of the compact space — the perturbative flop transition and the nonperturbative conifold transition. The final two sections deal with dualities of compactified theories, the first developing K3 compactification and the second the dualities of toroidally compactified heterotic strings.

19.1 The $N = 2$ superconformal algebra

The $N = 2$ superconformal algebra in operator product form, given in eq. (11.1.4), is repeated below:

$$T_B(z)T_F^{\pm}(0) \sim \frac{3}{2z^2}T_F^{\pm}(0) + \frac{1}{z}\partial T_F^{\pm}(0)\,, \tag{19.1.1a}$$

$$T_B(z)j(0) \sim \frac{1}{z^2}j(0) + \frac{1}{z}\partial j(0)\,, \tag{19.1.1b}$$

$$T_F^{+}(z)T_F^{-}(0) \sim \frac{2c}{3z^3} + \frac{2}{z^2}j(0) + \frac{2}{z}T_B(0) + \frac{1}{z}\partial j(0)\,, \tag{19.1.1c}$$

$$T_F^{+}(z)T_F^{+}(0) \sim T_F^{-}(z)T_F^{-}(0) \sim 0\,, \tag{19.1.1d}$$

$$j(z)T_F^{\pm}(0) \sim \pm\frac{1}{z}T_F^{\pm}(0)\,, \tag{19.1.1e}$$

$$j(z)j(0) \sim \frac{c}{3z^2}\,. \tag{19.1.1f}$$

In the examples of interest the current j is single-valued with respect to all vertex operators. The Laurent expansions are then

$$T_B(z) = \sum_{n\in\mathbf{Z}} \frac{L_n}{z^{n+2}} , \qquad j(z) = \sum_{n\in\mathbf{Z}} \frac{J_n}{z^{n+1}} , \tag{19.1.2a}$$

$$T_F^+(z) = \sum_{r\in\mathbf{Z}+\nu} \frac{G_r^+}{z^{r+3/2}} , \qquad T_F^-(z) = \sum_{r\in\mathbf{Z}-\nu} \frac{G_r^-}{z^{r+3/2}} , \tag{19.1.2b}$$

where the shift ν can take any real value. The OPEs (19.1.1) correspond to the $N = 2$ superconformal algebra

$$[L_m, G_r^\pm] = \left(\frac{m}{2} - r\right) G_{m+r}^\pm , \tag{19.1.3a}$$

$$[L_m, J_n] = -nJ_{m+n} , \tag{19.1.3b}$$

$$\{G_r^+, G_s^-\} = 2L_{r+s} + (r-s)J_{r+s} + \frac{c}{3}\left(r^2 - \frac{1}{4}\right)\delta_{r,-s} , \tag{19.1.3c}$$

$$\{G_r^+, G_s^+\} = \{G_r^-, G_s^-\} = 0 , \tag{19.1.3d}$$

$$[J_n, G_r^\pm] = \pm G_{r+n}^\pm , \tag{19.1.3e}$$

$$[J_m, J_n] = \frac{c}{3} m\delta_{m,-n} . \tag{19.1.3f}$$

It was shown in section 18.5 that every heterotic string theory with $d = 4$, $N = 1$ spacetime supersymmetry has a right-moving $N = 2$ superconformal algebra. In compactifications with the spin connection embedded in the gauge connection there is also a left-moving $N = 2$ algebra. Most of this final chapter deals with string theories having such (2,2) superconformal algebras. These are interesting for a number of reasons. First, they can also be taken as backgrounds for the type II string, where they lead to $d = 4$, $N = 2$ supersymmetry. This larger supersymmetry puts strong constraints on the dynamics, even nonperturbatively. Second, the large world-sheet superconformal algebra allows us to derive many general results concerning the low energy dynamics of heterotic string compactifications. Third, there are several additional constructions of (2,2) CFTs, and an interesting interplay between the different constructions. Finally, we have explained in the previous chapter that (0,2) CFTs have several phenomenological advantages over the more restricted (2,2) theories. However, many (0,2) theories are obtained from (2,2) theories by turning on Wilson lines or moduli. Also, many of the methods and constructions that we will develop for (2,2) theories can also be applied to the (0,2) case, though with more difficulty.

Heterotic string vertex operators

In this section we consider only the right-moving supersymmetry algebra, so that the results apply to all supersymmetric compactifications of the

heterotic string. We take $\tilde{c} = 9$, as is relevant to four-dimensional theories. The local symmetry $\tilde{T}_F$ of the heterotic string is embedded in the $N = 2$ algebra as

$$\tilde{T}_F = \tilde{T}_F^+ + \tilde{T}_F^- . \tag{19.1.4}$$

The separate generators $\tilde{T}_F^\pm$ must have the same periodicity as $\tilde{T}_F$: either NS ($\nu = \frac{1}{2}$) or R ($\nu = 0$). In addition to $N = 2$ superconformal symmetry, spacetime supersymmetry implies that all states have integer charge under the current (18.5.17). In a general vertex operator proportional to

$$\exp\left[l\tilde{\phi} + is_0\tilde{H}^0 + is_1\tilde{H}^1 + i\tilde{Q}(\tilde{H}/3^{1/2})\right] , \tag{19.1.5}$$

it must then be the case that

$$l + s_0 + s_1 + \tilde{Q} \in 2\mathbf{Z} . \tag{19.1.6}$$

Given the result $\tilde{\jmath} = 3^{1/2} i\bar{\partial}\tilde{H}$ from section 18.5, it follows that $\tilde{Q}$ is the eigenvalue of $\tilde{J}_0$.

The vertex operators for the graviton, dilaton, and axion depend only on the noncompact coordinates and so are independent of compactification. For the remaining scalars, the weight $(1, \frac{1}{2})$ vertex operator in the -1 picture comes entirely from the compact CFT. The condition (19.1.6) in this case implies that $\tilde{Q}$ is an odd integer. The weight of $\exp(i\tilde{Q}\tilde{H}/3^{1/2})$ is $\tilde{h} = \tilde{Q}^2/6$, so the only possible values are $\tilde{Q} = \pm 1$ and the vertex operator takes one of the two forms

$$\mathcal{U}\exp(i\tilde{H}/3^{1/2}) , \quad \bar{\mathcal{U}}\exp(-i\tilde{H}/3^{1/2}) , \tag{19.1.7}$$

with $\mathcal{U}$ having weight $(1, \frac{1}{3})$.

For fermions from the compact CFT, the internal part has weight $(1, \frac{3}{8})$. When the four-dimensional spinor is a **2**, then $s_0 + s_1$ is an odd integer and the allowed values of $\tilde{Q}$ are $\frac{3}{2}$ and $-\frac{1}{2}$, giving the vertex operators

$$j^a\exp(3^{1/2}i\tilde{H}/2) , \quad \mathcal{U}\exp[-i\tilde{H}/(2\times 3^{1/2})] . \tag{19.1.8}$$

For $\tilde{Q} = \frac{3}{2}$, the exponential saturates the right-moving weight $\frac{3}{8}$ and is identical to the compact part of the spacetime supercharge. The remaining factor j is a $(1,0)$ current, so this state is a gaugino. For $\tilde{Q} = -\frac{1}{2}$, the remaining factor $\mathcal{U}$ is of weight $(1, \frac{1}{3})$, just as for the scalar. Because these theories have spacetime supersymmetry there is an isomorphism between the scalar and fermionic spectra. The OPE with the compact part $\exp(\pm 3^{1/2}i\tilde{H}/2)$ of the supercharge, which has $\tilde{Q} = \pm\frac{3}{2}$, relates the bosonic states with $\tilde{Q} = +1$ to the fermionic states with $\tilde{Q} = -\frac{1}{2}$. Similarly when the four-dimensional spinor is a $\mathbf{2}'$, then $\tilde{Q}$ must be $\frac{1}{2}$ or $-\frac{3}{2}$, giving the vertex operators

$$\bar{\mathcal{U}}\exp[i\tilde{H}/(2\times 3^{1/2})] , \quad j^a\exp(-3^{1/2}i\tilde{H}/2) . \tag{19.1.9}$$

Chiral primary fields

The $N = 2$ superconformal algebra includes the anticommutators

$$\{\tilde{G}^+_{1/2}, \tilde{G}^-_{-1/2}\} = 2\tilde{L}_0 + \tilde{J}_0 \,, \tag{19.1.10a}$$

$$\{\tilde{G}^+_{-1/2}, \tilde{G}^-_{1/2}\} = 2\tilde{L}_0 - \tilde{J}_0 \,. \tag{19.1.10b}$$

We use the right-moving notation consistent with our convention for the heterotic string, but now allow arbitrary central charge. For central charge $\tilde{c}$, the bosonization of the $\tilde{J}\tilde{J}$ OPE implies that

$$\tilde{J} = i(\tilde{c}/3)^{1/2}\tilde{\partial}\tilde{H} \,. \tag{19.1.11}$$

Taking the expectation values of the anticommutators (19.1.10) in any state, the left-hand side is nonnegative and so

$$2\tilde{h} \geq |\tilde{Q}| \,. \tag{19.1.12}$$

Let us consider an NS state $|c\rangle$ that saturates this inequality with $\tilde{Q} = 2\tilde{h}$. Such a state has the properties

$$\tilde{G}^\pm_r|c\rangle = 0 \,, \quad r > 0 \,, \tag{19.1.13a}$$

$$\tilde{L}_n|c\rangle = \tilde{J}_n|c\rangle = 0 \,, \quad n > 0 \,, \tag{19.1.13b}$$

$$\tilde{G}^+_{-1/2}|c\rangle = 0 \,. \tag{19.1.13c}$$

The first two lines state that $|c\rangle$ is annihilated by all of the lowering operators in the $N = 2$ algebra and so is an $N = 2$ superconformal primary field. The additional property of being annihilated by $\tilde{G}^+_{-1/2}$ defines a *chiral primary field*. To derive (19.1.13), note that all of the lowering operators except for $\tilde{G}^-_{1/2}$ take $|c\rangle$ into a state that would violate the inequality (19.1.12), and so must annihilate it. The expectation value of the anticommutator (19.1.10b) further implies that $\tilde{G}^-_{1/2}$ and $\tilde{G}^+_{-1/2}$ annihilate $|c\rangle$, giving the rest of eq. (19.1.13). A state with $\tilde{Q} = -2\tilde{h}$ is similarly a superconformal primary field that is also annihilated by $\tilde{G}^-_{-1/2}$, and is known as an *antichiral primary field*. The free boson $\tilde{H}$ contributes $3\tilde{Q}^2/2\tilde{c}$ to the weight of any state, so chiral primaries are possible only if

$$\frac{3\tilde{Q}^2}{2\tilde{c}} \leq \frac{|\tilde{Q}|}{2} \;\Rightarrow\; |\tilde{Q}| \leq \frac{\tilde{c}}{3} \,. \tag{19.1.14}$$

In particular the NS vertex operators (19.1.7), with $\tilde{Q} = \pm 1$ and $\tilde{h} = \frac{1}{2}$, are chiral and antichiral primaries. This property will be useful later. For the present we just use it to complete an argument from the previous chapter. We have seen that the -1 picture massless vertex operators have $U(1)$ charge $\tilde{Q} = \pm 1$. Acting with $G_{-1/2}$ to obtain the 0 picture operators could give $\tilde{Q} = \pm 2$ or 0. However, the chiral and antichiral properties imply that the terms with $\tilde{Q} = \pm 2$ vanish, so that the 0 picture operator

must have $\tilde{Q} = 0$ and can depend on $\tilde{H}$ only through its derivative. Dimensionally it can then only be linear in $\bar{\partial}\tilde{H}$. Acting with $\tilde{J}_1$ picks out the coefficient of $\bar{\partial}\tilde{H}$,

$$\tilde{J}_1 \cdot \mathscr{V}^0 = \tilde{J}_1 G_{-1/2} \cdot \mathscr{V}^{-1} = (\tilde{G}_{-1/2}\tilde{J}_1 + \tilde{G}^+_{1/2} - \tilde{G}^-_{1/2}) \cdot \mathscr{V}^{-1} = 0 \,, \tag{19.1.15}$$

the final equality holding because $\mathscr{V}^{-1}$ is primary. The 0 picture vertex operator is the change in the world-sheet action when a modulus is varied. We have established that this is independent of $\tilde{H}$, as needed above eq. (18.7.2).

Spectral flow

Suppose that we have a representation of the $N = 2$ algebra (19.1.3) with some periodicity ν. Imagine shifting the $U(1)$ charge of every state by $-\tilde{c}\eta/3$, so that the free boson part of any vertex operator is shifted

$$\exp\left[i(3/\tilde{c})^{1/2}\tilde{Q}\tilde{H}\right] \to \exp\left[i(3/\tilde{c})^{1/2}\tilde{Q}\tilde{H} - i\eta(\tilde{c}/3)^{1/2}\tilde{H}\right] \,. \tag{19.1.16}$$

From their $U(1)$ charges we know that the $\tilde{T}^\pm_F$ depend on the free boson as $\exp[\pm i(3/\tilde{c})^{1/2}\tilde{H}]$. Then from the OPE of this factor with the exponential (19.1.16) it follows that the periodicity of $\tilde{T}^\pm_F$ with respect to any vertex operator shifts,

$$\nu \to \nu + \eta \,. \tag{19.1.17}$$

By this shift of the $U(1)$ charges, known as *spectral flow*, a representation with any periodicity can be converted to any other periodicity. The periodicities of the $U(1)$ current and energy-momentum tensor are unaffected. In the $d = 4$ heterotic string, the flow with $\eta = \frac{1}{2}$ converts a chiral primary into a $\tilde{Q} = -\frac{1}{2}$ R sector state, the flow with $\eta = -\frac{1}{2}$ converts an antichiral primary into a $\tilde{Q} = \frac{1}{2}$ R sector state, and vice versa: the superpartners are related to one another by spectral flow. The defining relations for chiral and antichiral primaries become

$$\tilde{G}^\pm_n|\psi\rangle = \tilde{L}_n|\psi\rangle = \tilde{J}_n|\psi\rangle = 0 \,, \quad n \geq 0 \,, \tag{19.1.18}$$

where $|\psi\rangle$ is the R sector state produced by the flow.

19.2 Type II strings on Calabi–Yau manifolds

Consider either type II string on a Calabi–Yau manifold. The compact CFT is the same as for the heterotic string, with the left-moving current algebra fermions λ^A for $A = 1, \ldots, 6$ replaced by fermions ψ^m and the remaining λ^A omitted. One can construct a right-moving spacetime supersymmetry precisely as in the heterotic string, and because the world-sheet

theory is now the same on the right and left, there is a second spacetime supercharge from the left-movers. Thus either type II theory will have $d = 4$, $N = 2$ supersymmetry. The argument from section 18.5 shows further that this will be true for any compact CFT with (2,2) superconformal symmetry, provided it satisfies the generalized GSO projection (19.1.6) on both sides.

For the IIA string on a Calabi–Yau manifold, the massless fields come from the NS–NS fluctuations g_{MN}, b_{MN}, ϕ and the R–R fluctuations c_M and c_{MNP}. For any Calabi–Yau manifold these will include the four-dimensional metric $g_{\mu\nu}$, dilaton ϕ, and axion $b_{\mu\nu} \cong a$. The field c_μ is a massless vector. In addition, every Calabi–Yau manifold has exactly one (3,0)-form and one harmonic (0,3)-form, giving additional scalars from c_{ijk} and $c_{\bar{i}\bar{j}\bar{k}}$. For each harmonic (1,1)-form there is a scalar from $g_{i\bar{j}}$, another scalar from $b_{i\bar{j}}$, and a vector from $c_{\mu i\bar{j}}$. For each harmonic (2,1)-form there are scalars from g_{ij} and $g_{\bar{i}\bar{j}}$ just as for the heterotic string, and also scalars from $c_{ij\bar{k}}$ and $c_{\bar{i}jk}$.

Let us see how these fit into multiplets of the $N = 2$ spacetime supersymmetry; the latter are summarized in section B.2. The metric $g_{\mu\nu}$ plus vector c_μ comprise the bosonic content of the supergravity multiplet. The remaining model-independent fields are four real scalars: ϕ, a, c_{ijk}, and $c_{\bar{i}\bar{j}\bar{k}}$. This is the bosonic content of one hypermultiplet. For each harmonic (1,1)-form there are two scalars and a vector, the bosonic content of a vector multiplet. For each harmonic (2,1)-form there are four scalars again forming a hypermultiplet. In all, there are

$$\text{IIA:}\quad h^{1,1} \text{ vector multiplets}\,, \quad h^{2,1} + 1 \text{ hypermultiplets}\,. \tag{19.2.1}$$

For the IIB string on a Calabi–Yau manifold, the massless fields come from the NS–NS fluctuations g_{MN}, b_{MN}, ϕ and the R–R fluctuations c, c_{MN}, and c_{MNPQ}. The model-independent fields are now the four-dimensional metric $g_{\mu\nu}$, dilaton ϕ, and axion $b_{\mu\nu} \cong a$, and also the scalar c, a second axion $c_{\mu\nu} \cong a'$, and a vector $c_{\mu ijk}$ from the (3,0)-form. For each harmonic (1,1)-form there is again a scalar from $g_{i\bar{j}}$ and one from $b_{i\bar{j}}$, and also one from $c_{i\bar{j}}$ and a fourth from the Poincaré dual of $c_{\mu\nu i\bar{j}}$. One might think that we should get additional scalars from $c_{ij\bar{k}\bar{l}}$ with the $h^{1,1}$ harmonic (2,2)-forms implied by the Hodge diamond (17.2.29), but because the 5-form field strength is self-dual these are actually identical to the states from $c_{\mu\nu i\bar{j}}$. For the same reason there is not an additional vector from $c_{\mu i\bar{j}\bar{k}}$. For each harmonic (2,1)-form there are scalars from g_{ij} and $g_{\bar{i}\bar{j}}$ and a vector from $c_{\mu ij\bar{k}}$. Again the self-duality means that the vectors $c_{\mu\bar{i}\bar{j}k}$ give the same vector states. The massless IIB states form the $N = 2$ supergravity multiplet plus

$$\text{IIB:}\quad h^{2,1} \text{ vector multiplets}\,, \quad h^{1,1} + 1 \text{ hypermultiplets}\,. \tag{19.2.2}$$

Table 19.1. *Relations between Calabi–Yau moduli and supersymmetry multiplets in the two type II theories.*

	IIA	IIB
Kähler (1,1):	vector	hyper
complex structure (2,1):	hyper	vector

For convenient reference we have summarized the Calabi–Yau moduli of the type II theories in table 19.1.

Low energy actions

In section B.7 we describe the general low energy theory allowed by $N = 2$ supergravity. An important result is that the potential is determined entirely by the gauge interactions. Since the gauge fields in the type II compactifications all come from the R–R sector, all strings states are neutral and so the potential vanishes. Thus we can conclude that all the scalars found above are moduli. Moreover, because this is a consequence of symmetry it remains true to all orders in string and world-sheet perturbation theory, and even nonperturbatively. This is different from the $N = 1$ case, where we saw that nonperturbative effects could produce a potential.

The low energy action is then determined by supersymmetry in terms of the kinetic terms for the moduli — the metric on moduli space. Supersymmetry further implies that the kinetic terms for the hypermultiplet scalars are independent of the vector multiplet scalars and the kinetic terms for the vectors and their scalar partners are independent of the hypermultiplet scalars. In other words, the moduli space is a product. The vector multiplet moduli space is a special Kähler manifold and the hypermultiplet moduli space a quaternionic manifold, both defined in section B.7.

Now let us compare the IIA and IIB theories compactified on the same Calabi–Yau manifold. A hypermultiplet has twice as many scalars as a vector multiplet, so the IIA and IIB moduli spaces (19.2.1) and (19.2.2) do not in general even have the same dimension. However, they are related in interesting ways. If the R–R scalars are set to zero the tree-level IIA and IIB theories become identical, and indeed this removes two states from each hypermultiplet. Thus at string tree level, the R–R-vanishing subspace of each hypermultiplet moduli space should be a product of the dilaton–axion moduli space and a space identical to the vector multiplet moduli space of the *other* type II theory on the same Calabi–Yau manifold.

We can also go the other way, constructing the larger hypermultiplet

moduli space from the smaller vector multiplet moduli space. Imagine compactifying one additional coordinate x^3 on a circle, going to $d = 3$. On a circle the IIA and IIB theories are T-dual, so the resulting moduli spaces should be identical. Indeed, each vector gives rise to two additional moduli, one from the vector component A_3 and one from the Poincaré dual of the $d = 3$ gauge field, so the dimensions are correct. Carrying out this reduction in detail gives the *c-map* from special Kähler manifolds to quaternionic manifolds. Since the hypermultiplet moduli spaces can be deduced in this way from the vector multiplet spaces, it follows that each can be characterized by a single holomorphic prepotential as in special Kähler geometry.

For the heterotic string we found a nonrenormalization theorem for the superpotential in world-sheet perturbation theory from the combination of holomorphicity and the symmetry $\delta T = i\epsilon$. It is interesting to apply these same constraints in the present case. Consider first a single Kähler modulus T representing the overall scale of the Calabi–Yau manifold. Just as for the heterotic string, eq. (17.5.4), one derives the Kähler potential

$$K = -3\ln(T + T^*) . \tag{19.2.3}$$

Up to a Kähler transformation, this is of the special geometry form (B.7.18),

$$K = -\ln \mathrm{Im}\left(\sum_I X^{I*}\partial_I F(X)\right) , \tag{19.2.4}$$

where

$$F(X) = \frac{(X^1)^3}{X^0} , \quad T = \frac{iX^1}{X^0} . \tag{19.2.5}$$

The PQ symmetry $\delta T = i\epsilon$ is

$$\delta X^1 = \epsilon X^0 . \tag{19.2.6}$$

The function F is not invariant under this but changes by

$$\delta F = 3\epsilon (X^1)^2 . \tag{19.2.7}$$

The Kähler potential is then invariant; more generally, it is invariant provided that

$$\delta F = c_{IJ} X^I X^J \tag{19.2.8}$$

with real coefficients.

The function F must be of degree 2 in the X^I, and so an n-loop world-sheet correction would scale as $T^{3-n}(X^0)^2$. The only such correction that is allowed by the PQ symmetry and is not of the trivial form (19.2.8) is

$$\Delta F = i\lambda (X^0)^2 , \tag{19.2.9}$$

a three-loop correction to the leading interaction. This does in general appear, as we will note later. Further, in parallel to the heterotic string, nonperturbative world-sheet corrections to the Kähler moduli space are allowed by this argument but corrections to the complex structure moduli space are forbidden because the Kähler modulus T cannot couple to the complex structure moduli.

For more than one hypermultiplet, the PQ symmetries $\delta T^A = i\epsilon^A$ again greatly constrain the function F. It can be shown that any symmetry of the Kähler metric must be of the form

$$\delta X^I = \omega^{IJ} X^J , \tag{19.2.10}$$

so that up to a field redefinition we must have

$$T^A = \frac{iX^A}{X^0} , \quad A = 1,\ldots,n , \tag{19.2.11}$$

and $\omega^{A0} = \epsilon^A$. Requiring that F transform as in eq. (19.2.8) determines that it is of the form

$$F = \frac{d_{ABC} X^A X^B X^C}{X^0} + i\lambda (X^0)^2 . \tag{19.2.12}$$

This is consistent with the explicit results in section 17.5 for Calabi–Yau compactification, which were stated for the heterotic string but also apply to the type II theories. The coefficients d_{ABC} are the intersection numbers discussed there. This is the moduli space of vector multiplets in the IIA string, or the R–R-vanishing subspace of the IIB hypermultiplet moduli space. Since it is derived using the (1,1) PQ symmetry, this F receives world-sheet instanton corrections of order $\exp(-n_A T^A/2\pi\alpha')$.

The complex structure moduli space must be a special Kähler manifold but is otherwise not restricted to a form as narrow as eq. (19.2.12). The one strong constraint is that the field-theory calculation of this moduli space receives no corrections from world-sheet interactions. The scale of the Calabi–Yau space, which governs these interactions, is a Kähler modulus. By the factorized property of the moduli space, it cannot appear in the complex structure metric. In section 19.6 we will describe the field theory calculation further.

The discussion of the moduli space metric thus far has been restricted to string tree level. For the potential, the $N = 2$ spacetime supersymmetry allowed us to draw strong conclusions that were valid even nonperturbatively. This is also the case for the metric: supersymmetry strongly constrains the form of possible string corrections, in the expansion parameter $g \sim e^{\Phi_4}$, as well as world-sheet corrections, in the expansion parameter $\alpha'/R_c^2 \sim 1/T$. The string coupling is governed by the dilaton, so any perturbative and nonperturbative corrections to the metric must depend

on the dilaton. For both IIA and IIB compactifications, we have argued above that the dilaton is in a hypermultiplet. We arrived at this conclusion by counting states, but one can also show it directly (exercise 19.1). The low energy action for the vector multiplet cannot depend on the dilaton because of the product structure, and so receives no corrections from string interactions, either perturbative or nonperturbative.

Referring to table 19.1, one can conclude from the nonrenormalization theorems that of the four moduli spaces appearing in the type II theories, the IIB complex structure moduli space receives neither world-sheet nor string corrections. The tree-level result one obtains in the field theory approximation is exact. The other moduli spaces receive corrections of one or both kinds. Later we will see various extensions and applications of these results.

Chiral rings

As a final point, let us consider compactification on a general (2,2) SCFT. In parallel to the discussion for the heterotic string, the vertex operators for the NS–NS moduli must be of one of the forms

$$|c,\tilde{c}\rangle\ ,\quad |c,\tilde{a}\rangle\ ,\quad |a,\tilde{c}\rangle\ ,\quad |a,\tilde{a}\rangle\ , \tag{19.2.13}$$

the states being chiral or antichiral primaries on each side. The corresponding operators are respectively denoted

$$\Phi^{++}\ ,\quad \Phi^{+-}\ ,\quad \Phi^{-+}\ ,\quad \Phi^{--}\ . \tag{19.2.14}$$

Now consider a product of operators of the same type, for example chiral–chiral operators Φ^{++} and Ψ^{++}. The minimum weight for an operator in the OPE is

$$h \geq \frac{1}{2}(Q_\Phi + Q_\Psi) = h_\Phi + h_\Psi\ , \tag{19.2.15}$$

and similarly for $\tilde{h}$ and $\tilde{Q}$. The OPE is therefore nonsingular,

$$\Phi^{++}(z,\bar{z})\Psi^{++}(0,0) \sim (\Phi\Psi)^{++}(0,0)\ . \tag{19.2.16}$$

The operator $(\Phi\Psi)^{++}$ has $(h,\tilde{h}) = \frac{1}{2}(Q,\tilde{Q})$ and so is again chiral–chiral. The $(c,\tilde{c})$ operators thus form a multiplicative *chiral ring* (not a *group*, because an operator with $Q > 0$ has no inverse). The $(a,\tilde{a})$ operators form the conjugate ring, and the $(c,\tilde{a})$ and $(a,\tilde{c})$ operators form a different ring and its conjugate.

Let us connect this with the Calabi–Yau example. The (2,2) $U(1)$ currents are

$$j = \psi^i\psi^{\bar{\imath}}\ ,\quad \tilde{\jmath} = \tilde{\psi}^i\tilde{\psi}^{\bar{\imath}}\ . \tag{19.2.17}$$

From any harmonic (p,q)-form we can construct the operator

$$b_{i_1\ldots i_p\bar{\jmath}_1\ldots\bar{\jmath}_q}(X)\psi^{i_1}\ldots\psi^{i_p}\tilde{\psi}^{\bar{\jmath}_1}\ldots\tilde{\psi}^{\bar{\jmath}_q}\ . \tag{19.2.18}$$

This has charges and weights

$$Q = p\ ,\quad h = \frac{p}{2}\ ,\quad \tilde{Q} = -q\ ,\quad \tilde{h} = \frac{q}{2}\ , \tag{19.2.19}$$

and so is a $(c,\tilde{a})$ chiral primary.[1] The weight comes entirely from the Fermi fields, because the form is harmonic. Naively multiplying two operators (19.2.18), the chiral ring is just the wedge product of the forms, which is the *cohomology ring*. This is correct at large radius, where the world-sheet interactions are weak, but the ring is corrected by world-sheet interactions. Note that the operator corresponding to a $(1,1)$-form is just the vertex operator for the Kähler modulus. The product of three such operators is proportional to the corresponding Yukawa coupling.

Topological string theory

Notice that

$$(G_0^+)^2 = 0\ . \tag{19.2.20}$$

This suggests that we think of G_0^+ as a BRST operator. The reader can show that the cohomology consists precisely of the chiral primary states, in the form (19.1.18), with vanishing spacetime momentum.

The operator G_0^+ is not conformally invariant, because the current T_F^+ has weight $(\frac{3}{2},0)$. Let us consider instead the energy-momentum tensor

$$T_B^{\rm top} \equiv T_B + \frac{1}{2}\partial j\ . \tag{19.2.21}$$

The reader can verify the following properties:

$$T_B^{\rm top}(z)T_B^{\rm top}(0) \sim \frac{2}{z^2}T_B^{\rm top}(z) + \frac{1}{z}\partial T_B^{\rm top}(z)\ , \tag{19.2.22a}$$

$$T_B^{\rm top}(z)T_F^+(0) \sim \frac{1}{z^2}T_F^+ + \frac{1}{z}\partial T_F^+\ . \tag{19.2.22b}$$

This shows that $T_B^{\rm top}$ generates a conformal symmetry of central charge 0, and that under this symmetry T_F^+ has weight $(1,0)$ and so G_0^+ is conformally invariant.

Starting with any (2,2) CFT, we can make a string theory by coupling the world-sheet metric to $T_B^{\rm top}$. Because the central charge already vanishes, no additional ghosts are needed; the OPE

$$T_F^+(z)T_F^-(0) = \ldots + \frac{1}{z}T_B^{\rm top}(0) + \ldots \tag{19.2.23}$$

[1] Often the sign convention for $\tilde{Q}$ is reversed so that the operator is $(c,\tilde{c})$.

shows that T_F^- plays the role of the b ghost. This theory has very few states because the cohomology is so small, and in particular has no dynamics because all physical states are time-independent. It is known as *topological string theory*, and its amplitudes are a special subset of the amplitudes of the related type II string theory.

19.3 Heterotic string theories with (2,2) SCFT

Now let us consider heterotic string theory with a general $c = \tilde{c} = 9$ (2,2) CFT. The remainder of the left-moving central charge for the compact theory comes from 26 free current algebra fermions. The noncompact fields are the usual X^μ and $\tilde{\psi}^\mu$. We continue to take the generalized GSO projection (19.1.6) on the right-moving side. On the left-moving side we will similarly generalize the GSO projection. We focus on the $E_8 \times E_8$ case. The current algebra fermions of interest will always be λ^A with $7 \leq A \leq 16$. For the second E_8, where $17 \leq A \leq 32$, we take the same GSO projection as in ten dimensions. The current algebra GSO projection then requires that the sum of the charge Q from the left-moving $N = 2$ SCFT and the charge for the current algebra number current

$$\sum_{K=4}^{8} \lambda^{K+}\lambda^{K-} \tag{19.3.1}$$

be an even integer.

From the current algebra fields and the free boson for the $U(1)$ of the left-moving superconformal algebra one can form the following (1,0) currents, all of which survive the GSO projection:

$$\lambda^A\lambda^B\,, \quad \Theta_{\mathbf{16}} \exp(3^{1/2}iH/2)\,, \quad \Theta_{\overline{\mathbf{16}}} \exp(-3^{1/2}iH/2)\,, \quad i\partial H\,. \tag{19.3.2}$$

Here $\Theta_{\mathbf{16}}$ and $\Theta_{\overline{\mathbf{16}}}$ are the R sector vertex operators for the current algebra fermions, with the subscript distinguishing the two spinor representations. These currents transform as

$$\mathbf{45} + \mathbf{16} + \overline{\mathbf{16}} + \mathbf{1} \tag{19.3.3}$$

under the manifest $SO(10)$ current algebra. The gauge group must have an $SO(10)$ subgroup under which the adjoint representation decomposes in this way; this identifies it as E_6, whose adjoint is the **78**. In addition there is another E_8 from the second set of current algebra fermions. This $E_6 \times E_8$ is the full gauge symmetry of generic (2,2) compactifications. In special cases there are additional gauge symmetries, such as the $SU(3)$ of the $\mathbf{Z}_3$ orbifold.

To find the scalar spectrum, we start with the operator Φ^{++} for a state $|c,\tilde{c}\rangle$ with $Q = \tilde{Q} = 1$. On the right-moving side this is in the $-\frac{1}{2}$ picture, but on the left the superconformal symmetry is just a global symmetry and there are no pictures. Rather, we need total weight $h = 1$. We can obtain this and also satisfy the GSO projection with an additional λ^A excitation; the vertex operator is

$$\mathscr{V} = \lambda^A\Phi^{++}\,. \tag{19.3.4}$$

This is a **10** of the $SO(10)$ that acts on λ^A. By spectral flow on the left-moving part of Φ^{++} we also obtain

$$\Theta_{\mathbf{16}}\Phi^{++}(1 \to -\tfrac{1}{2})\,. \tag{19.3.5}$$

The notation indicates the charge Q after spectral flow. The charge is shifted by $-\frac{3}{2}$ units, which moves Φ^{++} from the NS to the R sector of the (2,2) CFT. The effect of spectral flow is to give $\Phi^{++}(Q \to Q')$ a weight

$$h = \frac{Q}{2} + \frac{Q'^2 - Q^2}{6}\,, \quad \tilde{h} = \frac{\tilde{Q}}{2}\,. \tag{19.3.6}$$

This is $(\frac{3}{8}, \frac{1}{2})$ in the present case. We have also included an R sector vertex operator $\Theta_{\mathbf{16}}$ for the current algebra fermions, the subscript indicating its representation. The vertex operator then has the correct weight $(1, \frac{1}{2})$ and satisfies the GSO projection. Spectral flow also gives

$$\Phi^{++}(1 \to -2)\,. \tag{19.3.7}$$

This is now in the NS sector, with weight $(1, \frac{1}{2})$, and satisfies the GSO projection. These $SO(10)$ representations $\mathbf{10} + \mathbf{16} + \mathbf{1}$ add up to a **27** of E_6. As discussed in section 19.1, spectral flow on the right-moving side generates the fermionic partners of these scalars in the **2** of the four-dimensional Lorentz group.

There is one more massless scalar related to the above, with the weight $(1, \frac{1}{2})$ vertex operator

$$G^-_{-1/2} \cdot \Phi^{++}\,. \tag{19.3.8}$$

This is neutral under the gauge group. To see the significance of this state, consider using the same (2,2) CFT for compactification of one of the *type II* strings. In this case, Φ^{++} is the $(-1,-1)$ picture vertex operator for a modulus. The operator (19.3.8), which is in the heterotic -1 picture, is then identical to the zero-momentum vertex operator for the type II modulus in the $(0,-1)$ picture. Raising the right-moving picture in both theories, the 0 picture heterotic vertex operator is identical to the $(0,0)$ picture type II vertex operator. These are the pictures that we add to the action when we turn on a background, so we conclude that we get the same CFT in the heterotic theory with a background of the scalar (19.3.8) as in the type II theories with a nonzero $(c,\tilde{c})$ modulus.

This further implies that the massless state (19.3.8) is a modulus, with vanishing potential. The argument is that we know from $N = 2$ spacetime supersymmetry that the corresponding type II state has no potential, and so the world-sheet theory with this background is an exact CFT whichever string theory we have. This kind of argument, using the larger supersymmetry of the type II theory to make arguments indirectly about the heterotic compactification, is very effective. It is important to note that it is valid only at string tree level: we have used the statement that CFTs correspond to tree-level backgrounds. At higher orders there is no relation between the two theories, because different states run around the loops. Some quantities that are not renormalized in the type II theory do get corrections in the less supersymmetric heterotic theory. For example, we argued that for the type II string, $N = 2$ spacetime supersymmetry implies that the flat directions are flat even nonperturbatively. In the heterotic string we know that gluino condensation and other effects can produce a potential.

Starting with a state $|a, \tilde{a}\rangle$ leads to the antiparticles of the above states. Starting with states $|c, \tilde{a}\rangle$ and $|a, \tilde{c}\rangle$ leads to a modulus plus a generation of the opposite chirality, the spacetime **2** being correlated with the gauge $\overline{\mathbf{27}}$. This pairing between generations and moduli of one type, and anti-generations and moduli of another type, generalizes the association with (1,1) and (2,1) forms found in Calabi–Yau compactification. In chapter 17 we argued that the moduli were exact by appealing to a result on the detailed form of instanton amplitudes, and now we have come to the same conclusion by appealing to results on the general $N = 2$ spacetime supersymmetric action. This second method is more general. For example, it also implies that the blowing-up modes for the fixed points of orbifolds are moduli, a result argued for in section 16.4 by citing detailed studies of twisted-state amplitudes.

In Calabi–Yau compactification we found additional E_6 singlets. In the abstract (2,2) description, these are states of weight $(1, \frac{1}{2})$ and $Q = 0$ that are $N = 2$ superconformal primary fields on the left-moving side. This is in contrast to the states (19.3.8), which are not annihilated by $G^+_{1/2}$.

We have used the relation between heterotic and type II compactifications at string tree level, but let us note that any modular-invariant type II compactification also gives rise to a modular-invariant heterotic compactification. The modular transformation of the type II string theory mixes up the four sectors on each side, R vs NS and $\exp(\pi i F) = \pm 1$, in the same fashion as in the ten-dimensional theory in chapter 10. To make a heterotic theory we replace the two left-moving fermions $\psi^{2,3}$ with 26 left-moving current algebra fermions. The effect is independent of the (2,2) CFT and in particular is the same as in ten dimensions. Because the difference in the number of fermions is an odd multiple of eight, the signs in the type II and heterotic modular invariants differ (compare eq. (10.7.9) with eq. (11.2.13)), which is precisely as required by spacetime spin-statistics.

More on the low energy action

The argument that the scalars (19.3.8) are moduli is not self-contained, in that it uses results on $N = 2$ supergravity that we have not derived. To show these requires detailed analysis of the field theory actions and is beyond the scope of this book. One can also give a direct demonstration that the scalar (19.3.8) is a modulus, by extracting the effective action from an analysis of the heterotic string scattering amplitudes. The basic strategy is to consider a tree-level amplitude with any number of moduli (19.3.8), in any combination of the chiral and antichiral types. If the potential vanishes then this amplitude vanishes in the zero-momentum limit. Writing the operator $G^-_{-1/2}$ as a contour integral of T_F around Φ^{++}, one can deform the contour until it surrounds other vertex operators. It then takes one of the two forms

$$G^-_{-1/2} G^-_{-1/2} \cdot \Phi^{+\pm} = 0 \,, \tag{19.3.9a}$$

$$G^-_{-1/2} G^+_{-1/2} \cdot \Phi^{-\pm} = (2L_{-1} - G^+_{-1/2} G^-_{-1/2}) \cdot \Phi^{-\pm} = 2\partial \Phi^{-\pm} \,. \tag{19.3.9b}$$

We have used the relations

$$(G^-_{-1/2})^2 = 0 \,, \quad G^-_{-1/2} \cdot \Phi^{-\pm} = 0 \,. \tag{19.3.10}$$

The final result is a total derivative and so should integrate to zero. To complete the argument one needs to show that there are no surface terms from vertex operators approaching one another; this uses the fact that the same structure appears on the right-moving side as on the left-moving one. Also, the fixed vertex operators require some additional bookkeeping. These details are left to the references. Below we cite further results that are found from a careful study of string amplitudes. These are obtained by the same approach, but the details are lengthy and again are left to the references.

In the previous section we discussed the constraints from $N = 2$ supergravity on the metrics for the type II moduli spaces, that is, on the kinetic terms for the moduli. We have argued that the CFT is the same for the type II and heterotic theories, and so the metric on moduli space should be the same in both string theories. In particular, the Zamolodchikov metric (18.6.10) gives the moduli space metric in terms of data from the CFT. This conclusion is confirmed by a study of moduli scattering amplitudes, which to order k^2 are the same in the type II and heterotic theories. Thus the (1,1) and (2,1) moduli spaces for the heterotic string each are

special Kähler manifolds and are governed by a single holomorphic prepotential. This is in agreement with the explicit Calabi–Yau results in section 17.5. The analytic function $F_1(T)$ governing the Kähler moduli was there denoted $W(T)$, and the analytic function $F_2(Z)$ governing the complex structure moduli was there denoted $\mathcal{G}(Z)$.

For the **27**s and $\overline{\mathbf{27}}$s, the model-dependent factors $\Phi^{\pm\pm'}$ in the vertex operators are the same as for the corresponding moduli. One would therefore expect that their amplitudes would be related to the amplitudes for the moduli in a model-independent way. Indeed, the low energy action is completely determined in terms of the holomorphic prepotentials F_1 and F_2 governing the (1,1) and (2,1) moduli, except for the extra E_6 singlets. The $\overline{\mathbf{27}}$ metric and superpotential are

$$G'_{A\bar{B}} = \exp[\kappa^2(K_2 - K_1)/3]G_{A\bar{B}} \,, \tag{19.3.11a}$$

$$W(\phi) = \phi^A{}_{\bar{x}}\phi^B{}_{\bar{y}}\phi^C{}_{\bar{z}}d^{\bar{x}\bar{y}\bar{z}}\partial_A\partial_B\partial_C F_1(T) \,. \tag{19.3.11b}$$

The **27** metric and superpotential are

$$G'_{a\bar{b}} = \exp[\kappa^2(K_1 - K_2)/3]G_{a\bar{b}} \,, \tag{19.3.12a}$$

$$W(\chi) = \chi^a{}_x\chi^b{}_y\chi^c{}_z d^{xyz}\partial_a\partial_b\partial_c F_2(Z) \,. \tag{19.3.12b}$$

Unlike earlier results, these cannot be derived from $N = 2$ supergravity, as the **27**s and $\overline{\mathbf{27}}$s have no analogs in the type II theory. That the relations (19.3.11) and (19.3.12) are identical in form follows from the fact that the (1,1) and (2,1) states are essentially identical in CFT, differing only by a change in sign of the free scalar H from the superconformal algebra. The four-loop term (19.2.9) in F_1 does not affect $W(\phi)$.

These results generalize the Calabi–Yau results in section 17.5. We have also learned from the use of the PQ symmetries that the Kähler prepotential F_1 is of the form eq. (19.2.12) in world-sheet perturbation theory, and that F_2 cannot receive world-sheet corrections. Again we emphasize that the forms (19.3.11) and (19.3.12) are derived using CFT arguments and so are exact at string tree level, but that the relation between the different terms in the low energy action and the special form of the Kähler potential are not protected by the $N = 1$ supersymmetry of the heterotic string and so do not survive string loop corrections.

19.4 $N = 2$ minimal models

In chapter 15 we described the $N = 0$ and $N = 1$ minimal models. There is a similar family of solvable CFTs with $N = 2$ superconformal symmetry. It is interesting to consider heterotic string theories where the (2,2) CFT is a combination of these $N = 2$ minimal models, with total central charge

$(c, \tilde{c}) = 9$. This is another subject for which our treatment must be rather abbreviated. The full details of the constructions are lengthy and are left to the references.

A generalization of the method described in section 15.1 shows that unitary representations of the $N = 2$ superconformal algebra can exist only if $c \geq 3$, or at the discrete values

$$c = 3 - \frac{6}{k+2} = \frac{3k}{k+2} \,, \quad k = 0, 1, \ldots \,. \tag{19.4.1}$$

For the discrete theories, the allowed weights and $U(1)$ charges are

$$\text{NS:} \quad h = \frac{l(l+2) - q^2}{4(k+2)} \,, \quad Q = \frac{q}{k+2} \,, \tag{19.4.2a}$$

$$\text{R:} \quad h = \frac{l(l+2) - (q \pm 1)^2}{4(k+2)} + \frac{1}{8} \,, \quad Q = \frac{q \pm 1}{k+2} \mp \frac{1}{2} \,, \tag{19.4.2b}$$

where $0 \leq l \leq k$ and $-l \leq q \leq l$.

We showed that the $N = 0$ minimal models could be constructed as cosets starting from $SU(2)$ current algebras. There is a similar relation here. The central charge (19.4.1) is precisely the central charge of the $SU(2)$ current algebra at level k. The connection is as follows. Recall from section 15.5 that we can represent one current, say j^3, in terms of a free boson $i(k/2)^{1/2}\partial H$, and the CFT then separates into the free boson CFT and a so-called parafermionic theory. All other operators separate, for example

$$j^+ = \psi_1 \exp\left[i\left(\frac{2}{k}\right)^{1/2} H\right] \,, \quad j^- = \psi_1^\dagger \exp\left[-i\left(\frac{2}{k}\right)^{1/2} H\right] \,. \tag{19.4.3}$$

Now define

$$T_F^+ = \psi_1 \exp\left[i\left(\frac{k+2}{k}\right)^{1/2} H\right] \,, \quad T_F^- = \psi_1^\dagger \exp\left[-i\left(\frac{k+2}{k}\right)^{1/2} H\right] \,. \tag{19.4.4}$$

These operators have conformal weight

$$1 - \frac{1}{2}\left(\frac{2}{k}\right) + \frac{1}{2}\left(\frac{k+2}{k}\right) = \frac{3}{2} \,, \tag{19.4.5}$$

and one can show that they satisfy the $N = 2$ superconformal OPE. The parafermionic plus free-boson central charge remains at its original value.

Similarly, the current algebra primary fields factorize

$$\mathcal{O}^j_m = \psi^j_m \exp\left[im\left(\frac{2}{k}\right)^{1/2} H\right] \,. \tag{19.4.6}$$

Define now

$$\mathcal{O}'^j_m = \psi^j_m \exp\left[i\frac{2m}{k^{1/2}(k+2)^{1/2}} H\right] \,. \tag{19.4.7}$$

Relative to the current algebra primary, the exponent in $\mathcal{O}'^{j}_{m}$ is multiplied by $[2/(k+2)]^{1/2}$. The exponent in T_F is multiplied by the reciprocal factor relative to j^3, so the leading singularity z^{-1} in the current–primary OPE remains the same, and the operators $\mathcal{O}'^{j}_{m}$ are NS primaries under the $N=2$ algebra. Subtracting and adding the free-boson contributions, the weight of $\mathcal{O}'^{j}_{m}$ is

$$h = \frac{j(j+1)}{k+2} - \frac{m^2}{k} + \frac{2m^2}{k(k+2)} = \frac{j(j+1)-m^2}{k+2} . \tag{19.4.8}$$

This matches the weight of the NS primary (19.4.2a), with the identification $l = 2j$ and $q = 2m$. The ranges of l and q then match the ranges of the current algebra primaries. With the properly normalized $N=2$ current

$$j = i[k/(k+2)]^{1/2}\partial H , \tag{19.4.9}$$

the charge $Q = 2m/(k+2)$ also matches that of the current algebra primary. Similarly, the fields

$$\psi^{j}_{m} \exp\left[i\frac{2m \pm k/2}{k^{1/2}(k+2)^{1/2}}H\right] \tag{19.4.10}$$

have an additional factor $z^{\pm 1/2}$ in their OPEs with the currents. They are therefore primary fields in the R sector and are also annihilated by $G_0^{\pm}$, the sign correlating with that in the exponential. The weight and $U(1)$ charge agree with eq. (19.4.2b).

Landau–Ginzburg models

We now give a Lagrangian representation of the minimal models, the *Landau–Ginzburg* description. The rigid subgroup of the (2,2) superconformal algebra is (2,2) world-sheet supersymmetry. Having four supercharges, this is the dimensional reduction of $d=4$, $N=1$ supersymmetry. Any $d=4$, $N=1$ theory becomes a (2,2) world-sheet theory by dimensional reduction, requiring the fields to be independent of $x^{2,3}$.

In particular, let us take a single chiral superfield with superpotential

$$W(\Phi) = \Phi^{k+2} . \tag{19.4.11}$$

Consider a scale transformation

$$\sigma \to \lambda\sigma , \quad \phi = \lambda^{\omega}\phi , \tag{19.4.12a}$$

$$\psi \to \lambda^{\omega-1/2}\psi , \quad F = \lambda^{\omega-1}F , \tag{19.4.12b}$$

with ω as yet unspecified. The relation between the scaling of the various components of the superfield is determined by the fact that the supersymmetry transformation squares to a translation. Including the scaling

of $d^2\sigma$, the terms in the action (B.2.16) that are linear in W scale as

$$\lambda^{2-1+(k+2)\omega} , \tag{19.4.13}$$

and so are invariant if $\omega = -1/(k+2)$. With this value for ω, the kinetic terms scale as $\lambda^{-2/(k+2)}$ and are less important at long distance (large λ). Thus the theory at long distance is scale-invariant, and so also conformally invariant by the discussion of the c-theorem. Normally one must worry about quantum corrections to scaling, but not here because the superpotential is not renormalized. In this case the nonrenormalization theorem can be understood from symmetry. The theory with superpotential (19.4.11) has an R symmetry (defined in eq. (B.2.21)) under which ϕ has charge $2/(k+2)$. This allows no corrections to the superpotential. If we began with a superpotential which also had higher powers of Φ, their effect would scale away at long distance.

The combination of conformal invariance and rigid supersymmetry generates the full (2,2) superconformal theory. Thus, the long distance limit of the theory has this symmetry, and it is this limiting *critical* theory that can be used as a string compactification. Equivalently, but more in the language of renormalization, we can hold the distance fixed but take to zero the 'cutoff' length at which the original field theory is defined.

We expect the critical theory to be a minimal model. The chiral superfield without a superpotential is the usual $c=3$ free field representation. As in the discussion of $N=0$ Landau–Ginzburg theories in chapter 15, the superpotential should reduce the effective number of degrees of freedom and so reduce the central charge. To see which minimal model we have, let us note that the field ϕ is a $(c,\tilde{c})$ primary. Its supersymmetry transformation (B.2.14) contains a projection operator P_+ onto four-dimensional spinors with $s_0 + s_1 = \pm 1$. The value of s_0 determines which of $P^0 \pm P^1$ the supersymmetry squares to, and so whether it is left- or right-moving. The projection P_+ thus implies that one rigid supersymmetry on each side annihilates ϕ; by convention we call these $G^-_{-1/2}$ and $\tilde{G}^-_{-1/2}$, so ϕ is $(c,\tilde{c})$. The chiral–chiral property is also consistent with the weight and charge. The scale transformation (19.4.12) implies that $h_\phi + \tilde{h}_\phi = -\omega$, and ϕ is spinless so

$$h_\phi = \tilde{h}_\phi = \frac{1}{2(k+2)} . \tag{19.4.14}$$

The R symmetry, under which ϕ has charge $2/(k+2)$, acts on all components of the supercharge and so is equal to $Q + \tilde{Q}$. Thus

$$Q_\phi = \tilde{Q}_\phi = \frac{1}{k+2} , \tag{19.4.15}$$

and ϕ satisfies $Q = 2h$ and $\tilde{Q} = 2\tilde{h}$.

We can now identify the Landau–Ginzburg theory (19.4.11) with the minimal model at the same k. The minimal model field $\mathcal{O}^{l\,1/2}_{1/2}$ is a chiral primary, as one sees from the relation $Q = 2h$, and its weight agrees with that of ϕ at the same k. Also, by the chiral ring argument, we can make further primaries as powers ϕ^l. These correspond to $\mathcal{O}^{l\,l/2}_{l/2}$. However, the process terminates, because the equation of motion

$$\partial_\phi W(\phi) = (k+2)\phi^{k+1} = 0 \tag{19.4.16}$$

implies that $l \leq k$. This matches the minimal model bound on l, as well as the general bound that the maximum charge of a chiral primary is

$$Q = \frac{c}{3} = \frac{k}{k+2} \,. \tag{19.4.17}$$

An important role is played by the $\mathbf{Z}_{k+2}$ symmetry of the Landau–Ginzburg theory,

$$\Phi \to \exp\left(\frac{2\pi i}{k+2}\right)\Phi \,. \tag{19.4.18}$$

This acts in the same way on all components of the superfield and leaves the superpotential invariant. It is a discrete subgroup of the superconformal $U(1)$ generated by

$$\exp(2\pi i Q) \;; \tag{19.4.19}$$

this operator acts on ϕ as in (19.4.18), and it commutes with $T_F^{\pm}$ and so acts in the same way on all components of a world-sheet superfield. The operator $\exp(2\pi i \tilde{Q})$ is not an independent symmetry, because all fields in the Landau–Ginzburg theory are invariant under $\exp[2\pi i(Q - \tilde{Q})]$.

The Landau–Ginzburg theory is strongly interacting at long distance (since the interaction dominates the kinetic term) and so cannot be solved explicitly. Nevertheless, as in the examples we have seen, most of the quantities of interest in the low energy limit of string theory can be determined using constraints from supersymmetry. Much of the physics can then be rather directly understood from this representation, as opposed to the more abstract CFT construction of the minimal models. Landau–Ginzburg theories can be generalized to multiple superfields, where the classification of superpotentials uses methods from singularity theory. There are also more general current algebra constructions.

19.5 Gepner models

Now we wish to use the exact CFTs from the previous section to construct string theories. In order to obtain central charge (9,9) we need several

minimal/Landau–Ginzburg models, with

$$\sum_i \frac{k_i}{k_i+2} = 3 \,. \tag{19.5.1}$$

There are many combinations that satisfy this. Now consider the product of the Landau–Ginzburg path integrals, where we sum over common periodic or antiperiodic boundary conditions on all the fermions ψ_i and $\tilde{\psi}_i$ at once. The result is modular-invariant: the modular transformations mix the path integral sectors in the usual way, and the left–right symmetry guarantees the absence of anomalous phases. In terms of the abstract CFT description this is the diagonal invariant, taking the same $N = 2$ representation on the left and right and summing over representations; to be precise, one separates each representation into two halves according to $\exp(\pi i F)$ before combining left and right.

This is a consistent CFT for either the type II or heterotic string, but it is not yet spacetime supersymmetric. We must now impose the GSO projection (19.1.6), namely

$$l + s_0 + s_1 + Q \in 2\mathbf{Z} \,. \tag{19.5.2}$$

Normally this is imposed as a $\mathbf{Z}_2$ projection, beginning with a spectrum for which the combination $l + s_0 + s_1 + Q$ takes only integer values. It is therefore necessary first to twist by the group generated by

$$g_q = \exp(\pi i s + 2\pi i Q) = \exp(\pi i s) \prod_i \exp(2\pi i Q_i) \,, \tag{19.5.3}$$

where we define s to be even in the NS sector and odd in the R sector. The extra factor of $exp(\pi i s)$ is needed because $l + s_0 + s_1$ is integer in the NS sector but half-integer in the R sector. The operator (19.5.3) contains the product of the $\mathbf{Z}_{k_i+2}$ generators for the separate minimal model factors and so generates $\mathbf{Z}_p$, where p is the least common multiple of the $k_i + 2$.

There are two possible subtleties. First, since the projection (19.5.3) is not left–right symmetric, modular invariance is not guaranteed. The issue is the same as for the orbifold, discussed in section 16.1, and the necessary and sufficient condition is level matching just as in that case. Second, the phase of the operator (19.5.3) is determined by level matching and may not be that which we wanted. In the references it is shown that under rather general conditions, which include the case at hand, these subtleties do not arise and so the resulting theory is consistent and supersymmetric. This argument also applies to a more general set of (2,2) CFTs known as Kazama–Suzuki theories, which are also constructed from current algebras.

Let us illustrate these general results for the notationally simple case

of N minimal model factors having equal levels k. The central charge condition

$$\frac{Nk}{k+2} = 3 \tag{19.5.4}$$

has integer solutions

$$k^N = 1^9,\ 2^6,\ 3^5,\ 6^4\ . \tag{19.5.5}$$

Before twisting, the NS–NS primaries are

$$\prod_{i=1}^{N} \psi_{m_i}^{j_i} \tilde{\psi}_{m_i}^{j_i} \exp\left[i \frac{2m(H_i + \tilde{H}_i)}{k^{1/2}(k+2)^{1/2}}\right] . \tag{19.5.6}$$

We continue to use the $SU(2)$ notation, though the common notation in the literature on this subject is to use integer-valued labels $l = 2j$ and $q = 2m$ (or, confusingly, m equal to twice its $SU(2)$ value).

Now twist by g_q. An operator of charge $Q_i = l$ depends on the free scalar H_i from the ith factor as

$$\exp\left[il\left(\frac{k+2}{k}\right)^{1/2} H_i\right] . \tag{19.5.7}$$

This picks up an extra phase $\exp(2\pi i n l)$ when transported around a vertex operator in a sector twisted by g_q^n. It follows that the vertex operators in that sector contain an additional factor

$$\exp\left[in\left(\frac{k}{k+2}\right)^{1/2} H_i\right] . \tag{19.5.8}$$

Thus the untwisted vertex operator (19.5.6) becomes

$$\prod_{i=1}^{N} \psi_{m_i}^{j_i} \tilde{\psi}_{m_i}^{j_i} \exp\left[i \frac{(2m_i + nk)H_i + 2m_i \tilde{H}_i}{k^{1/2}(k+2)^{1/2}}\right] . \tag{19.5.9}$$

Using eq. (19.5.4), this has total $U(1)$ charge

$$Q = \frac{1}{k+2} \sum_{i=1}^{N} (2m_i + nk) = 3n + \frac{2}{k+2} \sum_{i=1}^{N} m_i\ . \tag{19.5.10}$$

The level mismatch is

$$\begin{aligned} L_0 - \tilde{L}_0 &= \frac{1}{2k(k+2)} \sum_{i=1}^{N} \left\{(2m_i + nk)^2 - (2m_i)^2\right\} \\ &= \frac{3n^2}{2} + \frac{2n}{k+2} \sum_{i=1}^{N} m_i\ . \end{aligned} \tag{19.5.11}$$

Thus, requiring the charge (19.5.10) to be an integer implies that the level mismatch is a multiple of $\frac{1}{2}$, which is the appropriate result for the NS–NS sector before GSO projecting. The other sectors work as well.

Now let us look for $(c, \tilde{c})$ states. On the right-moving side, the charge and weight are given by the untwisted values (19.4.2a), so

$$\tilde{h} - \frac{\tilde{Q}}{2} = \sum_{i=1}^{N} \frac{j_i(j_i+1) - m_i(m_i+1)}{k+2}\ . \tag{19.5.12}$$

The chiral primaries have $m_i = j_i$ for all i. In the untwisted sector, $n = 0$, these are paired with chiral primaries on the left. The number of such states having $Q = 1$ is given by all sets of j_i such that

$$\sum_i j_i = \frac{k+2}{2}\ , \quad |j_i| \le \frac{k}{2}\ . \tag{19.5.13}$$

From the structure of the $N = 2$ superconformal representations one can show that there are no $(c, \tilde{c})$ states in the twisted sectors; these moduli, or the **27**s in the heterotic string, come entirely from the untwisted sector.

The $(a, \tilde{c})$ states, or $\overline{\mathbf{27}}$s, come from primaries with opposite m on the right and left,

$$m_i = -\tilde{m}_i = -j_i\ ,\ \text{all } i\ . \tag{19.5.14}$$

These must come from the twisted sectors. Note that the states (19.5.9) are in general excited states in their representations, and the $(a, \tilde{c})$ states are obtained with lowering operators. A little thought shows that $(a, \tilde{c})$ states can arise only if $\tilde{m}_i$ is independent of i; one such state is consistent with the conditions (19.5.13) for the 3^5 and 6^4 cases. For example, the $(a, \tilde{c})$ state in the 3^5 model is obtained from the state with $m_i = \frac{1}{2}$ and $n = 1$ by acting with $G^-_{-1/2}$ in each of the five factors. The reader can check that this has the correct weight and charge, and that the OPE implies that it is nonzero.

In summary, the numbers $(n_{\mathbf{27}}, n_{\overline{\mathbf{27}}})$ for the k^N models are

$$1^9 : (84,0)\ ,\quad 2^6 : (90,0)\ ,\quad 3^5 : (101,1)\ ,\quad 6^4 : (149,1)\ . \tag{19.5.15}$$

Connection to Calabi–Yau compactification

An interesting point about the Gepner models is that most are in the same moduli space as Calabi–Yau compactifications. The simplest example is 3^5, five copies of the $k = 3$ model. The discrete symmetry is

$$S_5 \times \mathbf{Z}_5^4\ , \tag{19.5.16}$$

where the permutation group S_5 interchanges the various factors. The $\mathbf{Z}_5$s come from the separate minimal model factors,

$$\exp(2\pi i Q_i)\ ,\quad i = 2,3,4,5\ . \tag{19.5.17}$$

The symmetry $\exp(2\pi i Q_1)$ is not independent because the projection (19.5.3) relates it to the others.

Now consider our simple example of a Calabi–Yau model, the quintic in CP^4, for the special polynomial

$$G(z) = z_1^5 + z_2^5 + z_3^5 + z_4^5 + z_5^5 \,. \tag{19.5.18}$$

This is invariant under the same discrete symmetry (19.5.16); the permutation acts on the z_i, and the four $\mathbf{Z}_5$s are

$$z_i \to \exp(2\pi i n_i/5) z_i \,, \quad n_1 = 0 \,. \tag{19.5.19}$$

We can set $n_1 = 0$ because an overall phase rotation of the z_i is trivial by the projective equivalence. Further, this Gepner model has 101 **27**s and one $\overline{\mathbf{27}}$, the same as the Calabi–Yau theory. The generations in each theory can be shown to fall into the same representations of the discrete symmetry. The only difference in the massless spectrum is that the Gepner model has four extra $U(1)$ gauge symmetries. These come from the currents ∂H_i for the separate factors, minus one linear combination that is already part of E_6. Such enhancements are common at special points of moduli space, as for toroidal compactification at the self-dual point. There are also extra E_6 singlet $U(1)$ charged states. All the extra states become massive by the Higgs mechanism as we move away from the Gepner point.

This is strong evidence that the 3^5 Gepner model is the same theory as the quintic (19.5.18). The same is true of other Gepner models, though in many cases one needs a Calabi–Yau manifold constructed from *weighted projective space*, where the projection (17.2.34) that defines CP^n is generalized to allow different scalings for the different z_i. To understand the connection in more detail, note the suggestive fact that the total Landau–Ginzburg superpotential

$$\sum_{i=1}^{5} \Phi_i^5 \tag{19.5.20}$$

is the same as the defining polynomial (19.5.18) of the Calabi–Yau manifold.

To make this observation more precise we generalize the previous Landau–Ginzburg construction, starting again with a theory of (2,2) rigid supersymmetry obtained by dimensional reduction from a $d = 4$, $N = 1$ theory. We take the five superfields Φ_i and an additional superfield P, as well as a $U(1)$ gauge field. The superpotential is

$$W = P G(\Phi) \,, \tag{19.5.21}$$

where we take an arbitrary quintic polynomial as in the Calabi–Yau case. This is gauge-invariant with $U(1)$ gauge charges

$$q_\Phi = 1 \,, \quad q_P = -5 \,. \tag{19.5.22}$$

The gauge coupling is e, and if we start from the general four-dimensional action (B.2.16) then there is one more parameter at our disposal, a $U(1)$ Fayet–Iliopoulos term which we will denote $\xi = -r/2$.

The potential energy for this *linear sigma model* is

$$U = |G(\phi)|^2 + |p|^2 \sum_{i=1}^{5} \left| \frac{\partial G}{\partial \phi_i} \right| + \frac{e^2}{2} \left(r + 5|p|^2 - \sum_{i=1}^{5} |\phi_i|^2 \right) + (A_2^2 + A_3^2) \left(25|p|^2 + \sum_{i=1}^{5} |\phi_i|^2 \right) \,, \tag{19.5.23}$$

coming from the F-terms, the D-terms, and the dimensional reduction of the kinetic terms. We use lower case letters for the scalar components of superfields. We are interested in the low energy dynamics of this field theory, and so in those points in field space where the potential vanishes.

Let us first restrict attention to polynomials that are *transverse*, meaning that the five equations

$$\frac{\partial G}{\partial \phi_i} = 0 \tag{19.5.24}$$

have no simultaneous solutions except at $\phi = 0$. The reason for imposing this condition is that we are going to make contact with the Calabi–Yau manifold defined by the embedding $G(\phi) = 0$. If the gradient vanishes at any point, the condition $G(\phi) = 0$ degenerates and does not define a smooth manifold (if the gradient vanishes at some point ϕ_i, this point automatically lies on the submanifold $G = 0$ because $\phi_i \partial_i G = 5G$). These are actually five equations for four independent unknowns because of the projective equivalence (homogeneity of G). They therefore generically have no solutions other than $\phi = 0$; the case in which they do is very interesting and will be discussed in section 19.7.

Let us first consider the case $r > 0$. Transversality implies that the second term in the potential vanishes only if p vanishes and/or all the ϕ_i vanish. Combined with the vanishing of the third term this implies that

$$p = 0 \,, \quad \sum_{i=1}^{5} |\phi_i|^2 = r \,. \tag{19.5.25}$$

The fourth term forces $A_2 = A_3 = 0$, so finally we are left with

$$G(\phi) = 0 \,. \tag{19.5.26}$$

The manifold of vacua is identical to the Calabi–Yau manifold defined by $G = 0$ in CP^4. The condition (19.5.25) on ϕ can be regarded as a partial fixing of the projective invariance. The remaining invariance, a common phase rotation of the ϕ_i, is the $U(1)$ gauge invariance. The metric on this

space is induced by the flat metric in the kinetic term $|\partial_a \phi_i|^2$. In particular, the size of the Calabi–Yau manifold is

$$R_c^2 \propto r \,. \tag{19.5.27}$$

One can show that this classical analysis becomes quantitatively accurate for large r.

Now consider $r < 0$. The unique zero-energy point is

$$|p|^2 = \frac{r}{5} \,, \quad \phi_i = 0 \,, \quad A_2 = A_3 = 0 \,. \tag{19.5.28}$$

Although this is an isolated zero of the potential, the fields ϕ_i are massless because their potential is of order $|\phi|^8$. In fact, they are described by a generalized Landau–Ginzburg theory, with superpotential

$$W = \langle p \rangle \, G(\Phi) \,; \tag{19.5.29}$$

we can replace p with its mean value because the fluctuations are massive. This superpotential produces a nontrivial critical theory, by a generalization of the earlier argument.

We have seen that for positive values the parameter r has an interpretation as a modulus. It is the only Kähler modulus for this CFT. The complex structure moduli are the parameters in the polynomial G. Thus we conclude that the Landau–Ginzburg theories represent a different region in the same moduli space. The identification (19.5.27) would suggest that they correspond to unphysical negative values of R_c^2, but that identification is valid only at large r. There is an important distinction between the $r \to +\infty$ and $r \to -\infty$ limits. The former really represents an infinite distance in moduli space, corresponding to the fact that the Calabi–Yau space is becoming very big. As $r \to -\infty$, however, the low energy critical theory is determined by the superpotential (19.5.29). This depends on r through $\langle p \rangle$, but that can be absorbed in a rescaling of the fields. It follows that the low energy theory becomes independent of r as $r \to -\infty$. This point is actually at finite distance in moduli space, and the region of moduli space described by the Landau–Ginzburg theory is in the interior.

Recall that to construct a string theory from the Landau–Ginzburg theory we had to twist by the Z_5 symmetry g_q. It is interesting to see how this arises in the present construction. The expectation value of p breaks the $U(1)$ gauge symmetry, but a discrete subgroup

$$p \to p \,, \quad \phi_i \to \exp(2\pi i/5)\phi_i \,, \tag{19.5.30}$$

remains as an unbroken gauge symmetry of the low energy theory. As discussed in section 8.5, gauging of a discrete symmetry is one way to think about the twisting construction.

The analysis above breaks down at $r = 0$. The potential requires p and ϕ_i to vanish, and the fields $A_{2,3}$ then have no potential. This infinite volume in field space could produce a singularity that prevents continuation from positive to negative r. In fact, there is a singularity (to jump ahead a little, it is the mirror of the conifold singularity in the complex structure), but it does not prevent continuation between the two regions. The point is that the Kähler modulus is a complex field, and we have identified only its real part. To find the imaginary part, the $B_{i\bar{j}}$ background, recall that this gives a total derivative on the world-sheet. There is one natural total derivative to add to the present theory, namely

$$i\frac{\theta}{2\pi}\int F_2 \,, \tag{19.5.31}$$

where F_2 is the $U(1)$ field strength 2-form. This does indeed correspond to the imaginary part of the modulus. In the Calabi–Yau phase one can use the equation of motion for the gauge field A_I to show this.

At $r = 0$ but with θ nonzero, the world-sheet theory is nonsingular and so one can continue past the $r = \theta = 0$ singularity. The θ parameter in two dimensions has been extensively discussed in field theory, in part as a model for the instanton θ parameter in four dimensions. It does not change the equations of motion but changes the boundary conditions, so that there is a fractional electric flux

$$F_{12} = \frac{\theta}{2\pi} \,. \tag{19.5.32}$$

This flux will produce a nonzero energy density unless it is screened. A fractional flux cannot be screened by massive integer charged quanta. It can be screened by massless integer charges in two dimensions, or by the condensate if the $U(1)$ symmetry is spontaneously broken. In the present case a charged field, either p or ϕ_i, has an expectation value when r is nonzero and then the θ parameter has no effect. When r vanishes the $U(1)$ is unbroken. If $A_{2,3}$ are nonzero then all charged fields are massive and there is an energy density. Only at the point where all the fields vanish does the energy density go to zero, so the field space is effectively compact at low energy and the theory is nonsingular.

Thus the two parts of moduli space are smoothly connected. The term *phases* is often used to describe the two regions. Like the water/steam case, the two phases are continuously connected but display qualitatively different physics.

We have focused on the simplest example, but there are clearly many possible generalizations. It is interesting to note the following point. In the (2,2) algebra there are left- and right-moving $U(1)$s, under which the fields that move in the opposite direction are neutral. In the Landau–Ginzburg theory we identified the sum of these charges as an R symmetry.

To identify the separate symmetries we need also the rotation S_1 in the (2,3) plane, which in the dimensionally reduced theory becomes an internal symmetry. Since a four-dimensional Weyl spinor has $s_0 = -s_1$, the charge S_1 is correlated with the direction of motion, while the R charge is independent of it. By forming linear combinations of R and S_1 we can obtain symmetries Q and $\tilde{Q}$ under which either the left- or right-movers are neutral. To do this simultaneously for all fields, we must assign a common R-charge to all superfields. Since there is a gauge field we must be concerned about a possible anomaly in a current that acts on Fermi fields moving in one direction. The anomaly comes from a current–current OPE, as in section 12.2. By the above construction Q and $\tilde{Q}$ are the same for all superfields and so their anomalies are proportional to the sum of the gauge $U(1)$ charges. This is $-5 + (5 \times 1) = 0$ for the model at hand, so the anomalies vanish. If there were an anomaly, then one would not expect to have independent conserved Q and $\tilde{Q}$ and so there could be no (2,2) superconformal algebra. In fact one finds in this case quantum corrections that invalidate the classical analysis used above. In more general models, the anomaly cancellation condition turns out to be equivalent to the condition that in the Calabi–Yau phase the first Chern class vanishes, which was a necessary condition for conformal invariance.

19.6 Mirror symmetry and applications

In CFT it is arbitrary which states we call $(c, \tilde{c})$ and which $(a, \tilde{c})$. These just differ by a redefinition $H \to -H$ of the free scalar for the left-moving $U(1)$ current. However, for CFTs obtained from Calabi–Yau compactification these have very different geometric interpretations, in terms of the Kähler and complex structure moduli respectively. This suggests that Calabi–Yau manifolds might exist in *mirror pairs* $\mathcal{M}$ and $\mathcal{W}$, where

$$(h^{1,1}, h^{2,1})_{\mathcal{M}} = (h^{2,1}, h^{1,1})_{\mathcal{W}} , \tag{19.6.1}$$

and where the two CFTs are isomorphic, being related by $H \to -H$.

We can illustrate this for the analog of Calabi–Yau compactification with two compact dimensions. The holonomy is in $SU(1)$, which is trivial, so the compact dimensions must be a 2-torus. Calling the compact directions $x^{8,9}$, act with T-duality in the 9-direction. This flips the sign of $X_L^9(z)$ and so that of $\psi^9(z)$. Therefore it also flips the $U(1)$ current $i\psi^8\psi^9$. The 2-torus is thus its own mirror, but with different values of the moduli. Referring back to the discussion at the end of section 8.4, we noted there that T-duality on one axis interchanged the Kähler modulus ρ with the complex structure modulus τ. When the Kähler modulus ρ is large, the 2-torus is large; when the complex structure modulus τ is large, the 2-torus is long and thin.

The T-duality between the IIA and IIB theories implies that the IIA string on one 2-torus is the same as the IIB string on its mirror. This last will also be true for six-dimensional Calabi–Yau manifolds: the reversal of the $U(1)$ charge on one side also reverses the GSO projection on that side, interchanging the two type II strings. This is consistent with our results (19.2.1) and (19.2.2) for the moduli spaces. If we put the IIA theory on $\mathcal{M}$ and the IIB theory on $\mathcal{W}$, the number of vector multiplets $h^{1,1}_{\mathcal{M}} = h^{2,1}_{\mathcal{W}}$ is the same, and similarly the number of hypermultiplets.

The explicit construction of the mirror transformation for six-dimensional Calabi–Yau manifolds is less straightforward. Circumstantial evidence for the existence of mirror pairs was found when the $(h^{1,1}, h^{2,1})$ values were plotted for large classes of Calabi–Yau manifolds: if a given point was present, then a manifold with reversed Hodge numbers $(h^{2,1}, h^{1,1})$ usually also existed. This does not prove that the manifolds are mirrors, because the Hodge numbers do not determine the full CFT, but it is suggestive.

There is one class of Calabi–Yau manifolds where the mirror can be constructed explicitly, the ones that are related to Gepner models. Consider our usual example 3^5. The subgroup of the global $\mathbf{Z}_5^4$ symmetry that commutes with the spacetime supersymmetry is the group $\Gamma = \mathbf{Z}_5^3$ with elements

$$\exp\left\{2\pi i[r(Q_2 - Q_3) + s(Q_3 - Q_4) + t(Q_4 - Q_5)]\right\} \tag{19.6.2}$$

for integer r, s, and t. We claim that if the theory is twisted by Γ then something simple happens. Consider first a single periodic scalar X compactified at radius $R = (\alpha' n)^{1/2}$ for some integer n. The translation

$$X \to X + 2\pi(\alpha'/n)^{1/2} \tag{19.6.3}$$

generates a $\mathbf{Z}_n$. If we twist by this $\mathbf{Z}_n$ then we obtain the scalar at radius

$$R' = (\alpha'/n)^{1/2} . \tag{19.6.4}$$

This is T-dual to the original radius, so the result is isomorphic to the original CFT, differing only by $X_L(z) \to -X_L(z)$. We leave it to the reader to show that the twist by Γ has the same effect in the 3^5 model, turning the Gepner CFT into one that is isomorphic under $H(z) \to -H(z)$. The point of this exercise is that we now have a geometric relation between the original theory and its mirror. This Gepner model maps to the quintic, which we will denote $\mathcal{M}$. The group Γ acts on the CP^4 coordinates in the Calabi–Yau description as

$$(z_1, z_2, z_3, z_4, z_5) \to (z_1, \alpha^r z_2, \alpha^{s-r} z_3, \alpha^{t-s} z_4, \alpha^s z_5) , \tag{19.6.5}$$

where $\alpha = \exp(2\pi i/5)$. Twisting by Γ produces the coset space

$$\mathscr{W} = \mathscr{M}/\Gamma \,. \tag{19.6.6}$$

Some of the transformations have fixed points, so the space $\mathscr{W}$ is not a manifold but has orbifold singularities. These can be blown up, and the resulting smooth manifold indeed has Hodge numbers $(h^{1,1}, h^{2,1}) = (101, 1)$, the reverse of the $(1, 101)$ of the quintic.

The explicit twist can be carried out only at the Gepner point in moduli space, but the existence of mirror symmetry at this point is sufficient to imply it for the whole moduli space. The point is that the isomorphism of CFTs implies a one-to-one mapping of moduli, so the effect of turning on a modulus in one theory is identical to that of the equivalent modulus in the dual theory.

There have been many attempts to derive mirror symmetry in a more general way, with partial success. *Toric geometry* is a generalization of the projective identification that defines CP^n corresponding to the most general linear sigma model. It provides a framework for constructing many Calabi–Yau manifolds and their mirrors. In another direction, one might wonder whether a connection can be made to T-duality, as in the case of the 2-torus. Indeed, this has been done as follows. Put the IIA string on a Calabi–Yau manifold $\mathscr{M}$, and consider the manifold of states of a D0-brane: this is just the Calabi–Yau manifold itself, since the D0-brane can be anywhere. In the IIB string on the mirror manifold $\mathscr{W}$, BPS states come from Dp-branes with p odd, wrapped on nontrivial cycles of the mirror. Since $b_1 = b_5 = 0$, we must have $p = 3$. This immediately suggests a T-duality on three axes. Three of the coordinates of the D0-brane map to internal Wilson lines on the D3-brane, which therefore must be topologically a 3-torus. By following this line of argument one can show that $\mathscr{W}$ is a T^3 *fibration*. That is, it is locally a product $T^3 \times X$ with X a three-manifold, but with the shape of the T^3 *fiber* varying over X. The mirror transformation is T-duality on the three axes of T^3, and $\mathscr{M}$ is also a T^3 fibration. Any Calabi–Yau space with a Calabi–Yau mirror must be such a fibration; this property is not uncommon.

Moduli spaces

An important consequence of mirror symmetry is that it allows the full low energy field theory to be obtained at string tree level but exactly in world-sheet perturbation theory. We have argued that the field theory calculation of the complex structure moduli space is exact, but now we can also obtain the Kähler moduli space from the complex structure moduli space of the mirror.

Let us explain further how this works, taking our usual example of the

quintic. We focus on the Kähler moduli space, which has a single modulus T. The general polynomial invariant under the $\mathbf{Z}_5^3$ twist (19.6.5) is

$$G(z) = z_1^5 + z_2^5 + z_3^5 + z_4^5 + z_5^5 + 5\psi z_1 z_2 z_3 z_4 z_5 \,. \tag{19.6.7}$$

This polynomial is parameterized by one complex parameter ψ, which survives as the sole complex structure modulus of the mirror.

The low energy action for the complex structure modulus of the mirror can be obtained as described in section 17.4. The special coordinates and periods are defined by the integrals of the harmonic $(3,0)$-form Ω over closed cycles,

$$Z^I = \int_{A^I} \Omega_{3,0} \;, \quad \mathscr{G}_I(Z) = \int_{B_I} \Omega_{3,0} \,. \tag{19.6.8}$$

The range of I is from 1 to $h^{2,1} + 1$, which is 2 in this example. For this construction the cycles and $\Omega_{3,0}$ can be given explicitly and the integrals evaluated. The result is that Z^a and $\mathscr{G}_I$ are hypergeometric functions of ψ. These in turn determine the prepotential $\mathscr{G} = \frac{1}{2} Z^I \mathscr{G}_I$, and so the low energy action for ψ.

There are three special points in this space, $\psi = 0, 1$, and ∞. The Gepner point $\psi = 0$ is where the theory can be described by a product of minimal models. The *conifold* point $\psi = 1$ is a singular Calabi–Yau space. The singularity is very interesting, and will be described in detail in the next section. The *large complex structure limit* is $\psi = \infty$. It is the only point at infinite distance in the moduli space metric, and so must be related by mirror symmetry to the large-radius limit $T = \infty$.

To exploit mirror symmetry we need the precise mapping between ψ and T. As in section 19.2, T is related to the special coordinates on Kähler moduli space by $T = iX^1/X^0$. The Z^I are special coordinates on the complex structure moduli space. Special geometry allows only a symplectic transformation (B.7.20) between different sets of special coordinates. The precise form of the symplectic transformation (which depends on the basis of cycles used in eq. (19.6.8)) can be found by comparing the exact prepotential as $\psi \to \infty$ with the large-radius limit of the Kähler prepotential. To leading approximation at large radius the result is

$$T \approx \frac{5}{2\pi} \ln(5\psi) \,. \tag{19.6.9}$$

The full mapping gives the exact prepotential for T and so the low energy action. Expanded around large T it agrees with the general form in section 19.2,

$$F = (X^0)^2 \left[\frac{5i}{6} T^3 - \frac{25i}{2\pi^3} \zeta(3) + \sum_{k=1}^{\infty} C_k \exp(-2\pi k T) \right] . \tag{19.6.10}$$

The first term is the tree-level interaction, the normalization agreeing with the expression (17.4.16) in terms of the intersection number. The second term is the three-loop correction, related to the R^4 term in the effective action which has the distinctive coefficient $\zeta(3)$. The final term represents a sum over instantons, where k is the total winding number of the world-sheet over the nontrivial 2-cycle of the Calabi–Yau manifold. The contribution of each instanton is rather simple, so the numerical constant just counts the number n_k of instantons (holomorphic curves) of given winding number, up to some simple factors. Expanding out the result from the mirror map gives immediately the number of such curves, which grows rapidly:

$$n_k = 2875,\ 609250,\ 317206375,\ 242467530000,\ \ldots\ . \tag{19.6.11}$$

The direct geometric determination of n_k is much more involved. Initially only the first few values were known, but now the full series has been determined, in agreement with the mirror symmetry prediction.

All of the above applies to string tree level. The string corrections depend on which string theory is put on the Calabi–Yau space. For the IIA string, the Kähler moduli are in vector multiplets and their low energy action receives no corrections. For the IIB string the low energy action for the complex structure moduli receives no corrections. For the heterotic or type I string there is only $d = 4$, $N = 1$ supersymmetry and so both moduli spaces may be corrected, while the superpotential may receive nonperturbative corrections.

The flop

The integral of the Kähler form over a 2-cycle is

$$\mathrm{Re}(T^A) = \int_{N^A} J_{1,1} = \int_{N^A} d^2w\, G_{i\bar{\jmath}} \frac{\partial X^i}{\partial w} \frac{\partial X^{\bar{\jmath}}}{\partial \bar{w}} > 0\ . \tag{19.6.12}$$

This must be positive for every 2-cycle, and similarly for the integral of $J_{1,1} \wedge J_{1,1}$ over any 4-cycle and of $J_{1,1} \wedge J_{1,1} \wedge J_{1,1}$ over the whole Calabi–Yau space. These conditions define the Kähler moduli space as a cone in the space parameterized by T^A.

In combination with mirror symmetry, this presents a puzzle. The boundary of the cone has codimension 1, since $\mathrm{Re}(T^A) = 0$ is a single real condition on the geometry. This must agree with the structure of the complex structure moduli space of the mirror manifold. The puzzle is that such boundaries do not appear in the complex structure moduli space. All special points in the latter are determined by complex equations, and so lie on manifolds of even codimension; an example is the point $\psi = 1$ of the quintic.

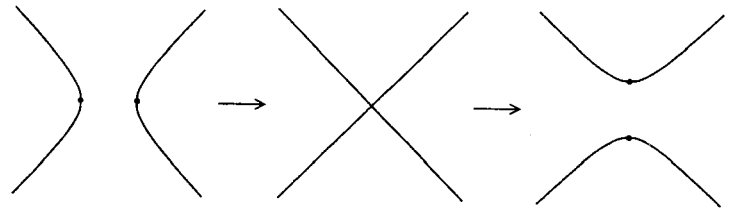


Fig. 19.1. The flop transition, projected onto the $\mathrm{Re}(\phi_1)$–$\mathrm{Re}(\rho_1)$ plane. The dots indicate the intersection of the minimal 2-spheres with this plane. The conifold transition is similar, but with a 3-sphere before the transition and a 2-sphere after.

The resolution of this puzzle is suggested by geometry. The integral of the Kähler form represents the minimum volume of a 2-sphere in the given homology class. This goes to zero at the boundary and would be negative beyond it. There is a sense in which the geometry can be continued to 'negative volumes.' A model for the region of the small sphere is given in terms of four complex scalars ϕ_1, ϕ_2, ρ_1, and ρ_2. To make a six-dimensional manifold we impose the condition

$$\phi^* \cdot \phi - \rho^* \cdot \rho - r = 0 \tag{19.6.13}$$

for some real parameter r, and also the identification

$$(\phi_i, \rho_i) \cong (e^{i\lambda}\phi_i, e^{-i\lambda}\rho_i)\ . \tag{19.6.14}$$

Let r first be positive, and look at the space parameterized by ϕ when $\rho_i = 0$. The condition (19.6.13) defines a 3-sphere and the identification (19.6.14) reduces this to a 2-sphere, with volume $4\pi r$. As ρ_i varies, the size of this 2-sphere grows, so $4\pi r$ is the minimum volume. For $r = 0$ the volume is zero and the space singular, but for $r < 0$ the space is perfectly smooth: the previous picture goes through with ϕ_i and ρ_i interchanged. The smallest 2-sphere has volume $4\pi|r|$, but it is a different 2-sphere from the one considered at positive r. This is shown schematically in figure 19.1. The transition from positive to negative r is known as a *flop*.

The mirror symmetry argument strongly suggests that the CFT at the $r = 0$ point is nonsingular, and that one can pass smoothly through it. One can check this in various ways. The flop transition does not change the Hodge numbers $h^{1,1}$ and $h^{2,1}$; this is consistent with the fact that nothing is happening in the mirror description. It does change the topology, however, as measured for example by the intersection numbers of various 2-cycles. These intersection numbers determine the low energy interactions of the Kähler moduli in the field-theory limit, so we need to understand how the

discontinuity of the intersection numbers is compatible with the continuity of the physics in the mirror description. The point is that as we approach the boundary of the Kähler cone, the action for instantons wrapped on the shrinking 2-cycle becomes very small and so the instanton corrections important. The smoothness of the transition has been checked in two ways. The first is by evaluating the instanton sum near the transition: the difference between the instanton contributions on the two sides of the transition just offsets the discontinuity in the intersection number. The second is by looking at points on either side of the transition but far from it: the calculation in the mirror is then found to reduce to the appropriate intersection number in the various limits. As a final argument for the smoothness of the transition we can use a linear sigma model. In fact, if we take four chiral superfields and gauge the $U(1)$ symmetry (19.6.14), the D-term condition and gauge equivalence just reproduce the above model of the flop. As in the earlier application of the linear sigma model, we can interpolate from positive r to negative r along a path of nonzero θ.

The full picture is that in each moduli space the only singularities are of codimension at least 2. In the complex structure description the topology is the same throughout. In the mirror-equivalent Kähler description the cones for the different topologies join smoothly. However, smooth Calabi–Yau manifolds do not cover the whole Kähler moduli space. Some regions have a description in terms of orbifolds of Calabi–Yau manifolds, or Landau–Ginzburg models, or a hybrid of the two.

We cannot illustrate the flop transition with the quintic. This has only one Kähler modulus, and when it vanishes the volume of the whole Calabi–Yau manifold goes to zero. Incidentally, the moduli space of the quintic (in either the Kähler or complex structure description) is multiply connected. One nontrivial path runs from ψ to $\exp(2\pi i/5)\psi$; these points are equivalent with the coordinate change $z_1 \to \exp(-2\pi i/5)z_1$. The Gepner model is a fixed point for this operation. A second nontrivial path circles the conifold point $\psi = 1$. Together these generate the full modular group. This acts in a complicated way in terms of the variable T, but has a fundamental region with $\mathrm{Re}(T)$ positive.

We could consider a situation in which the moduli are time-dependent, moving from one Kähler cone to another. From the four-dimensional point of view, this is just the smooth evolution of a scalar field. If we consider the same process with the radius of the manifold blown up to macroscopic scales, we would see a region of the compact space pinch down and then expand in a topologically distinct way.

In general relativity the geometry of spacetime is dynamical, but it is an old question as to whether the topology is as well: spacetime can bend, but can it break? String theory, as a complete theory of quantum gravity, should answer this, and it does. At least in the limited way considered here, and in the somewhat more drastic way that we are about to consider, topology can change. It remains to understand the full extent of this and to learn what ideas are to replace geometry and topology as the foundation of our understanding of spacetime.

19.7 The conifold

Following eq. (19.5.24) we have discussed the requirement that the polynomial defining the embedding of the Calabi–Yau space in CP^4 be transverse, its gradient nonvanishing. As explained there, the vanishing of the gradient gives five conditions for four unknowns and generically has no solutions. If we allow the complex structure moduli to vary we get additional unknowns, and there will in general be solutions having complex codimension 1 (real codimension 2). The conifold is a realization of this. The vanishing of the gradient implies that

$$z_i^5 = \psi z_1 z_2 z_3 z_4 z_5 , \quad i = 1,\dots,5 . \tag{19.7.1}$$

Multiplying these five equations together implies either that all the z_i vanish (which point is excluded from CP^4 by definition) or that

$$\psi^5 = 1 . \tag{19.7.2}$$

This has isolated solutions in the complex plane, consistent with the counting. We have noted above that ψ and $\exp(2\pi i/5)\psi$ are equivalent, so there is one possible singular point, $\psi = 1$. The singular manifold is known as a *conifold*, with $\psi = 1$ the conifold point in moduli space. The singularity, or node, on the manifold itself is at the point $z_1 = z_2 = z_3 = z_4 = z_5$.

Let us see the nature of the singularity. Generically, and in this example, the matrix of second derivatives of G is nonvanishing. We can then find complex coordinates $w = (w_1,\dots,w_4)$ such that near the singularity the manifold is of the form

$$\sum_i w_i^2 = 0 , \tag{19.7.3}$$

the gradient of the left-hand side vanishing at the point $w = 0$. These are ordinary, not projective coordinates: one can fix the projective invariance by $z_1 = 1$, and the w_i are linear functions of $z_2,\dots,z_5$. This equation then defines a space of $4 - 1$ complex dimensions as it should. The space is a cone, meaning that if w is on it then so is aw for any real a. To see the cross-section of the cone, consider the intersection with the 3-sphere

$$\sum_i |w_i^2| = 2\rho^2 . \tag{19.7.4}$$

Separating w_i into real and imaginary parts $w_i = x_i + iy_i$, this becomes

$$x \cdot x = \rho^2 , \quad y \cdot y = \rho^2 , \quad x \cdot y = 0 . \tag{19.7.5}$$

That is, x lies on a 3-sphere, and for given x the coordinate y lies on a 2-sphere. This is in fact a direct product, so the whole geometry near the singularity is

$$S^3 \times S^2 \times \mathbf{R}^+ . \tag{19.7.6}$$

If the complex structure is deformed away from the singular value then the embedding equation becomes

$$\sum_i w_i^2 = \psi - 1 . \tag{19.7.7}$$

There is now a minimum 3-sphere of radius $|\psi - 1|^{1/2}$. For example, taking $\psi - 1$ to be real, this would be given by

$$x \cdot x = |\psi - 1| , \quad y = 0 . \tag{19.7.8}$$

We have seen that for manifolds with orbifold or flop singularities, the CFT and associated string theory remain perfectly well-behaved. This is not the case at a conifold singularity. The exact calculation described in the previous section shows that there is a singularity at the conifold point in moduli space. Specifically, let us take the A^1 cycle to be the 3-sphere that is contracting to zero size at $\psi = 1$. The special coordinate Z^1 is defined by an integral (19.6.8) over A^1, so it must be that in terms of this coordinate the conifold singularity is at $Z^1 = 0$. The result of the exact calculation is then that the period has a singularity

$$\mathcal{G}_1 = \frac{1}{2\pi i} Z^1 \ln Z^1 + \text{holomorphic terms} . \tag{19.7.9}$$

This implies in turn a logarithmic singularity in the metric $G_{1\bar{1}}$ on the moduli space. The singularity (19.7.9) can be understood as follows. Observe that if Z^1 is taken once around the origin then the period is multivalued:

$$\mathcal{G}_1 \to \mathcal{G}_1 + Z^1 . \tag{19.7.10}$$

Now, this period is defined by an integral (19.6.8) over a cycle B_1 that intersects the shrinking cycle A^1 once. This does not define B_1 uniquely, and it is a general result that if we take a surface in the topological class of B_1 and follow it as we deform the complex structure through a cycle around the conifold point, then it ends up as a cycle topologically equivalent to $B_1 + A^1$, which also intersects A^1 once. This *monodromy* of the cycles translates into the monodromy (19.7.10) of the period.

We wish to understand the meaning of this singularity. We focus on the IIB string, where the issue is particularly sharp. In this case the complex structure moduli are in vector multiplets, and so the low energy action

does not receive quantum corrections. The conifold singularity is then a property of the exact low energy field theory.

A general physical principle, which seems to hold true even in light of all recent discoveries about dynamics, is that singularities in low energy actions are IR effects, arising because one or more particles is becoming massless. For a nonsingular description of the physics we must keep these extra massless particles explicitly in the effective theory. We then need to understand why a particle would become massless at the conifold point in moduli space. The fact that this point is associated with a 3-cycle shrinking to zero size suggests a natural mechanism. A 3-brane wrapped around this surface would at least classically have a mass proportional to its area, and so become massless at $Z^1 = 0$.

This classical reasoning could be invalidated by quantum corrections, which might add a zero-point energy to the mass of the soliton. This does not happen for the following reason. The vector multiplet Z^1 comes from the (2,1)-form ω_1 that has unit integral over the cycle A^1 and zero integral over the other basis 3-cycles. In other words, the R–R 4-form potential is

$$c_{\mu npq}(x, y) = c_\mu^1(x)\omega_{1npq}(y) , \tag{19.7.11}$$

with c_μ^1 the four-dimensional gauge field. For a D3-brane whose world-volume D is the product of the cycle A^1 on which the brane is wrapped and a path P in the noncompact dimensions, the coupling to the R–R 4-form is

$$\int_D c_4 = \int_P c_1^1 . \tag{19.7.12}$$

The D3-brane thus has unit charge under the $U(1)$ gauge symmetry associated with the vector multiplet of Z^1. There is a BPS bound that the mass of any state with $U(1)$ charge is at least the charge times $|Z_1|$, times an additional nonzero factor. A BPS state, which attains the bound, thus has a mass that vanishes at $Z_1 = 0$. For the wrapped D3-brane, an analysis far from the conifold point in moduli space, where it is large and its world-volume theory weakly coupled, shows that it has one hypermultiplet of BPS states. This is also consistent with the low energy supersymmetry algebra, which allows a mass term (B.7.11) proportional to $|Z^1|$ for a charged hypermultiplet.

Finally, the logarithm in the low energy effective action arises from loops of the light charged particles. By a standard field theory calculation a hypermultiplet of unit charge and mass M contributes

$$-\frac{1}{32\pi^2} \ln(\Lambda^2/M^2) F_{\mu\nu} F^{\mu\nu} \tag{19.7.13}$$

to the effective Lagrangian density. Here Λ is the effective cutoff on the momentum integral. This is in precise agreement with the singular-

ity (19.7.9): $N = 2$ supergravity implies that the gauge kinetic term is proportional to

$$\frac{1}{8\pi}\mathrm{Re}(i\partial_1\mathscr{G}_1)\,. \tag{19.7.14}$$

Thus the singular interactions at the conifold point, though they are found in a tree-level string calculation, can only be understood in terms of the full nonperturbative spectrum of the theory. This is another indication of the tight structure of nonperturbative string theory. We call the D-brane nonperturbative because it is not part of the ordinary string spectrum, and because at any fixed Z^1 the ratio of its mass to the masses of the string states goes to infinity as g is taken to zero. Note that in four dimensions the scale $g^{-1}\alpha'^{-1/2}$ of its mass is the four-dimensional Planck mass up to numerical factors. This is consistent with the fact that the BPS bound is derived in supergravity using only the gravitational and gauge part of the action, and so when written in units of the Planck scale cannot depend on the dilaton.

In some early papers the state that is becoming massless is referred to as a black hole. As discussed in section 14.8, the black hole and D-brane pictures apply in different regimes. In the present case the particle is singly charged and so the D-brane picture is the relevant one in the string perturbative regime $g < 1$. For $g > 1$ we would have to use a dual description of the IIB string; in this description the D-brane picture is again the relevant one.

Previously we encountered D-branes as large, essentially classical objects. It is not clear in what regimes it is sensible to sum over virtual D-branes, but clearly here where a D-brane becomes a light particle it is necessary to do so. One might think that as $g = e^{\Phi}$ goes to zero, the D-brane would have to decouple because it becomes very massive, meaning that its effect would go to zero. However, the complex structure action is independent of the dilaton Φ. Evidently we must take the upper cutoff Λ in the loop amplitude (19.7.13) also to scale as $1/g$ as compared to the string mass. This is another indication of the existence of distances shorter than the string scale.

The conifold transition

We should consider the possibility that at the point where the D-brane hypermultiplet becomes massless, there is another branch of moduli space where it acquires an expectation value. This does not happen in the example above because there is a quartic potential. In the notation of section B.7, where the two scalars in the hypermultiplet are denoted Φ_α,

the condition that the potential vanish is

$$\Phi^\dagger_\alpha\sigma^A_{\alpha\beta}\Phi_\beta = 0\,, \quad A = 1,2,3\,. \tag{19.7.15}$$

It is easy to show that the only solution is $\Phi_1 = \Phi_2 = 0$. However, in more intricate examples a new branch of moduli space does emerge from a conifold point.

Let us consider a different point in the complex structure moduli space of the quintic, where the embedding equation is

$$z^1H_1(z) + z^2H_2(z) = 0\,, \tag{19.7.16}$$

with H_1 and H_2 generic quartic polynomials in the z^i. This has singular points when

$$z^1 = z^2 = H_1(z) = H_2(z) = 0\,. \tag{19.7.17}$$

The simultaneous quartic equations generically have 16 solutions, so this is the number of singular points on the Calabi–Yau manifold. Sixteen 3-spheres have shrunk to zero size.

The new feature of this example is that the shrinking 3-spheres are not all topologically distinct. Their sum is trivial in homology, which is to say that there is a four-dimensional surface whose boundary consists of these sixteen 3-spheres. Thus there are only fifteen distinct homology cycles and so fifteen associated $U(1)$ gauge groups. However, there are sixteen charged hypermultiplets that become massless at the point (19.7.16) in moduli space, since a D3-brane can wrap each small 3-sphere. The fact that the sum of the cycles is trivial translates into the statement that the sum of the charges of the sixteen light hypermultiplets is zero. Labeling the hypermultiplets by $i = 1,\ldots,16$, we can take a basis $I = 1,\ldots,15$ for the $U(1)$s such that the charges are

$$q^I{}_i = \delta^I{}_i\,, \ i = 1,\ldots,15\,, \quad q^I{}_{16} = -1\,. \tag{19.7.18}$$

The condition that the potential for the charged hypermultiplets vanish is then

$$\Phi^\dagger_{i\alpha}\sigma^A_{\alpha\beta}\Phi_{i\beta} - \Phi^\dagger_{16\alpha}\sigma^A_{\alpha\beta}\Phi_{16\beta} = 0\,, \quad A = 1,2,3\,, \quad i = 1,\ldots,15\,. \tag{19.7.19}$$

This has nonzero solutions, namely

$$\Phi_{i\alpha} = \Phi_{16\alpha}\,, \quad i = 1,\ldots,15\,. \tag{19.7.20}$$

Thus there is a new branch of moduli space. The fifteen $U(1)$s are spontaneously broken, so the number of vector multiplets is reduced from 101 to 86, while the potential leaves one additional hypermultiplet modulus (19.7.20) for a total of two.

As with the flop transition, this stringy phenomenon is already hinted at in geometry. We have discussed blowing up the 3-sphere at the apex

of the cone (19.7.6), but it is also possible to blow up the 2-sphere. There are certain global obstructions to how this can be done; it cannot be done for the simple singularity at $\psi = 1$ but it can be done in the case (19.7.16). In fact these obstructions just coincide with the condition that the hypermultiplet potential has flat directions. The resulting Calabi–Yau manifold has just the Hodge numbers that would be deduced from the low energy field theory, which are

$$(h^{1,1}, h^{2,1}) = (2, 86) \tag{19.7.21}$$

in the present case. Thus the condensate of D-branes has a classical interpretation in terms of a change of the topology of the manifold.

This change of topology is more radical than the flop, in that the Hodge numbers change, and in particular the Euler number $\chi = 2(h^{1,1} - h^{2,1})$ changes. This is another example of the phenomenon, illustrated in figure 14.4, that the more we understand string dynamics the more we find that all theories and vacua are connected to one another. It appears that all Calabi–Yau vacua may be connected by conifold transitions.

The conifold transition is also more radical in that it is nonperturbative while the flop occurs in CFT, at string tree level. In fact the Euler number cannot change in CFT. One way to see this is by considering the dynamics of type II strings on the Calabi–Yau manifold. To have a potential that can give mass to some moduli, we need charged matter as above. However, the low energy gauge fields are all from the R–R sector and do not couple to ordinary strings. We can also see it by putting the heterotic string on the same space. At tree level the only way generations and antigenerations could become massive is in pairs, through a coupling $\mathbf{1} \cdot \mathbf{27} \cdot \overline{\mathbf{27}}$ when a singlet acquires an expectation value. This leaves the Euler number unchanged.

For a different string theory on the same Calabi–Yau manifold, the nonperturbative physics will be different. For the IIB theory there are no 3-branes that could become massless at the conifold point. Also in this case the complex structure moduli are in hypermultiplets, so the low energy effective action can receive string corrections. It is then possible and in fact likely that these corrections remove the singularity present in the tree-level action. This is similar to the way that world-sheet instantons remove the singularity at the edge of the Kähler cone in the flop transition. On the other hand, mirror symmetry relates the conifold singularity in complex structure moduli space to a singularity in the Kähler moduli space of the mirror. The IIB theory has the same behavior at this singularity as the IIA theory at the singularity in complex structure moduli space.

The two heterotic theories and the type I theory on a Calabi–Yau manifold have only $d = 4$, $N = 1$ supersymmetry, so there is less control over their nonperturbative behavior. One might think that the argument above about generations and antigenerations becoming massive in pairs would exclude any Euler number changing conifold transition in these cases. However, one of the things that has been learned from the recent study of nonperturbative dynamics in field and string theory is that at a nontrivial fixed point (meaning that the interactions remain nontrivial to arbitrarily long distances) one can have phase transitions that cannot be described by any classical Lagrangian. We will illustrate one such transition in the next section. There is no physical principle (such as an index theorem) to exclude the possibility that as one passes through such a fixed point to a new branch of moduli space, unpaired generations become massive due to strong interaction effects. It has been argued that this does actually occur, though in a somewhat different situation.

19.8 String theories on K3

A Calabi–Yau manifold of $2n$ real dimensions has $SU(n)$ holonomy. The number of six-dimensional Calabi–Yau manifolds is large, but in the discussion of mirror symmetry we saw that there is a unique two-dimensional example T^2. In four dimensions there are exactly two Calabi–Yau manifolds, the flat T^4 and the manifold K3, which has nontrivial $SU(2)$ holonomy. Compactification on K3 down to six noncompact dimensions is of interest for a number of reasons. The resulting six-dimensional theories have interesting dynamics but are highly constrained by Lorentz invariance and supersymmetry. Also, compactification on K3 often appears as an intermediate step to a four-dimensional theory, where the compact space is locally the product of K3 and a 2-manifold.

Compactification on K3 breaks half of the supersymmetry of the original theory. Under $SO(9,1) \to SO(5,1) \times SO(4)$, the ten-dimensional spinors decompose

$$\mathbf{16} \to (\mathbf{4}, \mathbf{2}) + (\mathbf{4}', \mathbf{2}') \,, \tag{19.8.1a}$$

$$\mathbf{16}' \to (\mathbf{4}, \mathbf{2}') + (\mathbf{4}', \mathbf{2}) \,. \tag{19.8.1b}$$

Under $SO(4) \to SU(2) \times SU(2)$, the $\mathbf{2}$ transforms under the first $SU(2)$ and the $\mathbf{2}'$ under the second, so if the holonomy lies in the first $SU(2)$ then a constant $\mathbf{2}'$ spinor is also covariantly constant and gives rise to an unbroken supersymmetry. The smallest $d = 6$ supersymmetry algebra (reviewed in section B.7) has eight supercharges, so each ten-dimensional supersymmetry gives rise to one six-dimensional supersymmetry. The decompositions (19.8.1) determine the chiralities: the IIA theory on K3 has nonchiral $d = 6$ (1,1) supersymmetry, the IIB theory has chiral (2,0) supersymmetry, and the heterotic or type I theory has (1,0) supersymmetry.

The Hodge diamond of K3 is

$$\begin{array}{ccccc} & & h^{2,2} & & \\ & h^{2,1} & & h^{1,2} & \\ h^{2,0} & & h^{1,1} & & h^{0,2} \\ & h^{1,0} & & h^{0,1} & \\ & & h^{0,0} & & \end{array} = \begin{array}{ccccc} & & 1 & & \\ & 0 & & 0 & \\ 1 & & 20 & & 1 \\ & 0 & & 0 & \\ & & 1 & & \end{array} . \tag{19.8.2}$$

In four spatial dimensions, the Poincaré dual squares to one, $** = 1$, so we can define self-dual or anti-self-dual 2-forms,

$$* \omega_2 = \pm \omega_2 . \tag{19.8.3}$$

On K3, 19 of the (1,1)-forms are self-dual, and the remaining (1,1)-form and the (2,0) and (0,2)-forms are anti-self-dual.

For the IIA string, the fluctuations without internal indices are $g_{\mu\nu}$, $b_{\mu\nu}$, ϕ, c_μ, and $c_{\mu\nu\rho}$, the last being related by Poincaré duality to a second vector c'_μ. Each (1,1)-form gives rise to a Kähler modulus $g_{i\bar{j}}$ and an axion $b_{i\bar{j}}$. An additional scalar arises from each of b_{ij} and $b_{\bar{i}\bar{j}}$. The complex structure moduli arise from (1,1)-forms by using the (2,0)-form $\Omega_{2\beta}$, in parallel to their connection with (2,1)-forms in four-dimensional theories:

$$g_{ij} = \Omega_{[i}{}^{\bar{k}} \omega_{j]\bar{k}} , \tag{19.8.4}$$

and similarly for $g_{\bar{i}\bar{j}}$. This vanishes when ω_2 is the Kähler form so there are a total of $19 + 19 = 38$ complex structure moduli. The total number of moduli for the K3 surface is then 80, of which 58 parameterize the metric and 22 the antisymmetric tensor background. Finally, $c_{\mu np}$ gives 22 vectors, one for each 2-form. In all the spectrum consists of the (1,1) supergravity multiplet (B.6.7) and 20 vector multiplets (B.6.8).

For the IIB string, there are the same NS–NS fluctuations $g_{\mu\nu}$, $b_{\mu\nu}$, ϕ, g_{mn} and b_{mn}. There is also an R–R scalar c, another from the dual of $c_{\mu\nu\rho\sigma}$, and another antisymmetric tensor $c_{\mu\nu}$. The components c_{mn} give an additional 22 scalars from the 2-forms, while $c_{\mu\nu pq}$ give 22 tensors. For the latter we must be careful about the duality properties. The ten-dimensional field strength is

$$H_{\mu\nu\sigma pq} = H_{\mu\nu\sigma} \omega_{pq} . \tag{19.8.5}$$

The ten-dimensional $*$ factorizes

$$*_{10} = *_4 *_6 . \tag{19.8.6}$$

Since the ten-dimensional field strength is self-dual in the IIB string, the four-dimensional field strength transforms in the same way as the internal form ω_2. The tensors $b_{\mu\nu}$ and $c_{\mu\nu}$ have both self-dual and anti-self-dual parts, so the total spectrum contains 21 self-dual tensors and 5 anti-self-dual tensors. The bosonic fields add up to the (2,0) supergravity multiplet (B.6.10) and 21 tensor multiplets (B.6.11).

It is interesting that the properties of the cohomology of K3 can be deduced entirely from physical considerations. The (2,0) supergravity theory is chiral and so potentially anomalous. We leave the discussion of anomalies in six dimensions to the references, but the result is that the anomaly from the supergravity multiplet can only be canceled if there are exactly 21 tensor multiplets. This determines the cohomology, and so indirectly the spectrum of the nonchiral IIA theory on the same manifold.

Before going on to the heterotic string, let us note some further properties of K3 and the associated CFT. First, there are various orbifold limits. Two were developed in the exercises to chapter 16, namely $T^4/\mathbf{Z}_2$ and $T^4/\mathbf{Z}_3$. The spectra of the type II theories on each of these orbifolds are the same as those that we have just found. Second, the manifold K3 is hyper-Kähler. In section B.7 hyper-Kähler geometry is defined in the context of field space, but the idea also applies to the spacetime geometry: the holonomy $SU(2) \subset SO(4)$ is the case $m = 1$ of the discussion in the appendix. In fact, spacetime and moduli space are not so distinct. Consider a Dp-brane for $p < 5$, oriented so that it is extended in the noncompact directions and at a point in the K3. We leave it to the reader to show that this breaks half the supersymmetries of the type II theory on K3, leaving eight unbroken. The four collective coordinates for the motion of the Dp-brane within K3 lie in a hypermultiplet and so their moduli space geometry is hyper-Kähler. However, the moduli space of the collective coordinates is just the space in which the Dp-brane moves, K3. This is an elementary example of a very fruitful idea, the interrelation between spacetime geometry and the moduli spaces of quantum field theories on branes. Third, the 80-dimensional moduli space of the NS–NS fields on K3 is guaranteed by supersymmetry to be of the form (B.6.1), namely

$$\frac{SO(20,4,\mathbf{R})}{SO(20,\mathbf{R}) \times SO(4,\mathbf{R})} , \tag{19.8.7}$$

up to a right identification under some discrete T-duality group.

Finally, the CFT of the string on K3 has (4,4) world-sheet superconformal invariance. This is closely related to the condition for $d = 4$, $N = 2$ supersymmetry cited in section 18.5. In geometric terms it comes about as follows. The basic world-sheet supercurrent is

$$T_F = \psi_m \partial X^m = \psi_r e^r{}_m \partial X^m , \tag{19.8.8}$$

and similarly for right-movers. We have used the tetrad $e^r{}_m$ to convert the index on ψ to tangent space. This tangent space index transforms as a $\mathbf{4} = (\mathbf{2},\mathbf{2})$ of $SO(4) = SU(2) \times SU(2)$. The curvature of K3 lies entirely within the first $SU(2)$, so rotations of ψ_r in the second $SU(2)$ leave the action

invariant. However, they do not leave the supercurrent invariant, and the three infinitesimal $SU(2)$ rotations generate three additional conserved supercurrents.

For the heterotic string we need to specify the gauge background. We start by embedding the spin connection in the gauge connection. This breaks the gauge symmetry to $E_7 \times E_8$ or $SO(28) \times SU(2)$. The bosonic spectrum includes the same states $g_{\mu\nu}$, $b_{\mu\nu}$, ϕ, g_{mn}, and b_{mn} found in the NS–NS spectrum of the type II theories. These comprise the bosonic content of the $d = 6$, $N = 1$ supergravity multiplet, one tensor multiplet, and 20 hypermultiplets. In addition there are vector multiplets in the adjoint of the gauge group. Finally there are additional hypermultiplets not related to the cohomology, which come from varying the gauge connection so that it is no longer equal to the spin connection. For the $E_8 \times E_8$ theory these hypermultiplets lie in the representations

$$(\mathbf{56},\mathbf{1})^{10} + (\mathbf{1},\mathbf{1})^{65} \tag{19.8.9}$$

of $E_7 \times E_8$. For the $SO(32)$ theory they lie in

$$(\mathbf{28},\mathbf{2})^{10} + (\mathbf{1},\mathbf{1})^{65} \tag{19.8.10}$$

of $SO(28) \times SU(2)$. Let us mention another result from the analysis of anomalies. A necessary condition for anomaly cancellation is that the numbers of hyper, tensor, and vector multiplets satisfy

$$n_H + 29n_T - n_V = 273\ . \tag{19.8.11}$$

In both of the present theories this is $625 + 29 - 381 = 273$. The full story of anomaly cancellation is more involved, because of the possibility of multiple tensors, and is left to the references.

The potential for the charged hypermultiplets has flat directions, and there is a nice geometric description of the resulting moduli space. The conditions (17.1.12), namely $F_{ij} = F_{\bar{i}\bar{j}} = F_i^{\bar{i}} = 0$, translate for four compact dimensions into the statement that the field strength is self-dual,

$$F = *F\ . \tag{19.8.12}$$

This is the condition that defines instantons in Yang–Mills theory; K3 is a four-dimensional Euclidean manifold, which is the usual setting for Yang–Mills instantons. The integral of the Bianchi identity (17.1.13),

$$\int_{\mathrm{K3}} \mathrm{tr}(R_2 \wedge R_2) = \int_{\mathrm{K3}} \mathrm{Tr}_v(F_2 \wedge F_2)\ , \tag{19.8.13}$$

determines the instanton charge: the number works out to 24. Thus the moduli space parameterized by the charged hypermultiplets is the space of gauge fields of instanton number 24 on K3. Supersymmetry guarantees that these lowest order solutions are exact. One can think of the moduli as representing the sizes of the instantons, their positions on K3, and their orientations within the gauge group; these parameters are not completely independent for the different instantons because there are constraints in order for the gauge field to be globally well defined. With spin connection equal to gauge connection all instantons are in the same $SU(2)$ subgroup and the unbroken symmetry is rather large. Generically they have various gauge orientations and the unbroken symmetry is smaller, E_8 for the $E_8 \times E_8$ theory and $SO(8)$ for the $SO(32)$ theory. For the $E_8 \times E_8$ theory, the gauge field with spin connection equal to gauge connection lies entirely in one E_8. By varying the moduli one can break the first E_8 entirely but the gauge field in the second E_8 remains zero. This is because the instanton numbers in the respective groups start at $(n_1, n_2) = (24, 0)$ and cannot change continuously. There are other branches of moduli space with different values of (n_1, n_2) such that $n_1 + n_2 = 24$.

Finally, it is very interesting to consider what happens when one or more instantons shrink to zero size. Note that all of these instantons are 5-branes, in that they are localized on K3 but the fields are independent of the six noncompact dimensions. We have discussed small instantons for the type I string in section 14.3: a new $SU(2)$ gauge symmetry appears on the 5-brane. The type I theory is the dual of the $SO(32)$ heterotic theory so the same must happen in the latter case. The gauge symmetry in the core cannot change as we go from weak to strong coupling by varying the neutral dilaton. Independent of duality, some of the arguments that were used in the type I case to derive the existence of the $SU(2)$ gauge symmetry apply also in the heterotic case — the ones based on the instanton moduli space and on the need for complete hypermultiplet representations. The group grows to $Sp(m)$ for m coincident zero-size instantons.

For the $E_8 \times E_8$ theory the same analysis leads to a very different result. To understand what happens, let us remember that the $E_8 \times E_8$ heterotic string is M-theory compactified on a segment of length $\alpha'^{1/2}g$. The eleven-dimensional spacetime is bounded by two ten-dimensional walls, with one E_8 living in each wall. The claim is that when an instanton in one of the walls shrinks to zero size, it can detach from the wall and move into the eleven-dimensional bulk. It remains extended in the noncompact directions so must be some 5-brane; there is a natural candidate, the 5-brane of M-theory discussed in section 14.4.

We have seen a similar phenomenon in section 13.6, where an instanton constructed from the gauge fields on a D4-brane could be contracted to a point and then detached from the D4-brane as a D0-brane. In fact, the present situation is dual to this, as shown in figure 19.2. If we compactify one of the noncompact dimensions with a small radius and regard this as the eleventh direction, we get the IIA string compactified to five dimensions on $\mathrm{K3} \times S_1/\mathbf{Z}_2$. The gauge fields live on D8-branes, and the instanton detaches as a D4-brane; this is T-dual to the D4–D0 system.

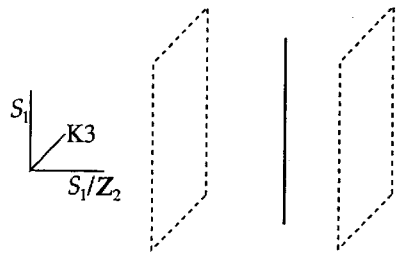


Fig. 19.2. A schematic picture of M-theory on K3 × $(S_1/\mathbf{Z}_2)$ × S_1. The 4 + 1 noncompact dimensions are suppressed, and K3 is represented by a single dimension. An M5-brane, extended in the noncompact and S_1 directions, is shown. When the S_1 is small, this is the IIA theory with a D4-brane. When the $S_1/\mathbf{Z}_2$ is small it is the $E_8 \times E_8$ heterotic string with a detached M5-brane. The M5-brane can move to either boundary and become an instanton in one of the E_8s.

In all, there can be some number n_5 of M5-branes, and this and the instanton numbers now satisfy

$$n_1 + n_2 + n_5 = 24 \,. \tag{19.8.14}$$

The different (n_1, n_2) moduli spaces discussed above are now connected, as an instanton can detach from one wall, move across the bulk, and attach to the other. In chapter 14 we argued that the world-volume of the M5-brane includes a massless tensor and five scalars. Here the M5-brane is extended in the noncompact dimensions, so these become massless fields in the six-dimensional low energy field theory. Four of the scalars, forming a hypermultiplet, represent the position of the brane within K3. The fifth scalar, in a tensor multiplet, represents the position in the $S_1/\mathbf{Z}_2$ direction. The total number of tensor multiplets is $n_T = n_5 + 1$.

The instanton and M5-brane branches meet at a point, and the nature of the transition is quite interesting. As the vacuum moves onto the M5-brane branch, the number n_T of tensor multiplets increases by one. The anomaly cancellation condition $n_H + 29n_T - n_V = 273$ requires a compensating change in the number of hyper or vector multiplets. Typically, the number of hypermultiplets associated with the gauge background decreases by 30 when the instanton number goes down by one, offsetting the contribution of the tensor and hypermultiplets on the M5-brane.

The ordinary Higgs mechanism preserves the anomaly cancellation by giving mass to a vector and hypermultiplet. For the Higgs mechanism there is a familiar classical Lagrangian description. There is no classical Lagrangian that exhibits this new phase transition where a tensor multiplet becomes massless and a net of 29 hypermultiplets massive, or the reverse. In this respect it is like the generation-changing transition discussed in the previous section, and so would until recently have been considered impossible. We now understand that such transitions can occur at nontrivial fixed points. In fact, the Δn_T transition point is similar to the tensionless string theory that arises on coincident $M5$-branes. We have not discussed in detail the boundary conditions on the ends of the $S_1/\mathbf{Z}_2$, but an M2-brane can end on them, as well as on an M5-brane as before (using duality, the reader can derive this fact in various ways). An M2-brane stretched between an M5-brane and the wall is a string with tension proportional to the separation, becoming tensionless when the M5-brane reaches the wall.

19.9 String duality below ten dimensions

In chapter 14 we focused on the nonperturbative dynamics of string theories in ten dimensions, and in a few toroidal compactifications. In this chapter we have seen some further phenomena that arise in compactified theories, in particular the conifold transition and the instanton/5-brane transition.

We should emphasize that many things that are impossible in CFT (string tree level) can happen nonperturbatively. One is the conifold transition itself, as we have explained. Another is heterotic string theory with $n_T > 1$, which we have just found. To get a massless tensor from a perturbative string state requires exciting a right-moving vector oscillator and a left-moving vector oscillator, and there is exactly one way to do this. The vacua with $n_5 > 0$ then do not have a perturbative string description. A third concerns the maximum rank of the gauge group in the heterotic string. Focusing on the maximal commuting subgroup $U(1)^r$, each $U(1)$ contributes 1 to the central charge, or $\frac{3}{2}$ for a right-mover, for a maximum of $r = 16 + 2k$, where k is the number of compactified dimensions. On the other hand, the $SO(32)$ theory in the limit that all instantons are pointlike has gauge group $SO(32) \times Sp(1)^{24}$, or as large as $SO(32) \times Sp(24)$ if the instantons are coincident. Each of these has rank 40, exceeding the 24 allowed in CFT. A fourth is the no-go theorem for the Standard Model in type II theory. This was proved in section 18.2, but the possibility of nonperturbative breakdown was also discussed.

This does not mean that the various results obtained in CFT are valueless. First, an understanding of the tree-level spectrum is a necessary step toward determining the nonperturbative dynamics. Second, Calabi–Yau compactification of the weakly coupled $E_8 \times E_8$ string resembles

the grand unified Standard Model sufficiently closely to suggest that our vacuum may be of this type, at least approximately.

The conifold and small instanton transitions both occurred in theories with eight supersymmetries; such theories can have rich dynamics. Presumably theories with four and fewer supersymmetries have dynamics that is at least as rich, though the understanding of these is less complete. However, even with 16 supersymmetries there are some important phenomena that come with compactification. In particular, the toroidal compactifications of the heterotic string have this supersymmetry, and these are the main subject of this final section.

Heterotic strings in $7 \leq d \leq 9$

We would like to determine the strong-coupling behavior of the heterotic string compactified on T^k. The answer would seem to be obvious, because we know the duals in ten dimensions and we can just compactify these. To see what the issue is, recall the $SO(32)$ heterotic–type I relations (14.3.4),

$$G_{\mathrm{I}\mu\nu} = g_{\mathrm{h}}^{-1} G_{\mathrm{h}\mu\nu} , \tag{19.9.1a}$$

$$g_{\mathrm{I}} = g_{\mathrm{h}}^{-1} . \tag{19.9.1b}$$

This symmetry acts locally on the fields, and so should take a given spacetime into the same spacetime in the dual theory. However, the metric is rescaled; therefore, for toroidal compactification, the radii are rescaled

$$R_{m\mathrm{I}} \propto g_{\mathrm{h}}^{-1/2} R_{m\mathrm{h}} . \tag{19.9.2}$$

As the heterotic coupling becomes large the k-torus in the type I theory becomes small. As usual, we seek a description where the compact manifold is fixed in size or large, because g is not an accurate measure of the effective coupling with a very small compact manifold. Thus we will follow a succession of dualities, as we did in section 14.5 in deducing the dual of the $E_8 \times E_8$ heterotic string.

The obvious next step is T-duality. This gives

$$g' \propto V_{\mathrm{I}}^{-1} g_{\mathrm{I}} \propto V_{\mathrm{h}}^{-1} g_{\mathrm{h}}^{(k-2)/2} , \tag{19.9.3a}$$

$$R'_m \propto R_{m\mathrm{I}}^{-1} \propto g_{\mathrm{h}}^{1/2} R_{m\mathrm{h}}^{-1} . \tag{19.9.3b}$$

We have defined the volume $V = \prod_{m=10-k}^{9} R_m$ in each theory. The compact space is now an orientifold as discussed in section 13.2,

$$T^k/\mathbf{Z}_2 , \quad \mathbf{Z}_2 = \{1, \Omega\hat{\beta}\} . \tag{19.9.4}$$

Here $\hat{\beta}$ is essentially a reflection in the compact directions, to be studied in more detail below.

At strong heterotic coupling the compact space is now large, while the coupling is proportional to $g_{\mathrm{h}}^{(k-2)/2}$. For $k = 1$ the theory that we have arrived at is weakly coupled, but even here there is a subtlety. If we begin with a compactification that has vanishing Wilson lines, we know from the discussion in chapter 8 that in the T-dual theory the 16 D$(9-k)$-branes will be at a single fixed point. The R–R and dilaton charges of the fixed points and D-branes cancel globally, but not locally. The dilaton, and therefore the effective coupling, is position-dependent. It diverges at the fixed points without D-branes when $k \geq 2$, and even for $k = 1$ it will diverge if the dual spacetime is too large. To keep things simple we will always start with a configuration of Wilson lines such that the D-branes are distributed equally among the fixed points. The number of fixed points is 2^k, so that by using half-D-branes we can do this for k as large as 5.

For $k = 2$ the coupling in the dual theory is V_{h}^{-1}. If the original 2-torus is larger than the string scale then we have reached a weakly coupled description, and if it is smaller then we simply start with an additional T-duality. If it is of order the string scale then the coupling g' is of order 1 and this is the simplest description that we can reach.

For $k = 3$ the coupling g' is strong, suggesting a further weak–strong duality. The bulk physics for k odd is that of the IIA theory, so strong coupling gives an eleven-dimensional theory. The necessary transformations (12.1.9) were obtained from the dimensional reduction of $d = 11$ supergravity, giving

$$R_{10\mathrm{M}} \propto g'^{2/3} \propto g_{\mathrm{h}}^{1/3} V_{\mathrm{h}}^{-2/3} , \tag{19.9.5a}$$

$$R_{m\mathrm{M}} \propto g'^{-1/3} R'_m \propto g_{\mathrm{h}}^{1/3} R_{\mathrm{h}}^{-1} V_{\mathrm{h}}^{1/3} . \tag{19.9.5b}$$

All the radii grow with g_{h}, so the strongly coupled theory is eleven-dimensional. We will make some further remarks about the $k = 3$ case after the discussion of $k = 4$.

Heterotic–type IIA duality in six dimensions

The case $k = 4$ is interesting for a number of reasons, and we will discuss it in some detail. The description (19.9.3) is strongly coupled and the bulk physics is described by the IIB string, so we make a further IIB weak–strong transformation to obtain

$$g'' \propto g'^{-1} \propto g_{\mathrm{h}}^{-1} V_{\mathrm{h}} , \tag{19.9.6a}$$

$$R''_m \propto g'^{-1/2} R'_m \propto R_{m\mathrm{h}}^{-1} V_{\mathrm{h}}^{1/2} . \tag{19.9.6b}$$

We now have a weakly coupled description on a space of fixed volume as g_{h} becomes large.

To be more precise about the nature of the dual theory we must determine the $\mathbf{Z}_2$ identification. This is related to the Ω of the type I theory by a T-duality and then a IIB S-duality. The T-duality is a redefinition by β_R, a reflection in the compact directions acting only on the right-movers. Then $\Omega\hat{\beta}$ is the image of Ω under this,

$$\Omega\hat{\beta} = \beta_R^{-1}\Omega\beta_R = \Omega\beta_L^{-1}\beta_R \,. \tag{19.9.7}$$

For $k = 2n$ even, β has a convenient definition $\exp(\pi i J)$ as a rotation by π in n planes. Then

$$\hat{\beta} = \beta_L^{-1}\beta_R = \beta_L^{-2}\beta_L\beta_R = \exp(-2\pi i J_L)\beta = \exp(\pi i n \mathbf{F}_L)\beta \,. \tag{19.9.8}$$

In other words, this differs from the simple parity operation β by an extra $(-1)^n$ in the left-moving R sector. For T^4 this is simply β, and the $\mathbf{Z}_2$ is $\Omega\beta$. We must now consider the effect of the IIB weak–strong duality. The image of β is β, because duality commutes with the Lorentz group. To determine the image of Ω let us note its effect on the massless fields of the IIB theory, as discussed in section 10.6,

$$G_{\mu\nu}+ \,, \quad B_{\mu\nu}- \,, \quad \Phi+ \,, \quad C- \,, \quad C_{\mu\nu}+ \,, \quad C_{\mu\nu\rho\sigma}- \,. \tag{19.9.9}$$

The weak–strong duality interchanges $B_{\mu\nu}$ and $C_{\mu\nu}$ and inverts $e^{-\Phi} + iC$. Conjugating the operation (19.9.9) by this results in

$$G_{\mu\nu}+ \,, \quad B_{\mu\nu}+ \,, \quad \Phi+ \,, \quad C- \,, \quad C_{\mu\nu}- \,, \quad C_{\mu\nu\rho\sigma}- \,. \tag{19.9.10}$$

This acts as $+1$ on NS–NS fields and -1 on R–R fields. This identifies it as $\exp(i\pi\mathbf{F}_L)$ (or $\exp(i\pi\mathbf{F}_R)$ — which one we choose is arbitrary). As another check, Ω commutes with one of the two supercharges (the sum of the left- and right-movers), as does $\exp(\pi i\mathbf{F}_L)$ (the supercharge in the NS–R sector).

Thus our dual to the heterotic theory on T^4 is the IIB theory on

$$T^k/\mathbf{Z}_2 \,, \quad \mathbf{Z}_2 = \{1, \exp(\pi i \mathbf{F}_L)\beta\}. \tag{19.9.11}$$

We can bring this to a more familiar form by a further T-duality transformation on a single coordinate, say X^9; since the radii are independent of the heterotic coupling this still defines a good dual. This gives the IIA theory with

$$g_{\mathrm{A}} = g''R_9'^{-1} = g_{\mathrm{h}}^{-1}R_{9\mathrm{h}}V_{\mathrm{h}}^{1/2} \,, \tag{19.9.12a}$$

$$R_{9\mathrm{A}} = R_{9\mathrm{B}'}^{-1} = V_{\mathrm{h}}^{-1/2}R_{9\mathrm{h}} \,, \tag{19.9.12b}$$

$$R_{m\mathrm{A}} = R_{m\mathrm{B}'} = V_{\mathrm{h}}^{1/2}R_{m\mathrm{h}}^{-1} \,, \quad m = 6,7,8 \,. \tag{19.9.12c}$$

The T-duality adds or deletes a 9-index on each R–R field so that β will act with the opposite sign. This cancels the action of $\exp(\pi i\mathbf{F}_L)$, so the IIA image of the space (19.9.11) is the ordinary orbifold

$$T^4/\mathbf{Z}_2 \,, \quad \mathbf{Z}_2 = \{1, \beta\} \,. \tag{19.9.13}$$

This orbifold is a special case of K3.

Thus the dual of the heterotic string on T^4 is the IIA string on K3; this is often termed *string–string duality*. In particular we have found that a special configuration of Wilson lines in the heterotic theory maps to an orbifold K3, but since the duality at this point implies an isomorphism between the respective moduli we can in the usual way extend this to the full moduli space. Indeed, the moduli space (19.8.7) of the IIA string on K3 is identical to the Narain moduli space (11.6.14) of the heterotic string. The coset structure is just a consequence of the 16 supercharges, but the number 20 in each case is a nontrivial check. Also, a careful analysis of the discrete T-duality of the K3 CFT has shown that it is identical to that of the heterotic theory on T^4.

In perturbation theory the gauge group of the IIA string on K3 is $U(1)^{24}$. This is also the gauge group at generic points in heterotic moduli space, only the currents ∂X and $\tilde{\psi}$ surviving. At special points non-Abelian symmetries appear, the low energy physics being the usual Higgs mechanism. These same symmetries must appear on the IIA side. The $U(1)$s all come from the R–R sector, so the charged gauge bosons must arise from D-branes. In particular, the gauge fields associated with 2-forms couple to D2-branes wrapped around the corresponding 2-cycles. These must become massless at the enhanced symmetry points, and we know from the conifold example that this can occur if one or more 2-cycles shrinks to zero size. Indeed, the possible singularities of K3 are known to have an A–D–E classification, meaning that they are associated with the Dynkin diagrams of the simply-laced Lie algebras. At such a singularity the charges of the massless D-brane states are the roots of the associated algebra. Thus nonperturbative string theory provides a connection between the A–D–E classification of singularities and the corresponding algebra.

A single collapsed 2-sphere gives a $\mathbf{Z}_2$ orbifold singularity. The orbifold CFT is solvable and nonsingular. One expects that if a CFT is nonsingular then string perturbation theory should be a good description at weak coupling, meaning that there should not be massless nonperturbative states. This seems to contradict the argument that the collapsed 2-sphere gives rise to massless wrapped D-brane states. In fact, the massless D-brane should appear only when both the real and imaginary parts of the Kähler modulus $T = v + ib$ for the 2-sphere vanish. A careful analysis shows that the solvable theory is the orbifold limit with $T = i\pi$. The modulus b is a twisted state in the orbifold theory, so to reach the point of enhanced symmetry one must turn on a twisted state background and the CFT is no longer solvable.

Define the six-dimensional dilaton by

$$e^{-2\Phi_6} = Ve^{-2\Phi} . \tag{19.9.14}$$

Tracing through the various dualities, the map between heterotic and IIA fields is

$$\Phi_6 \to -\Phi_6 , \quad G_{\mu\nu} \to e^{-2\Phi_6} G_{\mu\nu} , \tag{19.9.15a}$$

$$\tilde{H}_3 \to e^{-2\Phi_6} *_6 \tilde{H}_3 , \quad F_2^a \to F_2^a . \tag{19.9.15b}$$

The transformation takes the same form in both directions, heterotic → IIA and IIA → heterotic. The tensors and forms (19.9.15) are all in the noncompact directions. In the special case of a $\mathbf{Z}_2$ orbifold, the mapping of the moduli is given in eq. (19.9.12). The dimensionally reduced six-dimensional action for the fields (19.9.15) in the heterotic string is

$$S_{\text{het}} = \frac{1}{2\kappa_6^2} \int d^6x (-G_6)^{1/2} e^{-2\Phi_6} \Big(R + 4\partial_\mu \Phi_6 \partial^\mu \Phi_6 - \frac{1}{2}|\tilde{H}_3|^2 - \frac{\kappa_6^2}{2g_6^2}|F_2|^2 \Big) . \tag{19.9.16}$$

The same action for the IIA theory is

$$S_{\text{IIA}} = \frac{1}{2\kappa_6^2} \int d^6x (-G_6)^{1/2} \Big(e^{-2\Phi_6} R + 4e^{-2\Phi_6} \partial_\mu \Phi_6 \partial^\mu \Phi_6 - \frac{1}{2}|\tilde{H}_3|^2 - \frac{\kappa_6^2}{2g_6^2} e^{-2\Phi_6} |F_2|^2 \Big) . \tag{19.9.17}$$

We have omitted the kinetic terms for the moduli and the dependence of g_6 on the moduli; it is left to the reader to include these. The transformation (19.9.15) converts one theory to the other.

We should mention that the strategy that we used to find the dual of the ten-dimensional type I and IIB theories, following the D-string to strong coupling, was first applied to the six-dimensional heterotic–IIA duality. Consider the IIA NS5-brane, with four of its dimensions wrapped around K3. This is extended in one noncompact direction, and so is a string. A study of its fluctuations shows that they are the same as those of a heterotic string. The ratio of the tensions of the solitonic and fundamental strings is g^{-2}, as compared to the g^{-1} of the D-string. This again becomes small at strong coupling, so we can make the same duality argument as for the D-string. Similarly the fluctuations of the heterotic NS5-brane wrapped on T^4 are the same as those of the fundamental IIA string, and so this argument yields an element of the U-duality group.

Let us return to the case $k = 3$. To deduce the spacetime geometry, we need to understand how the $\mathbf{Z}_2$ identification acts on the M-theory circle. Again the $\mathbf{Z}_2$ arises via T-duality from the Ω projection of the type I theory. Recall from section 10.6 that the field $C_{\mu 789}$ is odd under the latter. In the T-dual description the field C_μ is then odd. Since this couples to 10-momentum, it must be that the $\mathbf{Z}_2$ reflects the M-theory circle as well as the original T^3. Thus, the $d = 7$ heterotic string is dual to M-theory on

$$T^4/\mathbf{Z}_2 = \text{K3}. \tag{19.9.18}$$

Recall from the Narain description that the moduli space of this heterotic compactification, including the dilaton, is locally

$$\frac{SO(19,3,\mathbf{R})}{SO(19,\mathbf{R}) \times SO(3,\mathbf{R})} \times \mathbf{R}^+ . \tag{19.9.19}$$

This 58-parameter space is identical to the space of metrics on K3. This is different from string theory on K3: M-theory has no 2-form field, so there are fewer moduli. The enhanced gauge symmetries on the heterotic side come from M2-branes wrapped on collapsed cycles of the K3.

Heterotic S-duality in four dimensions

The six-dimensional duality that we have just found can be used to find duals of four-dimensional theories. Let us consider the most supersymmetric case, compactification on a further 2-torus, to give the heterotic string on T^6 and the IIA string on $T^2 \times K3$. The four-dimensional dilaton

$$e^{-2\Phi_4} = R_5 R_6 e^{-2\Phi_6} \tag{19.9.20}$$

transforms as

$$e^{\Phi_4} \to (R_5 R_6)^{-1/2}, \tag{19.9.21}$$

where again the transformation is the same in both directions. The 3-form field strength transforms as

$$*_4 \tilde{H} \to e^{-2\Phi_4} dB_{56} , \tag{19.9.22}$$

but also in each theory this field strength is related to the axion by

$$*_4 \tilde{H} \propto e^{-2\Phi_4} da . \tag{19.9.23}$$

It follows that the dilaton–axion field $S = e^{-2\Phi_4} + ia$ is related to the scalar $\rho \sim B_{56} + iR_5 R_6$ introduced in section 8.4 by

$$S \to i\rho^* . \tag{19.9.24}$$

From this we learn something interesting. The T-duality in the (5,6)-directions acts by the usual $SL(2,\mathbf{Z})$ transformation on ρ in each theory. It follows from the duality (19.9.24) that in each theory there is also an $SL(2,\mathbf{Z})$ acting on S (and hence called S-duality). This includes a weak–strong duality $S \to 1/S$, as well as discrete shifts of the axion. Thus we

have deduced the strong-coupling dual of the heterotic string on T^4: itself. We see that the heterotic string compactified on tori has a complicated but consistent pattern of duals in different dimensions. In no case does one find two *different* weakly coupled duals of the same theory: that would be a contradiction.

In the heterotic theory on T^6, the interactions at energies far below the Planck scale reduce to $d = 4$, $N = 4$ gauge theory, and the $SL(2,\mathbf{Z})$ reduces to the Montonen–Olive symmetry of gauge theory, discussed in section 14.1. In both theories the full moduli space is

$$\frac{SU(1,1)}{SU(1,1,\mathbf{Z})} \times \frac{O(22,6,\mathbf{R})}{O(22,\mathbf{R}) \times O(6,\mathbf{R}) \times O(22,6,\mathbf{Z})} . \tag{19.9.25}$$

In the heterotic string the first factor is from the dilaton superfield and the second from the moduli of Narain compactification. As is usually the case, the integer subgroup of the symmetry in the numerator of each factor is a symmetry of the full theory. In the IIA theory the first factor is from the ρ field, while the dilaton–axion field, the K3 moduli, and additional moduli from the T^2 compactification combine to give a single coset. The $O(22,6,\mathbf{Z})$ duality then includes the perturbative duality of the K3, the S-duality of the dilaton–axion field, and U-dualities that mix these.

The six-dimensional duality is also useful in constructing dual pairs with less supersymmetry. Many Calabi–Yau manifolds are *K3 fibrations*, locally a product of K3 with a two-dimensional manifold. Applying the heterotic–IIA duality locally, the IIA theory on such a space is dual to the heterotic string on the corresponding T^4 fibration. For heterotic string compactifications with $d = 4$, $N = 2$ supersymmetry, the dilaton is in a vector multiplet. To see this, note that the dilaton is obtained by exciting one left- and one right-moving oscillator and so is of the form $|1,-1\rangle$ or $|-1,1\rangle$, where the notation refers to the helicity s_1 carried on each side. Spacetime supersymmetry acts only on the right, generating a multiplet of four states. A helicity $\pm\frac{3}{2}$ on one side is not possible at the massless level, as the conformal weight would be at least $\frac{9}{8}$. The supermultiplet must then consist of

$$|1,-1\rangle\,, \quad |1,-\tfrac{1}{2}\rangle^2\,, \quad |1,0\rangle \tag{19.9.26}$$

and the CPT conjugates. This is the helicity content of a vector multiplet. It follows that the hypermultiplet moduli space does not have string loop or nonperturbative corrections in $d = 4$, $N = 2$ compactifications of the heterotic string, just as the vector multiplet moduli space does not have such corrections in the dual type II theory. This is analogous to the constraints from mirror symmetry, but for string rather than world-sheet corrections. In some cases one can combine mirror symmetry and string duality to determine the exact low energy action for a $d = 4$, $N = 2$ compactification. As with mirror symmetry, comparing the exact result in one theory with the loop and instanton corrections in its dual leads to unexpected mathematical connections.

19.10 Conclusion

Especially in this final chapter, we have only been able to scratch the surface of many important and beautiful ideas. String theory is a rich structure, whose full form is not yet understood. It is a mathematical structure, but deeply grounded in physics. It incorporates and unifies the central principles of physics: quantum mechanics, gauge symmetry, and general relativity, as well as anticipated new principles: supersymmetry, grand unification, and Kaluza–Klein theory. Undoubtedly there are many remarkable discoveries still to be made.

Exercises

19.1 Verify directly that the type II dilaton is in a hypermultiplet, by the method of eq. (19.9.26).

19.2 Fill in the details of the counting of $(a,\tilde{c})$ states in the Gepner models, as discussed below eq. (19.5.14).

19.3 Show explicitly that the net effect of the twist (19.6.2) on the spectrum is to reverse the sign of the left-moving $U(1)$ charge.

19.4 For compactification of the type I string on T^k for $k \leq 5$, give explicitly the Wilson line configuration such that in the T-dual theory there is an equal number of D-branes coincident with each orientifold fixed plane. What is the unbroken gauge group in each case?

19.5 By composing S, T, S, and T dualities as discussed in section 19.9, show that in both directions the string–string duality transformation takes the form (19.9.15). Show that this transforms the heterotic action into the IIA action. Find the action for the moduli R_m and show that it is invariant.

Appendix B

Spinors and supersymmetry in various dimensions

Results about spinors and supersymmetry in various spacetime dimensions are used throughout this volume. This appendix provides an introduction to these subjects. The appropriate sections of the appendix should be read as noted at various points in the text.

B.1 Spinors in various dimensions

We develop first the Dirac matrices, which represent the Clifford algebra

$$\{\Gamma^\mu, \Gamma^\nu\} = 2\eta^{\mu\nu} \,. \tag{B.1.1}$$

We then go on to representations of the Lorentz group. To be specific we will take signature $(d-1,1)$, so that $\eta^{\mu\nu} = \mathrm{diag}(-1,+1,\ldots,+1)$. The extension to signature $(d,0)$ (and to more than one timelike dimension) will be indicated later. Throughout this appendix the dimensionality of spacetime is denoted by d; we generally reserve D to designate the total spacetime dimensionality of a string theory.

We begin with an even dimension $d = 2k+2$. Group the Γ^μ into $k+1$ sets of anticommuting raising and lowering operators,

$$\Gamma^{0\pm} = \frac{1}{2}(\pm\Gamma^0 + \Gamma^1) \,, \tag{B.1.2a}$$

$$\Gamma^{a\pm} = \frac{1}{2}(\Gamma^{2a} \pm i\Gamma^{2a+1}) \,, \quad a = 1,\ldots,k \,. \tag{B.1.2b}$$

These satisfy

$$\{\Gamma^{a+}, \Gamma^{b-}\} = \delta^{ab} \,, \tag{B.1.3a}$$

$$\{\Gamma^{a+}, \Gamma^{b+}\} = \{\Gamma^{a-}, \Gamma^{b-}\} = 0 \,. \tag{B.1.3b}$$

In particular, $(\Gamma^{a+})^2 = (\Gamma^{a-})^2 = 0$. It follows that by acting repeatedly

with the Γ^{a-} we can reach a spinor annihilated by all the Γ^{a-},

$$\Gamma^{a-}\zeta = 0 \quad \text{for all } a \,. \tag{B.1.4}$$

Starting from ζ one obtains a representation of dimension 2^{k+1} by acting in all possible ways with the Γ^{a+}, at most once each. We will label these by with $\mathbf{s} \equiv (s_0, s_1, \ldots, s_k)$, where each of the s_a is $\pm\frac{1}{2}$:

$$\zeta^{(\mathbf{s})} \equiv (\Gamma^{k+})^{s_k+1/2} \ldots (\Gamma^{0+})^{s_0+1/2}\zeta \,. \tag{B.1.5}$$

In particular, the original ζ corresponds to all $s_a = -\frac{1}{2}$.

Taking the $\zeta^{(\mathbf{s})}$ as a basis, the matrix elements of Γ^μ can be derived from the definitions and the anticommutation relations. Increasing d by two doubles the size of the Dirac matrices, so we can give an iterative expression starting in $d = 2$, where

$$\Gamma^0 = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \,, \quad \Gamma^1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \,. \tag{B.1.6}$$

Then in $d = 2k+2$,

$$\Gamma^\mu = \gamma^\mu \otimes \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \,, \quad \mu = 0,\ldots,d-3 \,, \tag{B.1.7a}$$

$$\Gamma^{d-2} = I \otimes \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \,, \quad \Gamma^{d-1} = I \otimes \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix} \,, \tag{B.1.7b}$$

with γ^μ the $2^k \times 2^k$ Dirac matrices in $d-2$ dimensions and I the $2^k \times 2^k$ identity. The 2×2 matrices act on the index s_k, which is added in going from $2k$ to $2k+2$ dimensions.

The notation $\mathbf{s}$ reflects the Lorentz properties of the spinors. The Lorentz generators

$$\Sigma^{\mu\nu} = -\frac{i}{4}[\Gamma^\mu, \Gamma^\nu] \tag{B.1.8}$$

satisfy the $SO(d-1,1)$ algebra

$$i[\Sigma^{\mu\nu}, \Sigma^{\sigma\rho}] = \eta^{\nu\sigma}\Sigma^{\mu\rho} + \eta^{\mu\rho}\Sigma^{\nu\sigma} - \eta^{\nu\rho}\Sigma^{\mu\sigma} - \eta^{\mu\sigma}\Sigma^{\nu\rho} \,. \tag{B.1.9}$$

The generators $\Sigma^{2a,2a+1}$ commute and can be simultaneously diagonalized. In terms of the raising and lowering operators,

$$S_a \equiv i^{\delta_{a,0}}\Sigma^{2a,2a+1} = \Gamma^{a+}\Gamma^{a-} - \frac{1}{2} \tag{B.1.10}$$

so $\zeta^{(\mathbf{s})}$ is a simultaneous eigenstate of the S_a with eigenvalues s_a. The half-integer values show that this is a spinor representation. The spinors form the 2^{k+1}-dimensional *Dirac* representation of the Lorentz algebra $SO(2k+1,1)$.

The Dirac representation is reducible as a representation of the Lorentz algebra. Because $\Sigma^{\mu\nu}$ is quadratic in the Γ matrices, the $\zeta^{(\mathbf{s})}$ with even and

odd numbers of $+\frac{1}{2}$s do not mix. Define

$$\Gamma = i^{-k}\Gamma^0\Gamma^1\ldots\Gamma^{d-1}\,, \tag{B.1.11}$$

which has the properties

$$(\Gamma)^2 = 1\,, \quad \{\Gamma,\Gamma^\mu\} = 0\,, \quad [\Gamma,\Sigma^{\mu\nu}] = 0\,. \tag{B.1.12}$$

The eigenvalues of Γ are ± 1. The conventional notation for Γ in $d = 4$ is Γ_5, but this is inconvenient in general d. Noting that

$$\Gamma = 2^{k+1}S_0S_1\ldots S_k\,, \tag{B.1.13}$$

we see that $\Gamma_{ss'}$ is diagonal, taking the value $+1$ when the s_a include an even number of $-\frac{1}{2}$s and -1 for an odd number of $-\frac{1}{2}$s. The 2^k states with Γ eigenvalue (*chirality*) $+1$ form a *Weyl* representation of the Lorentz algebra, and the 2^k states with eigenvalue -1 form a second, inequivalent, Weyl representation. For $d = 4$, the Dirac representation is the familiar four-dimensional one, which separates into 2 two-dimensional Weyl representations,

$$\mathbf{4}_{\text{Dirac}} = \mathbf{2} + \mathbf{2}'\,. \tag{B.1.14}$$

Here we have used a common notation, labeling a representation by its dimension (in boldface). In $d = 10$ the representations are

$$\mathbf{32}_{\text{Dirac}} = \mathbf{16} + \mathbf{16}'\,. \tag{B.1.15}$$

For an odd dimension $d = 2k + 3$, simply add $\Gamma^d = \Gamma$ or $\Gamma^d = -\Gamma$ to the Γ matrices for $d = 2k + 2$. This is now an irreducible representation of the Lorentz algebra, because $\Sigma^{\mu d}$ anticommutes with Γ. Thus there is a single spinor representation of $SO(2k + 2, 1)$, which has dimension 2^{k+1}.

Majorana spinors

The above construction of the irreducible representation of the Γ matrices shows that in even dimensions $d = 2k + 2$ it is unique up to a change of basis. The matrices $\Gamma^{\mu *}$ and $-\Gamma^{\mu *}$ satisfy the same Clifford algebra as Γ^μ, and so must be related to Γ^μ by a similarity transformation. In the basis s, the matrix elements of $\Gamma^{a\pm}$ are real, so it follows from the definition (B.1.2) that $\Gamma^3, \Gamma^5, \ldots, \Gamma^{d-1}$ are imaginary and the remaining Γ^μ real. This is also consistent with the explicit expression (B.1.7). Defining

$$B_1 = \Gamma^3\Gamma^5\ldots\Gamma^{d-1}\,, \quad B_2 = \Gamma B_1\,, \tag{B.1.16}$$

one finds by anticommutation that

$$B_1\Gamma^\mu B_1^{-1} = (-1)^k\Gamma^{\mu *}\,, \quad B_2\Gamma^\mu B_2^{-1} = (-1)^{k+1}\Gamma^{\mu *}\,. \tag{B.1.17}$$

For either B_1 or B_2 (and only for these two matrices),

$$B\Sigma^{\mu\nu}B^{-1} = -\Sigma^{\mu\nu *}\,. \tag{B.1.18}$$

It follows from eq. (B.1.18) that the spinors ζ and $B^{-1}\zeta^*$ transform in the same way under the Lorentz group, so the Dirac representation is its own conjugate. Acting on the chirality matrix Γ, one finds

$$B_1\Gamma B_1^{-1} = B_2\Gamma B_2^{-1} = (-1)^k\Gamma^*\,, \tag{B.1.19}$$

so that either form for B will change the eigenvalue of Γ when k is odd and not when it is even. For k even ($d = 2$ mod 4) each Weyl representation is its own conjugate. For k odd ($d = 0$ mod 4) each Weyl representation is conjugate to the other. Thus in $d = 4$ we can designate the representations as $\mathbf{2}$ and $\bar{\mathbf{2}}$ rather than $\mathbf{2}$ and $\mathbf{2}'$, but in $d = 10$, only as $\mathbf{16}$ and $\mathbf{16}'$

Just as the gravitational and gauge fields are real, various spinor fields satisfy a *Majorana* condition, which relates ζ^* to ζ. This condition must be consistent with Lorentz transformations and so must have the form

$$\zeta^* = B\zeta \tag{B.1.20}$$

with B satisfying (B.1.18). Taking the conjugate gives $\zeta = B^*\zeta^* = B^*B\zeta$, so such a condition is consistent if and only if $B^*B = 1$. Using the reality and anticommutation properties of the Γ-matrices one finds

$$B_1^*B_1 = (-1)^{k(k+1)/2}\,, \quad B_2^*B_2 = (-1)^{k(k-1)/2}\,. \tag{B.1.21}$$

A Majorana condition using B_1 is therefore possible only if $k = 0$ or 3 (mod 4), and using B_2 only if $k = 0$ or 1 (mod 4). If $k = 0$ both conditions are possible but they are physically equivalent, being related by a similarity transformation.

A Majorana condition can be imposed on a Weyl spinor only if $B^*B = 1$ *and* the Weyl representation is conjugate to itself. For k odd, which is $d = 0$ or 4 (mod 8), it is therefore not possible to impose both the Majorana and Weyl conditions on a spinor: one can impose one or the other. Precisely for $k = 0$ mod 4, which is $d = 2$ (mod 8), a spinor can simultaneously satisfy the Majorana and Weyl conditions. Majorana–Weyl spinors in $d = 10$ play a key role in the spacetime theory of the superstring, and Majorana–Weyl spinors in $d = 2$ (ψ^μ and $\tilde{\psi}^\mu$) play a key role on the world-sheet.

Extending to odd dimensions, $\Gamma^d = \pm\Gamma$, and so the conjugation (B.1.19) of Γ^d is compatible with the conjugation (B.1.17) of the other Γ^μ only for B_1, so that $k = 0$ or 3 (mod 4). In all, a Majorana condition is possible if $d = 0$, 1, 2, 3, or 4 (mod 8). When the Majorana condition is allowed, there is a basis in which B is either 1 or Γ and so commutes with all the $\Sigma^{\mu\nu}$. In this basis the $\Sigma^{\mu\nu}$ are imaginary.

All these results are summarized in the table B.1. The number of real parameters in the smallest representation is indicated in each case. This is twice the dimension of the Dirac representation, reduced by a factor of 2 for a Weyl condition and 2 for a Majorana condition. The derivation

Table B.1. *Dimensions in which various conditions are allowed for $SO(d-1,1)$ spinors. A dash indicates that the condition cannot be imposed. For the Weyl representation, it is indicated whether these are conjugate to themselves or to each other (complex). The final column lists the smallest representation in each dimension, counting the number of real components. Except for the final column the properties depend only on d mod 8.*

d	Majorana	Weyl	Majorana–Weyl	min. rep.
2	yes	self	yes	1
3	yes	-	-	2
4	yes	complex	-	4
5	-	-	-	8
6	-	self	-	8
7	-	-	-	16
8	yes	complex	-	16
9	yes	-	-	16
10=2+8	yes	self	yes	16
11=3+8	yes	-	-	32
12=4+8	yes	complex	-	64

implies that the properties are periodic in d with period 8, except the dimension of the representation which increases by a factor of 16.

For d a multiple of 4, a spinor may have the Majorana or Weyl property but not both: conjugation changes one Weyl representation into the other. In fact, the two cases are physically identical, there being a one-to-one mapping between them. Define the chirality projection operators

$$P_{\pm} = \frac{1 \pm \Gamma}{2} . \tag{B.1.22}$$

Given a Majorana spinor ζ or a Weyl spinor χ, the maps

$$\zeta \to P_{+}\zeta , \quad \chi \to \chi + B\chi^{*} \tag{B.1.23}$$

give a spinor of the other type, and these maps are inverse to one another.

The matrices $-\Gamma^{\mu T}$ also satisfy the Clifford algebra. The *charge conjugation* matrix has the property

$$C\Gamma^{\mu}C^{-1} = -\Gamma^{\mu T} . \tag{B.1.24}$$

Using the hermiticity property

$$\Gamma^{\mu\dagger} = \Gamma_{\mu} = -\Gamma^{0}\Gamma^{\mu}(\Gamma^{0})^{-1} , \tag{B.1.25}$$

this implies that

$$C\Gamma^{0}\Gamma^{\mu}(C\Gamma^{0})^{-1} = \Gamma^{\mu *} . \tag{B.1.26}$$

Then for even d,

$$C = B_1\Gamma^0 , \ d = 2 \bmod 4 ; \quad C = B_2\Gamma^0 , \ d = 4 \bmod 4 . \tag{B.1.27}$$

For odd $d = 2k+3$, again only $C = B_1\Gamma^0$ acts uniformly on Γ^{μ} for all μ; with this definition $C\Gamma^{\mu}C^{-1} = (-1)^{k+1}\Gamma^{\mu T}$. In all cases,

$$C\Sigma^{\mu\nu}C^{-1} = -\Sigma^{\mu\nu T} . \tag{B.1.28}$$

Additional properties of the matrices B and C are developed in exercise B.1.

Product representations

We now wish to develop the decomposition of a product of spinor representations. A product of spinors ζ and χ will have integer spins and so can be decomposed into tensor representations. Recall the standard spinor invariant

$$\bar{\zeta}\chi = \zeta^{\dagger}\Gamma^{0}\chi . \tag{B.1.29}$$

Similarly

$$\bar{\zeta}\Gamma^{\mu_1}\Gamma^{\mu_2}\ldots\Gamma^{\mu_m}\chi \tag{B.1.30}$$

transforms as the indicated tensor. However, this involves conjugation of the spinor ζ. From the properties of C it follows that $\zeta^{T}C$ transforms in the same way as $\bar{\zeta}$, so for the product of spinors without conjugation

$$\zeta^{T}C\Gamma^{\mu_1}\Gamma^{\mu_2}\ldots\Gamma^{\mu_m}\chi \tag{B.1.31}$$

transforms as a tensor.

Starting now with the case of $d = 2k+3$ odd, we claim that

$$\zeta^{T}C\Gamma^{\mu_1\mu_2\ldots\mu_m}\chi \tag{B.1.32}$$

for $m \le k+1$ comprise a complete set of independent tensors. Here

$$\Gamma^{\mu_1\mu_2\ldots\mu_m} = \Gamma^{[\mu_1}\Gamma^{\mu_2}\ldots\Gamma^{\mu_m]} \tag{B.1.33}$$

is the completely antisymmetrized product. Without the antisymmetry these would not be independent, as the anticommutation relation would allow a pair of Γ matrices to be removed. The restriction $m \le k+1$ comes about as follows. The definition of Γ implies in even dimensions that

$$\Gamma^{\mu_1\ldots\mu_s}\Gamma = -\frac{i^{-k+s(s+1)}}{(d-s)!}\epsilon^{\mu_1\ldots\mu_d}\Gamma_{\mu_{s+1}\ldots\mu_d} . \tag{B.1.34}$$

In odd dimensions, where $\Gamma^{d} = \pm\Gamma$, it follows that the antisymmetrized products (B.1.33) for m and $d-m$ are linearly related. There are no further

restrictions, and the dimensions agree: $2^{k+1} \cdot 2^{k+1}$ in the product of spinors and 2^{2k+2} from the binomial expansion. Thus

$$\mathbf{2^{k+1}} \times \mathbf{2^{k+1}} = [0] + [1] + \ldots + [k+1] \,, \tag{B.1.35}$$

where $[m]$ denotes the antisymmetric m-tensor.

For even $d = 2k + 2$, the products of m and $d - m$ Γ matrices are independent, and the same construction leads to

$$\begin{aligned} \mathbf{2^{k+1}_{Dirac}} \times \mathbf{2^{k+1}_{Dirac}} &= [0] + [1] + \ldots + [2k+2] \\ &= [0]^2 + [1]^2 + \ldots + [k]^2 + [k+1] \,. \end{aligned} \tag{B.1.36}$$

In the second line we have used the equivalence $[m] = [d - m]$ from contraction with the ϵ-tensor. Again the dimensionality is correct.

To find the products of the separate Weyl representations, use

$$\zeta^T C \Gamma^{\mu_1 \mu_2 \ldots \mu_m} \Gamma \chi = (-1)^{k+m+1} (\Gamma \zeta)^T C \Gamma^{\mu_1 \mu_2 \ldots \mu_m} \chi \,, \tag{B.1.37}$$

as follows from the definition of C. The tensor (B.1.32) is then nonvanishing if $k + m$ is odd and the chiralities of ζ and χ are the same, or if $k + m$ is even and the chiralities are opposite. This allows us to separate the product (B.1.36):

$$\mathbf{2^k} \times \mathbf{2^k} = \begin{cases} [1] + [3] + \ldots + [k+1]_+ \,, & k \text{ even} \,, \\ [0] + [2] + \ldots + [k+1]_+ \,, & k \text{ odd} \,, \end{cases} \tag{B.1.38a}$$

$$\mathbf{2^{k'}} \times \mathbf{2^{k'}} = \begin{cases} [1] + [3] + \ldots + [k+1]_- \,, & k \text{ even} \,, \\ [0] + [2] + \ldots + [k+1]_- \,, & k \text{ odd} \,, \end{cases} \tag{B.1.38b}$$

$$\mathbf{2^k} \times \mathbf{2^{k'}} = \begin{cases} [0] + [2] + \ldots + [k] \,, & k \text{ even} \,, \\ [1] + [3] + \ldots + [k] \,, & k \text{ odd} \,. \end{cases} \tag{B.1.38c}$$

The relation (B.1.34) implies that the tensors of rank $k + 1 = d/2$ satisfy a self-duality condition with a sign that depends on the chirality of the spinor. A self-dual tensor representation can only be real for k even.

Some of the facts that we have deduced can also be verified quickly by considering the eigenvalues s_a. Consider the reality properties of the Weyl spinors. Conjugation flips the rotation eigenvalues $s_1, \ldots, s_k$ but not the boost eigenvalue s_0. For k even, this is an even number of flips and gives a state of the same chirality; for k odd it reverses the chirality. This is consistent with the third column of table B.1. For the tensor products of Weyl representations, note that the even-rank tensors $[2n]$ (e.g. the invariant $[0]$) always contain a component with eigenvalues $s_a = (0, 0, \ldots, 0)$, while the odd-rank tensors do not. This would be obtained, for example, from the product of spinor components $s_a = (\frac{1}{2}, \frac{1}{2}, \ldots, \frac{1}{2})$ and $s_a = (-\frac{1}{2}, -\frac{1}{2}, \ldots, -\frac{1}{2})$. For k even these have opposite chirality, as in the product (B.1.38c). For k odd they have the same chirality, as in the products (B.1.38a) and (B.1.38b).

Table B.2. *Dimensions in which various conditions are allowed for $SO(N)$ spinors.*

N mod 8	real	Weyl	real and Weyl
0	yes	self	yes
1	yes	-	-
2	yes	complex	-
3	pseudo	-	-
4	pseudo	self	-
5	pseudo	-	-
6	yes	complex	-
7	yes	-	-

Spinors of $SO(N)$

For $SO(N)$ the analysis is quite parallel. For $N = 2l$, there is a 2^l-dimensional representation of the Γ-matrices which reduces to two 2^{l-1}-dimensional spinor representations of $SO(2l)$, while for $SO(2l+1)$ there is a single representation of dimension 2^l. The reality properties can be analyzed as in the Minkowski case. Essentially one ignores $\mu = 0, 1$, so $SO(N)$ is analogous to $SO(N + 1, 1)$, with the results shown in table B.2. Here *real* means the algebra can be written in terms of purely imaginary matrices. The term *pseudoreal* is often used for $N = 3, 4, 5$ mod 8, where the representation is conjugate to itself but cannot be written in terms of imaginary matrices.

The familiar case of a pseudoreal representation is the **2** of $SO(3)$. This is conjugate to itself because it is the only two-dimensional representation, but it must act on a complex doublet. It should be noted, however, that two wrongs make a right — the product of two pseudoreal representations is real. Let the indices on u_{ij} both be $SU(2)$ doublets, either of the same or different $SU(2)$s. Then the reality condition

$$u^*_{ij} = \epsilon_{ii'} \epsilon_{jj'} u_{i'j'} \tag{B.1.39}$$

is invariant. With just a single index, the analogous condition $u^*_i = \epsilon_{ii'} u_{i'}$ would force u to vanish. Incidentally, one can impose a Majorana condition on the **2** of $SO(2, 1)$, consistent with table B.1. A real basis for the Γ-matrices is

$$\Gamma^0 = i\sigma^2 \,, \quad \Gamma^1 = \sigma^1 \,, \quad \Gamma^2 = \sigma^3 \,. \tag{B.1.40}$$

Product representations are obtained as in the Minkowski case, with

the result in $N = 2l$

$$\mathbf{2^{l-1}} \times \mathbf{2^{l-1}} = \begin{cases} [0]+[2]+\ldots+[l]_+ , & l \text{ even} , \\ [1]+[3]+\ldots+[l]_+ , & l \text{ odd} , \end{cases} \tag{B.1.41a}$$

$$\mathbf{2^{l-1'}} \times \mathbf{2^{l-1'}} = \begin{cases} [0]+[2]+\ldots+[l]_- , & l \text{ even} , \\ [1]+[3]+\ldots+[l]_- , & l \text{ odd} , \end{cases} \tag{B.1.41b}$$

$$\mathbf{2^{l-1}} \times \mathbf{2^{l-1'}} = \begin{cases} [1]+[3]+\ldots+[l-1] , & l \text{ even} , \\ [0]+[2]+\ldots+[l-1] , & l \text{ odd} . \end{cases} \tag{B.1.41c}$$

For more than one timelike dimension, the analog of table B.1 or B.2 depends on the *difference* of the number of spacelike and timelike dimensions.

Decomposition under subgroups

We frequently consider subgroups such as

$$SO(9,1) \to SO(3,1) \times SO(6) . \tag{B.1.42}$$

We can directly match representations by comparing the eigenvalues of S_a. In particular, for the case in which all the dimensions are even,

$$SO(2k+1,1) \to SO(2l+1,1) \times SO(2k-2l) , \tag{B.1.43}$$

the Weyl spinors decompose

$$\mathbf{2^k} \to (\mathbf{2^l}, \mathbf{2^{k-l-1}}) + (\mathbf{2^{l'}}, \mathbf{2^{k-l-1'}}) , \tag{B.1.44a}$$

$$\mathbf{2^{k'}} \to (\mathbf{2^{l'}}, \mathbf{2^{k-l-1}}) + (\mathbf{2^{l}}, \mathbf{2^{k-l-1'}}) . \tag{B.1.44b}$$

Another subgroup that has particular relevance for the superstring is

$$SO(2n) \to SU(n) \times U(1) . \tag{B.1.45}$$

To describe this subgroup, consider again the complex linear combinations (B.1.2) of Γ-matrices, where $a = 1,\ldots,n$. A general $SO(2n)$ rotation will mix the Γ^{a+} both among themselves and with the Γ^{a-}. The subgroup that mixes the Γ^{a+} only among themselves is $U(n) = SU(n) \times U(1)$. Now let us consider how the spinor representation decomposes. Again we start with the spinor ζ annihilated by all the Γ^{a-}. This condition $\Gamma^{a-}\zeta = 0$ is invariant under $U(n)$ rotations so that ζ rotates at most by a phase. Thus

$$\zeta \in \mathbf{1}_{-n} , \tag{B.1.46}$$

where the $U(n)$ charge, indicated by the subscript, has been normalized to $2\sum S_a$. Acting with a raising operator adds an $SU(n)$ index and increases the $U(1)$ charge by 2, giving

$$\mathbf{2^n} \to [0]_{-n} + [1]_{2-n} + [2]_{4-n} + \ldots + [n]_n , \tag{B.1.47}$$

where $[k]$ refers to the k-times antisymmetrized $\mathbf{n}$ of $SU(n)$. The completely antisymmetrized $[n]$ is the same as $[0] = \mathbf{1}$, while $[n-1] = \overline{[1]} = \bar{\mathbf{n}}$, and so on. Decomposing further into the Weyl representations, the last term $[0]_n$ is in the $\mathbf{2^{n-1}}$, and the successive terms alternate. Thus in particular for

$$SO(6) \to SU(3) \times U(1) , \tag{B.1.48}$$

we have

$$\mathbf{4} \to \mathbf{1}_3 + \mathbf{3}_{-1} , \tag{B.1.49a}$$

$$\bar{\mathbf{4}} \to \mathbf{1}_{-3} + \bar{\mathbf{3}}_1 . \tag{B.1.49b}$$

A relation that arises often is

$$SO(4) = SU(2) \times SU(2) . \tag{B.1.50}$$

To see this, combine the four components of a vector into a 2×2 matrix

$$x = x^4 I + i x^i \sigma^i , \; i = 1,2,3 ; \quad \det x = \sum_{m=1}^{4} (x^m)^2 . \tag{B.1.51}$$

The length of x is invariant under independent left- and right-hand $SU(2)$ rotations

$$x' = g_1 x g_2^{-1} , \tag{B.1.52}$$

giving the decomposition (B.1.50). Then

$$\mathbf{4} = (\mathbf{2},\mathbf{2}) , \tag{B.1.53a}$$

$$\mathbf{2} = (\mathbf{2},\mathbf{1}) , \tag{B.1.53b}$$

$$\mathbf{2'} = (\mathbf{1},\mathbf{2}) . \tag{B.1.53c}$$

The decomposition of the vector is just eq. (B.1.52), while those of the spinors can be derived in various ways.

B.2 Introduction to supersymmetry: $d = 4$

The familiar conserved quantities, such as energy-momentum, angular momentum, and charge, transform as vectors, tensors, and scalars under the Lorentz group. It is also possible for a conserved quantity to transform as a spinor. Such a *supersymmetry (SUSY)* will relate the properties of fermions to those of bosons. Supersymmetry is a feature of all consistent string theories. Further, as discussed in section 16.2, there is good reason to expect that it will be found with particle accelerators.

In this appendix we summarize the various results that will be needed in the text. We are interested in the algebras, their representations, the transformations of the fields, and the invariant actions. The reader should be able to follow the derivation of the various representations (massless,

standard massive, and BPS massive). However, the transformations and actions require detailed calculation, and so for these we simply cite for reference some of the key results.

$d = 4$, $N = 1$ *supersymmetry*

According to table B.1, the smallest spinor in four dimensions has four real degrees of freedom. As shown in eq. (B.1.23) this can be described either as a Weyl spinor, with two complex components, or as a Majorana spinor, with four components satisfying a reality condition.

The smallest $d = 4$ supersymmetry algebra would have one Weyl or Majorana spinor of supercharges. Again these are identical, the same four linearly independent supercharges described in two different notations; we will use the Majorana description. A more general supersymmetry algebra in $d = 4$ would have $4N$ supercharges. For $N > 1$ this is known as *extended supersymmetry*. In any number of dimensions the ratio of the number of supercharges to the smallest spinor representation is denoted by N. However, the structure of the theory depends more on the actual number of supercharges than on the ratio N, so subsequent sections are organized according to this total number. For pedagogic purposes we find it convenient in this section to start with the smallest algebra and build up, but later we will start with the largest algebra and work downwards, from 32 to 16 to 8. The number of supercharges need not be a power of 2, but in the great majority of examples it is and so these are the cases on which we focus.

The $N = 1$ supersymmetry algebra is uniquely determined to be

$$\{Q_\alpha, \bar{Q}_\beta\} = -2P_\mu \Gamma^\mu_{\alpha\beta} , \tag{B.2.1a}$$

$$[P^\mu, Q_\alpha] = 0 , \tag{B.2.1b}$$

where P_μ is the spacetime momentum. The minus sign is due to our metric signature $(- + \ldots +)$. Recall that from the Majorana property, $\bar{Q} \equiv Q^\dagger \Gamma^0 = Q^T C$.

It is easy to work out the representations of this algebra. The massless and massive representations differ, and we consider the former first. For massless states choose a frame in which $k^1 = k^0$. The supersymmetry algebra becomes

$$\{Q_\alpha, Q^\dagger_\beta\} = 2k^0(1 + \Gamma^0\Gamma^1)_{\alpha\beta} = 2k^0(1 + 2S_0)_{\alpha\beta} . \tag{B.2.2}$$

In the s-basis, the Majorana condition becomes $Q^\dagger_{s_0 s_1} = Q_{s_0,-s_1}$ and the anticommutator becomes

$$\{Q_{s'_0 s'_1}, Q^\dagger_{s_0 s_1}\} = 4k^0 \delta_{s_0,1/2}\delta_{ss'} . \tag{B.2.3}$$

The matrix elements of $Q_{-1/2,s_1}$ must vanish in these momentum eigenstates because

$$\begin{aligned} 0 &= \langle\psi|\{Q_{-1/2,s_1}, Q^\dagger_{-1/2,s_1}\}|\psi\rangle \\ &= \|Q_{-1/2,s_1}|\psi\rangle\|^2 + \|Q^\dagger_{-1/2,s_1}|\psi\rangle\|^2 . \end{aligned} \tag{B.2.4}$$

The remaining supercharges form a fermionic oscillator algebra. Defining

$$b = (4k^0)^{-1/2} Q_{1/2,-1/2} , \quad b^\dagger = (4k^0)^{-1/2} Q_{1/2,1/2} , \tag{B.2.5}$$

the supersymmetry algebra becomes

$$\{b, b^\dagger\} = 1 , \quad b^2 = b^{\dagger 2} = 0 . \tag{B.2.6}$$

Starting from a state $|\lambda\rangle$ such that

$$S_1|\lambda\rangle = \lambda|\lambda\rangle , \quad b|\lambda\rangle = 0 , \tag{B.2.7}$$

the algebra generates exactly one additional state

$$b^\dagger|\lambda\rangle = |\lambda + \tfrac{1}{2}\rangle , \qquad S_1|\lambda + \tfrac{1}{2}\rangle = (\lambda + \tfrac{1}{2})|\lambda + \tfrac{1}{2}\rangle . \tag{B.2.8}$$

The massless irreducible multiplets thus each consist of two states with helicities differing by $\frac{1}{2}$: one state in each multiplet is a fermion and one a boson. These are also representations of Poincaré symmetry. However, CPT, which appears to be an exact symmetry of string theory as it is of field theory, requires that each multiplet be accompanied by its conjugate with opposite helicities and quantum numbers. Thus we have the following $(\lambda, \lambda + \frac{1}{2})$ multiplets:

- The *chiral multiplet* consists of a $(0, \frac{1}{2})$ multiplet and its CPT conjugate $(-\frac{1}{2}, 0)$, corresponding to a Weyl fermion and a complex scalar.
- The *vector multiplet* $(\frac{1}{2}, 1)$ plus $(-1, -\frac{1}{2})$ contains a gauge boson and a Weyl fermion, both necessarily in the adjoint of the gauge group.
- The *gravitino multiplet* $(1, \frac{3}{2})$ plus $(-\frac{3}{2}, -1)$ contains an additional spin-$\frac{3}{2}$ gravitino and so is not relevant since there is only one supersymmetry and so only the gravitino in the graviton multiplet. This multiplet would be relevant if we had a larger supersymmetry and decomposed it into $N = 1$ representations.
- The *graviton multiplet* $(\frac{3}{2}, 2)$ plus $(-2, -\frac{3}{2})$ contains the graviton and gravitino.
- Massless particles with helicities greater than 2 are believed to be impossible to couple to gravity, and have not arisen in string theory.

In an $N = 1$ supersymmetric extension of the Standard Model, the Higgs boson and spin-$\frac{1}{2}$ fermions are in chiral multiplets. The Standard Model fermions cannot be in vector multiplets because the latter must be in the adjoint representation.

For massive representations, the anticommutator in the rest frame is

$$\{Q_{s_0 s_1'}, Q^\dagger_{s_0 s_1}\} = 2m\delta_{ss'} \,. \tag{B.2.9}$$

This is now *two* copies of the fermionic oscillator algebra,

$$b_1 = (2m)^{-1/2} Q_{1/2,-1/2} \,, \quad b_2 = (2m)^{-1/2} Q_{-1/2,-1/2} \,, \tag{B.2.10a}$$

$$\{b_i, b_j^\dagger\} = \delta_{ij} \,, \quad \{b_i, b_j\} = \{b_i^\dagger, b_j^\dagger\} = 0 \,. \tag{B.2.10b}$$

Starting again from a state

$$S_1|\lambda\rangle = \lambda|\lambda\rangle \,, \quad b_i|\lambda\rangle = 0 \,, \tag{B.2.11}$$

the algebra generates the additional three states

$$b_1^\dagger|\lambda\rangle,\ b_2^\dagger|\lambda\rangle,\ b_1^\dagger b_2^\dagger|\lambda\rangle \,, \quad S_1 = \lambda + \tfrac{1}{2},\ \lambda + \tfrac{1}{2},\ \lambda + 1 \,. \tag{B.2.12}$$

For example, the massive chiral multiplet is $\lambda = -\frac{1}{2}, 0, 0, \frac{1}{2}$, the same as the CPT-extended massless multiplet. The multiplet $\lambda = 0, \frac{1}{2}, \frac{1}{2}, 1$ is incomplete, even without CPT, because massive states must be a representation of the rotation group $SU(2)$. Adding in $\lambda = -1, -\frac{1}{2}, -\frac{1}{2}, 0$, we obtain a spin-1, two spin-$\frac{1}{2}$, and one spin-0 particle. These are the same states as a massless vector plus chiral multiplet, and can be obtained from them via the Higgs mechanism.

Actions with $d = 4$, $N = 1$ SUSY

From section 16.4 on, we need some results about $d = 4$, $N = 1$ supersymmetry transformations and invariant actions. We collect these here, without derivation. A general renormalizable theory will contain a number of massless chiral and vector multiplets; the larger massive multiplets can always be decomposed into these. The particle content of the massless chiral multiplet corresponds to a complex scalar field ϕ and a Majorana (or Weyl) spinor ψ. That of a massless vector multiplet corresponds to a gauge field A_μ and a Majorana (or Weyl) spinor λ. In each case it is useful, though not essential, to add a complex auxiliary field, a complex field F in the chiral multiplet and a real field D in the vector multiplet. We then have the following *superfields*

$$\Phi^i\colon\ \phi^i \,,\ \psi^i \,,\ F^i \,, \tag{B.2.13a}$$

$$V^a\colon\ A_\mu^a \,,\ \lambda^a \,,\ D^a \,. \tag{B.2.13b}$$

These have the supersymmetry transformations

$$\delta\phi^i/2^{1/2} = i\bar\zeta \mathrm{P}_+\psi^i = i\bar\psi^i \mathrm{P}_+\zeta \,, \tag{B.2.14a}$$

$$\delta(\mathrm{P}_+\psi^i)/2^{1/2} = \mathrm{P}_+\zeta F^i + \Gamma^\mu \mathrm{P}_-\zeta D_\mu\phi^i \,, \tag{B.2.14b}$$

$$\delta F^i/2^{1/2} = -i\bar\zeta\Gamma^\mu D_\mu \mathrm{P}_+\psi^i \,, \tag{B.2.14c}$$

and

$$\delta A_\mu^a = -i\bar\zeta\Gamma_\mu\lambda^a \,, \tag{B.2.15a}$$

$$\delta\lambda^a = \frac{1}{2}\Gamma^{\mu\nu}\zeta F_{\mu\nu}^a + i\Gamma\zeta D^a \,, \tag{B.2.15b}$$

$$\delta D^a = -\bar\zeta\Gamma\Gamma^\mu D_\mu\lambda^a \,, \tag{B.2.15c}$$

in terms of a Majorana SUSY parameter ζ.

The most general renormalizable action is determined by the gauge couplings g_a (which of course must be equal within each simple group) and the *superpotential* $W(\Phi)$, which is a *holomorphic* function of the superfields. Also, for each $U(1)$ gauge group there is an additional parameter ξ_a, the *Fayet–Iliopoulos term*. The Lagrangian density is

$$\mathscr{L} = \mathscr{L}_1 + \mathscr{L}_2 \,, \tag{B.2.16}$$

where

$$\mathscr{L}_1 = -D_\mu\phi^{i*}D^\mu\phi^i - \frac{i}{2}\bar\psi^i\Gamma^\mu D_\mu\psi^i - \frac{1}{4g_a^2}F_{\mu\nu}^a F^{a\mu\nu} - \frac{i}{2g_a^2}\bar\lambda^a\Gamma^\mu D_\mu\lambda^a$$
$$- \frac{1}{2}\left[iW_{,ij}(\phi)\bar\psi^i\mathrm{P}_+\psi^j + 2^{1/2}\phi^{i*}t_{ij}^a\bar\lambda^a\mathrm{P}_+\psi^j\right] + \text{c.c.} \,, \tag{B.2.17}$$

and

$$\mathscr{L}_2 = F^{i*}F^i + \frac{1}{2g_a^2}D^{a2} + W_{,i}(\phi)F^i + \text{c.c.} + \frac{1}{2}D^a(2\xi_a + \phi^{i*}t_{ij}^a\phi^j) \,. \tag{B.2.18}$$

In $\mathscr{L}_1$ are the kinetic terms, fermion masses and Yukawa couplings, while in $\mathscr{L}_2$ are all terms involving the auxiliary fields. The t_{ij}^a are the gauge group representation matrices. Renormalizability requires the superpotential W to be at most cubic in the fields. Carrying out the Gaussian path integration over the auxiliary fields gives a scalar potential

$$-\mathscr{L}_2' = V = |F^i(\phi)|^2 + \frac{1}{2g_a^2}[D^a(\phi,\phi^*)]^2 \,, \tag{B.2.19}$$

where

$$F^i(\phi) = -W_{,i}(\phi)^* \,, \tag{B.2.20a}$$

$$D^a(\phi,\phi^*) = -\frac{g_a^2}{2}(2\xi_a + \phi^{i*}t_{ij}^a\phi^j) \,. \tag{B.2.20b}$$

The two terms in the potential are known respectively as the F-term, from the superpotential, and the D-term, from the gauge interaction.

An important nonrenormalization theorem states that the tree-level superpotential does not receive perturbative corrections. It is also important that this is only a perturbative statement, and that there can be nonperturbative corrections to the superpotential. An example of this arises in chapter 18.

Two kinds of internal symmetry are possible in supersymmetry. The first is a unitary rotation U_{ij} acting uniformly on all fields ϕ^i, $P_+\psi^i$ and F^i in a given chiral multiplet. This is a symmetry if W is invariant. The gauge fields A_μ^a couple to such a symmetry. The second, known as an *R symmetry*, acts differently on different components:

$$\phi^i \to \exp(iq_i\alpha)\phi^i\,, \quad P_+\psi^i \to \exp[i(q_i-1)\alpha]P_+\psi^i\,, \tag{B.2.21a}$$

$$F^i \to \exp[i(q_i-2)\alpha]F^i\,, \tag{B.2.21b}$$

$$A_\mu^a \to A_\mu^a\,, \quad P_+\lambda^a \to \exp(i\alpha)P_+\lambda^a\,, \quad D^a \to D^a\,. \tag{B.2.21c}$$

Examining the action, for example the Yukawa terms, one sees that this is a symmetry provided the superpotential transforms as

$$W(\phi) \to \exp(2i\alpha)W(\phi)\,. \tag{B.2.22}$$

In addition, the R symmetry must commute with the gauge symmetry.

Spontaneous supersymmetry breaking

As with an ordinary internal symmetry, spontaneous breaking of supersymmetry is signified by certain nonvanishing vacuum expectation values. In particular, consider

$$\langle 0|\{Q_\alpha, \chi_\beta\}|0\rangle\,, \tag{B.2.23}$$

where Q_α is some component of the supercharge. We can assume the operator χ_β to be fermionic; otherwise, the expectation value vanishes automatically by Lorentz invariance. If supersymmetry is unbroken, $Q_\alpha|0\rangle = \langle 0|Q_\alpha = 0$ and all such vacuum expectation values vanish. Classically the condition for unbroken supersymmetry becomes

$$\delta\psi^i = \delta\lambda^a = 0\,. \tag{B.2.24}$$

From the variations (B.2.14) and (B.2.15), it follows that a configuration is supersymmetric if the fields are position-independent, the gauge field is zero, and

$$F^i(\phi) = D^a(\phi) = 0\,. \tag{B.2.25}$$

Moreover, we see from the potential (B.2.19) that if such a configuration exists it will be a minimum of the energy. Supersymmetry will be spontaneously broken if there are no solutions to eqs. (B.2.25). The simplest example of a system with broken supersymmetry is a single superfield with superpotential

$$W = f\phi_1\,, \tag{B.2.26}$$

with f a nonzero constant; then $F = -f^* \neq 0$. This is rather trivial as it stands, but by coupling ϕ_1 appropriately to other fields, for example

$$W = f\phi_1 + m\phi_2\phi_3 + g\phi_1\phi_2^2\,, \tag{B.2.27}$$

one obtains a theory with a nonsupersymmetric spectrum.

Higher corrections and supergravity

In the usual power counting in four dimensions, the scalar field and vector potential have dimension l^{-1} and the spinors dimension $l^{-3/2}$, l being length. These are determined by the kinetic terms. It follows from the transformations (B.2.14) and (B.2.15) that the supersymmetry parameter ζ has dimension $l^{1/2}$, consistent with the product of two supersymmetry transformations being a translation. Also, the auxiliary fields F^i and D^a have dimension l^{-2}. Including the l^4 from d^4x, the renormalizable action retains all terms that are relevant at long distance, that is, all terms of dimension l^n with $n \geq 0$.

Power counting in renormalization theory is based on the scaling of the quantum fluctuations of the fields. However, in string theory we have encountered the phenomenon of moduli, scalar fields with flat potentials. These can have large *classical* values. In order to write an effective Lagrangian valid in all of moduli space,[1] we need a different power counting that assigns scalars scaling l^0. Supersymmetry then assigns their fermionic partners scaling $l^{-1/2}$. We wish to keep all terms of the same order as the kinetic terms for these fields, and therefore all terms in the Lagrangian density having dimension l^m with $m \geq -2$. In order to keep the kinetic terms for the gauge multiplet, assign A_μ scaling l^0 and λ scaling $l^{-1/2}$. Finally, we assign the metric scaling l^0, since it has a classical expectation value. Incidentally, this 'moduli space' power counting is the same in all dimensions, whereas the renormalization power counting is dimension-dependent.

[1] To be precise, the effective Lagrangian will still break down at particular points in moduli space, namely those points where extra massless fields occur. In the neighborhood of such a point, one needs an effective Lagrangian which includes these additional fields.

In this approximation, the low energy effective action includes all the earlier terms plus additional ones. It depends now on three functions:

- The superpotential $W(\Phi)$, which is still holomorphic but need no longer be cubic.
- An arbitrary holomorphic function $f_{ab}(\Phi)$ replacing the gauge coupling g_a^{-2}.
- The *Kähler* potential $K(\Phi, \Phi^*)$, which is a general function of the superfields.

Again there is a Fayet–Iliopoulos parameter ξ_a for each $U(1)$. The full Lagrangian density is quite lengthy, so we give only the purely bosonic terms,

$$\frac{\mathscr{L}_{\text{bos}}}{(-G)^{1/2}} = \frac{1}{2\kappa^2}R - K_{,\bar{i}j}D_\mu\phi^{i*}D^\mu\phi^j - \frac{1}{4}\text{Re}(f_{ab}(\phi))F^a_{\mu\nu}F^{b\mu\nu} - \frac{1}{8}\text{Im}(f_{ab}(\phi))\epsilon^{\mu\nu\sigma\rho}F^a_{\mu\nu}F^b_{\sigma\rho} - V(\phi,\phi^*) . \quad \text{(B.2.28)}$$

The potential is

$$V(\phi,\phi^*) = \exp(\kappa^2 K)(K^{\bar{i}j}W^*_{\bar{i}}W_{;j} - 3\kappa^2 W^*W) + \frac{1}{2}f_{ab}D^aD^b . \quad \text{(B.2.29)}$$

Here $K^{\bar{i}j}$ is the inverse matrix to $\partial_j\partial_{\bar{k}}K$ and

$$W_{;i} = \partial_i W + \kappa^2\partial_i K\, W \quad \text{(B.2.30a)}$$

$$\text{Re}(f_{ab}(\phi))D^b = -2\xi_a - K_{,i}t^a_{ij}\phi^j . \quad \text{(B.2.30b)}$$

The negative term proportional to κ^2 is a supergravity effect. The other terms generalize the earlier potential (B.2.19).

The kinetic term for the scalars is now field-dependent. The second derivative

$$K_{,\bar{i}j} = \frac{\partial^2 K(\phi,\phi^*)}{\partial\phi^{i*}\partial\phi^j} \quad \text{(B.2.31)}$$

plays the role of a metric for the space of scalar fields, generalizing the flat metric $\delta_{\bar{i}j}$ of the renormalizable theory. The flat metric is the special case $K = \phi^{i*}\phi^i$. A metric of the form (B.2.31) is known as a *Kähler metric*. In a similar way, the function $f_{ab}(\phi)$ gives rise to a field-dependent (*nonminimal*) kinetic term for the gauge fields, as well as a field-dependent $F_2 \wedge F_2$ coupling. The metric (B.2.31) is invariant under *Kähler transformations*,

$$K(\phi,\phi^*) \to K(\phi,\phi^*) + f(\phi) + f(\phi)^* . \quad \text{(B.2.32)}$$

This is an invariance of the whole action provided also that the superpotential transforms as

$$W(\phi) \to \exp[-\kappa^2 f(\phi)]W(\phi) . \quad \text{(B.2.33)}$$

This is important because in interesting examples the field space has a nontrivial topology and the Kähler potential is not globally defined.

The supersymmetry transformations of the fermions are

$$\delta \text{P}_+\psi^i/2^{1/2} = -K^{\bar{i}j}W^*_{;j}\text{P}_+\zeta + \Gamma^\mu \text{P}_-\zeta D_\mu\phi^i , \quad \text{(B.2.34a)}$$

$$\delta\lambda^a = \frac{1}{2}\Gamma^{\mu\nu}\zeta F^a_{\mu\nu} + i\Gamma\zeta D^a , \quad \text{(B.2.34b)}$$

$$\delta\psi_\mu = D_\mu\zeta + \frac{1}{2}\Gamma_\mu\zeta\exp(\kappa^2 K/2)W . \quad \text{(B.2.34c)}$$

Here ψ_μ is the gravitino. The covariant derivative of the spinor ζ includes the spin connection. The variations (B.2.34) all vanish if the metric is flat, the gauge field zero, the scalars and ζ constant, and $\partial_i W = D^a = W = 0$.

Extended supersymmetry in d = 4

With several supersymmetries Q^A_α for $A = 1,\ldots,N$, the straightforward generalization of the earlier algebra is

$$\{Q^A_\alpha, \bar{Q}^B_\beta\} = -2\delta^{AB}P_\mu\Gamma^\mu_{\alpha\beta} , \quad \{P^\mu, Q^A_\alpha\} = 0 . \quad \text{(B.2.35)}$$

This is not the most general algebra, but we analyze it first. Fôr massless particles, the earlier fermionic oscillator is replaced by N oscillators b_A. These generate 2^N states in a binomial distribution from helicity λ to helicity $\lambda + \frac{1}{2}N$. For example, for $N = 2$ the following massless multiplets are important:

hypermultiplet:	$(-\frac{1}{2}, 0^2, \frac{1}{2}) + (-\frac{1}{2}, 0^2, \frac{1}{2})$,
vector multiplet:	$(-1, -\frac{1}{2}^2, 0) + (0, \frac{1}{2}^2, 1)$,
supergravity multiplet:	$(-2, -\frac{3}{2}^2, -1) + (1, \frac{3}{2}^2, 2)$.

In each case there are two SUSY multiplets, related by CPT. For states that are their own CPT conjugates, a half-hypermultiplet is allowed.

Let us note an important feature of these multiplets. If we just look at the SUSY multiplets, not making use of CPT, then all states in the multiplet have the same gauge quantum numbers because the supersymmetry charges commute with the gauge symmetries.[2] It follows that an $N = 2$ theory cannot have chiral gauge interactions. The only SUSY multiplets

[2] There is an exception to this known as gauged supergravity, but it is not relevant to the gauge interactions of the Standard Model.

with spin-$\frac{1}{2}$ states are the half-hypermultiplet and the vector multiplet. The former contains states of helicities $\pm\frac{1}{2}$ with the same gauge quantum numbers and so is nonchiral. The latter is necessarily in the (real) adjoint representation and so is also nonchiral.

For $N = 4$ the multiplets are larger:

vector multiplet: $(-1, -\frac{1}{2}^4, 0^6, \frac{1}{2}^4, 1)$;

supergravity multiplet: $(-2, -\frac{3}{2}^4, -1^6, -\frac{1}{2}^4, 0) + (0, \frac{1}{2}^4, 1^6, \frac{3}{2}^4, 2)$.

Finally, for $N = 8$ there is only a single possible representation:

supergravity multiplet: $(-2, -\frac{3}{2}^8, -1^{28}, -\frac{1}{2}^{56}, 0^{70}, \frac{1}{2}^{56}, 1^{28}, \frac{3}{2}^8, 2)$.

Larger algebras would require helicities greater than 2, which is believed to be impossible (there are some uninteresting exceptions, such as free field theories). String theory has several times turned up loopholes in such statements, but not yet here.

Massive representations of extended supersymmetry similarly contain 2^{2N} states generated by b_{1A} and b_{2A}.

The most general extended supersymmetry algebra allowed by Lorentz invariance is

$$\{Q_\alpha^A, \bar{Q}_\beta^B\} = -2\delta^{AB} P_\mu \Gamma^\mu_{\alpha\beta} - 2iZ^{AB}\delta_{\alpha\beta} , \tag{B.2.36a}$$

$$\{P^\mu, Q_\alpha^A\} = [Z^{AB}, Q_\alpha^C] = [Z^{AB}, P_\mu] = [Z^{AB}, Z^{CD}] = 0 . \tag{B.2.36b}$$

Here Z^{AB} is some set of conserved charges. It must be antisymmetric in AB due to the Majorana property and the antisymmetry of the charge conjugation matrix C.

To be precise, this is the most general algebra if we include only charges that can be carried by point particles. Including charges that can be carried by *extended objects*, additional terms appear. Rather than explain this here, we introduce it in its natural physical context: first in section 11.6, and then in more variety in chapter 13. The same caveat applies to the higher-dimensional algebras to be introduced later in this appendix.

To see the effect of the additional term consider a particle in its rest frame, for which the algebra becomes

$$\{Q_\alpha^A, Q_\beta^{B\dagger}\} = 2m\delta^{AB}\delta_{\alpha\beta} + 2iZ^{AB}\Gamma^0_{\alpha\beta} . \tag{B.2.37}$$

Taking an eigenstate of the Z^{AB}, we can go to a basis in which

$$Z^{AB} = \begin{bmatrix} 0 & q_1 & 0 & 0 & \\ -q_1 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & q_2 & \\ 0 & 0 & -q_2 & 0 & \\ & \vdots & & & \ddots \end{bmatrix} , \tag{B.2.38}$$

with $q_i \geq 0$. The left-hand side of the algebra (B.2.37) is nonnegative as a matrix in $(A\alpha, B\beta)$. The eigenvalues $m \pm q_i$ on the right-hand side must therefore also be nonnegative, implying the *Bogomolnyi–Prasad–Sommerfield (BPS) bound*

$$m \geq q_i . \tag{B.2.39}$$

Thus the mass is bounded below by the charges, and in particular massless states must be neutral. If m is strictly greater than all the q_i, the massive representations are unaffected and contain 2^{2N} states. If the largest k q_is are equal to one another and to m, the algebra requires $2k$ pairs of fermionic oscillators to annihilate the states, just as half the oscillators do for a massless representation. This gives a *short* or *BPS* representation with $2^{2(N-k)}$ states. If all the q_i are equal to one another and to m, the result is an *ultrashort* representation of dimension 2^N (for N even), the same as the massless representation.

B.3 Supersymmetry in $d = 2$

In this section we briefly make the connection with the world-sheet algebras of string theory. The smallest spinor representation in two dimensions is Majorana–Weyl and has one Hermitean component. The general $(N, \tilde{N})$ algebra would have N Hermitean left-moving supercharges Q_L^A and $\tilde{N}$ Hermitean right-moving supercharges Q_R^A. The algebra is

$$\{Q_L^A, Q_L^B\} = \delta^{AB}(P^0 - P^1) , \quad \{Q_R^A, Q_R^B\} = \delta^{AB}(P^0 + P^1) , \tag{B.3.1a}$$

$$\{Q_L^A, Q_R^B\} = Z^{AB} , \tag{B.3.1b}$$

where now Z^{AB} need have no special symmetry. The superconformal generators G_0 and $\tilde{G}_0$ satisfy this algebra. Thus the R sector of the $(N, \tilde{N})$ superconformal theory contains the $(N, \tilde{N})$ supersymmetry algebra. In fact, this was one of several independent routes by which supersymmetry was first discovered.

The dimensional reduction of the $d = 4$, $N = 1$ supersymmetry algebra gives the $d = 2$ (2,2) algebra.

B.4 Differential forms and generalized gauge fields

Various antisymmetric tensor fields appear in supergravity and string theory. Differential forms are a convenient notation to minimize the bookkeeping of indices and combinatoric factors. A *p-form* A is simply a completely antisymmetric p-index tensor $A_{\mu_1\ldots\mu_p}$ with the indices omitted. Because we encounter many different forms, we will denote the rank of any form by an italicized subscript, A_p. The product of a p-form A_p and a q-form B_q is written $A_p \wedge B_q$ or simply A_pB_q, and is defined

$$(A_p \wedge B_q)_{\mu_1\ldots\mu_{p+q}} = \frac{(p+q)!}{p!q!} A_{[\mu_1\ldots\mu_p} B_{\mu_{p+1}\ldots\mu_{p+q}]} \,. \tag{B.4.1}$$

Again, [ ] denotes antisymmetrization, averaging over permutations with a ± 1 for odd and even perturbations. The wedge product of a p-form A and q-form B has the property

$$A_p \wedge B_q = (-1)^{pq} B_q \wedge A_p \,. \tag{B.4.2}$$

The exterior derivative d takes a p-form into a $(p+1)$-form:

$$(dA_p)_{\mu_1\ldots\mu_{p+1}} = (p+1)\partial_{[\mu_1} A_{\mu_2\ldots\mu_{p+1}]} \,. \tag{B.4.3}$$

It has the important property $d^2 = 0$.

The integral of a d-form is coordinate-invariant,

$$\int d^dx\, A_{01\ldots d-1} \equiv \int A_d \,, \tag{B.4.4}$$

the transformation of the tensor offsetting that of the measure. Because of the antisymmetry, one must specify an orientation. Similarly, a p-form can be integrated over any p-dimensional submanifold. For a manifold with boundary one has Stokes's theorem,

$$\int_{\mathcal{M}} dA_{p-1} = \int_{\partial\mathcal{M}} A_{p-1} \tag{B.4.5}$$

where $\mathcal{M}$ is p-dimensional.

None of the above constructions requires a metric. In particular d contains only the ordinary derivative, but it is invariant due to the antisymmetry. One construction that does require a metric is the Poincaré dual. It is defined as

$$*A_{\mu_1\ldots\mu_{d-p}} = \frac{1}{p!} \epsilon_{\mu_1\ldots\mu_{d-p}}{}^{\nu_1\ldots\nu_p} A_{\nu_1\ldots\nu_p} \,. \tag{B.4.6}$$

The Levi–Civita symbol $\epsilon_{\mu_1\ldots\mu_d}$ is defined to transform as a tensor. Thus with all lower indices its components are $\pm(-G)^{1/2}$ and 0, while with all upper indices its components are $\pm(-G)^{-1/2}$ and 0. One can check that

on a p-form,

$$** = (-1)^{p(d-p)+1} \,, \tag{B.4.7}$$

the $+1$ coming from the Minkowski signature.

One can also represent the above by introducing an algebra of d anticommuting *differentials* dx^μ, writing

$$A_p = \frac{1}{p!} A_{\mu_1\ldots\mu_p} dx^{\mu_1}\ldots dx^{\mu_p} \,. \tag{B.4.8}$$

The factorial just offsets the sum over permutations so that each independent component appears once. The product of a p-form A and a q-form B is then (B.4.1), and the exterior derivative is $d = dx^\nu \partial_\nu$.

In this notation, an Abelian field strength, vector potential, and gauge transformation are written

$$F_2 = dA_1 \,, \quad \delta A_1 = d\lambda \,. \tag{B.4.9}$$

In the non-Abelian case, writing the fields as matrices, these become

$$F_2 = dA_1 + A_1 \wedge A_1 \equiv dA_1 + A_1^2 \,, \quad \delta A_1 = d\lambda + A_1\lambda - \lambda A_1 \,. \tag{B.4.10}$$

In the Abelian case there is a straightforward generalization to a p-form gauge transformation

$$F_{p+2} = dA_{p+1} \,, \quad \delta A_{p+1} = d\lambda_p \,. \tag{B.4.11}$$

The action is

$$-\frac{1}{2}\int d^dx\,(-G)^{1/2}|F_{p+2}|^2 = -\frac{1}{2}\int d^dx\,\frac{(-G)^{1/2}}{(p+2)!} F_{\mu_1\ldots\mu_{p+2}} F^{\mu_1\ldots\mu_{p+2}} \,. \tag{B.4.12}$$

A given component, say $A_{1\ldots p+1}$, then appears with the canonical normalization for a real scalar, $-\frac{1}{2}\partial_\mu A_{1\ldots p+1}\partial^\mu A_{1\ldots p+1}$. There is no straightforward non-Abelian generalization. For $p = -1$, the gauge invariance is trivial and this describes a massless scalar.

Using the gauge invariance (B.4.11), we can set $n^\mu A_{\mu\nu_1\ldots\nu_p} = 0$. The field equation then also implies $k^\mu A_{\mu\nu_1\ldots\nu_p} = 0$ and $k^2 = 0$. The potential A_{p+1} thus gives rise to a massless particle in the representation $[p+1]$ of the spin $SO(d-2)$.

Since $[p+1] = [d-p-3]$ for $SO(d-2)$, a $(p+1)$-form potential and a $(d-p-3)$-form potential describe the same particle states. For $d = 4$ and $p = 1$, this is the familiar fact that $B_{\mu\nu}$ describes the axion. We can also show this at the level of the fields. The Bianchi identity from $F_{p+2} = dA_{p+1}$ and the equation of motion from the action (B.4.12) are

$$dF_{p+2} = 0 \,, \quad d*F_{p+2} = 0 \,. \tag{B.4.13}$$

There is an obvious symmetry here: defining

$$F'_{d-p-2} = *F_{p+2} \tag{B.4.14}$$

simply switches the field equation and Bianchi identity, and in particular one can solve the new Bianchi identity in terms of a new potential A'_{d-p-3}, where $dA'_{d-p-3} = F'_{d-p-2}$. These theories are therefore equivalent, and one need consider only potentials of rank up to $\frac{1}{2}d - 1$. Again note that it is the field strength, not the potential, that is dualized. One can also see the equivalence in the action:

$$-\frac{1}{2}\int d^dx\,(-G)^{1/2}|dA_{p+1}|^2 \to -\frac{1}{2}\int d^dx\,(-G)^{1/2}|F_{p+2}|^2 + \int A'_{d-p-3}\wedge dF_{p+2} \to -\frac{1}{2}\int d^dx\,(-G)^{1/2}|dA'_{d-p-3}|^2\,. \quad \text{(B.4.15)}$$

In the first action the potential A_{p+1} is the variable of integration. In the second, F_{p+2} is the variable of integration; the Bianchi identity is no longer automatic so a Lagrange multiplier A'_{d-p-3} has been introduced to enforce it. In the final form the original F_{p+2} has been integrated out, leaving a gauge action for A'_{d-p-3}. In $d = 4$, this is electric–magnetic duality of Maxwell's equations. In $d = 3$, it implies that a vector potential is equivalent to a massless scalar. In $d = 2$ a massless scalar is equivalent to a dual scalar; in fact, this is equivalent to the world-sheet T-duality $X \to X'$. Again, note that it is the field strength to which the Poincaré duality is applied, not the potential.

For $d = 2$ mod 4, where $*^2 = 1$ on $(d/2)$-forms, it is consistent with the field equation and Bianchi identity to impose one of

$$F_{d/2} = \pm *F_{d/2}\,. \quad \text{(B.4.16)}$$

These are consistent theories with half as many components. In $d = 2$ they correspond to the left- or right-moving parts of a massless scalar. The action (B.4.12) no longer gives the field equation, as

$$|F_{d/2}|^2 = \pm F_{d/2}\wedge F_{d/2} = \pm d(A_{d/2-1}\wedge F_{d/2}) \quad \text{(B.4.17)}$$

is a total derivative. There are more complicated actions which are not manifestly covariant.

B.5 Thirty-two supersymmetries

We now begin a survey of some of the supersymmetric theories that arise as low energy limits in string theory. A more complete treatment can be found in the references.

$d = 11$ *supergravity*

In four dimensions the largest supersymmetry algebra, $N = 8$, contains 32 supercharges. This same limit holds in higher dimensions, since we could reduce to four by compactifying on tori. Table B.1 then implies that $d = 11$ is the maximum in which supersymmetry can exist,[3] since the spinor representations are too large for $d \geq 12$. Although this exceeds by one the critical dimension of superstring theory, we will start with this case.

The Majorana spinor supercharge again satisfies the algebra

$$\{Q_\alpha, \bar{Q}_\beta\} = -2P_\mu\Gamma^\mu_{\alpha\beta}\,. \quad \text{(B.5.1)}$$

The massless irreducible representation contains $2^8 = 256$ states, half fermions and half bosons. By calculating the spins $S_1,\ldots,S_4$ one finds that the graviton multiplet contains two bosonic representations of $SO(9)$: a traceless symmetric tensor (the graviton) with $\frac{1}{2}\times 9\times 10 - 1 = 44$ components and a completely antisymmetric three-index tensor with $9\times 8\times 7/3! = 84$ components for 128 in all. There is a single fermionic vector-spinor representation. The spinor index takes 16 values and the vector 9 values; 16 components vanish by a trace condition as in eq. (10.5.19), leaving $16\times 9 - 16 = 128$ fermionic components.

With two or fewer derivatives there is a unique supersymmetric action, whose bosonic part is

$$S_{11} = \frac{1}{2\kappa^2}\int d^{11}x\,(-G)^{1/2}\left(R - \frac{1}{2}|F_4|^2\right) - \frac{1}{12\kappa^2}\int A_3\wedge F_4\wedge F_4 \quad \text{(B.5.2)}$$

with A_3 a 3-form potential and F_4 its 4-form field strength. The final *Chern–Simons* term is gauge-invariant in spite of the explicit appearance of A_3 because the term from the variation $\delta A_3 = d\lambda_2$ vanishes by parts.

$d = 10$ *IIA supergravity*

By compactifying the $d = 11$ theory on a torus and keeping only the massless fields (*dimensional reduction*), we obtain a $d = 10$ theory with 32 supercharges. The $d = 11$ Majorana spinor becomes a $d = 10$ Majorana spinor, which reduces to one Majorana–Weyl spinor of each chirality,

$$Q^1_\alpha \in \mathbf{16}\,, \quad Q^2_\alpha \in \mathbf{16'}\,. \quad \text{(B.5.3)}$$

The product of two spinors of the same chirality contains a vector, while the product of spinors of opposite chirality contains a scalar (eq. (B.1.38)),

[3] With *two* timelike dimensions a Majorana–Weyl spinor with 32 components is allowed at $d = 12$, but we will not try to figure out what this might mean.

so from the eleven-dimensional algebra we deduce

$$\{Q^1_\alpha, \bar{Q}^1_\beta\} = -2P_\mu(\mathrm{P}_+\Gamma^\mu)_{\alpha\beta}\ , \quad \{Q^2_\alpha, \bar{Q}^2_\beta\} = -2P_\mu(\mathrm{P}_-\Gamma^\mu)_{\alpha\beta}\ , \tag{B.5.4a}$$

$$\{Q^1_\alpha, \bar{Q}^2_\beta\} = -2P_{10}(\mathrm{P}_+\Gamma)_{\alpha\beta}\ . \tag{B.5.4b}$$

Here $\Gamma = \Gamma^{10}$ is from the toroidal dimension. A notable feature is the appearance of a central charge proportional to the Kaluza–Klein momentum. This is one of the ways that a central charge in the supersymmetry algebra can arise; additional central charges carried by extended objects are introduced in their physical context in section 13.2.

The dimensional reduction of the $d = 11$ theory leaves a scalar from $G_{10\,10}$, a Kaluza–Klein vector from $G_{\mu 10}$, a 2-form potential from $B_{\mu\nu 10}$ and a 3-form from $B_{\mu\nu\sigma}$. This is the same as the massless content of the IIA superstring, the scalar dilaton and the 2-form being from the NS–NS sector and the 1- and 3-forms from the R–R sector. This is no surprise because the large amount of supersymmetry determines the massless particle content completely. What is a surprise is that there really is an eleventh dimension hidden in the IIA string, invisible in perturbation theory but visible at strong coupling. This is discussed in chapter 14.

The action can be obtained by dimensional reduction; further details are given in section 12.1.

$d = 10$ *IIB supergravity*

There is another ten-dimensional supergravity, which is not obtained by compactifying an eleven-dimensional theory. This has two supercharges of the same chirality, which we can define to be **16**. The algebra is

$$\{Q^A_\alpha, \bar{Q}^B_\beta\} = -2\delta^{AB}P_\mu(\mathrm{P}_+\Gamma^\mu)_{\alpha\beta}\ . \tag{B.5.5}$$

The graviton multiplet contains two scalars, the traceless symmetric graviton, two antisymmetric 2-forms, and a 4-form with self-dual field strength, for

$$2 + 35 + 28 + 28 + 35 = 128 \tag{B.5.6}$$

bosonic states in all. This is the same as the massless content of the IIB superstring. More details are given in chapters 10 and 12.

$d < 10$ *supergravity*

The supergravities with 32 supercharges in $d < 10$ can be obtained by dimensional reduction of the IIA string, or equivalently of $d = 11$ supergravity. In this section we discuss some of the main features; this subject is relevant in particular to section 14.2. We need not consider the IIB string separately below $d = 10$, because after compactification on a circle it is T-dual to the IIA string (chapter 13).

Table B.3. *Supergravities with 32 supercharges. The group G is a symmetry of the low energy supergravity theory, and the moduli space is locally G/H.*

d	scalars	vectors	G	H
10A	1	1	$SO(1,1,\mathbf{R})$	-
10B	2	0	$SL(2,\mathbf{R})$	$SO(2,\mathbf{R})$
9	3	3	$SL(2,\mathbf{R}) \times SO(1,1,\mathbf{R})$	$SO(2,\mathbf{R})$
8	7	6	$SL(2,\mathbf{R}) \times SL(3,\mathbf{R})$	$SO(2,\mathbf{R}) \times SO(3,\mathbf{R})$
7	14	10	$SL(5,\mathbf{R})$	$SO(5,\mathbf{R})$
6	25	16	$SO(5,5,\mathbf{R})$	$SO(5,\mathbf{R}) \times SO(5,\mathbf{R})$
5	42	27	$E_{6(6)}(\mathbf{R})$	$Sp(4,\mathbf{R})$
4	70	28	$E_{7(7)}(\mathbf{R})$	$SU(8,\mathbf{R})$
3	128	-	$E_{8(8)}(\mathbf{R})$	$SO(16,\mathbf{R})$

The first issue we wish to consider is the number of scalars. Compactifying k of the dimensions of $d = 11$ supergravity, there are

$$\frac{1}{2}k(k+1) \tag{B.5.7}$$

scalars from G_{mn} and

$$\frac{1}{3!}k(k-1)(k-2) \tag{B.5.8}$$

from B_{mnp}. Again m, n, p are compactified and μ, ν are noncompact. In addition, in $d = 5$, the Poincaré dual $*(H_{\mu\nu\rho\sigma})$ gives the field strength (gradient) for an extra scalar, as discussed at the end of section B.4. In $d = 4$, $*(H_{\mu\nu\rho m})$ gives 7 extra scalars. In $d = 3$, $*(H_{\mu\nu mn})$ gives $\frac{1}{2}8 \times 7 = 28$ extra scalars. Also in $d = 3$ the duals of the 8 Kaluza–Klein vectors give additional scalars. The total number is indicated in table B.3.

The second issue is the number of vectors: k from $G_{\mu n}$ and $\frac{1}{2}k(k-1)$ from $B_{\mu mn}$. In addition there is one in $d = 6$ from $*(H_{\mu\nu\rho\sigma})$ and six in $d = 5$ from $*(H_{\mu\nu\rho m})$. In $d = 4$, $*(H_{\mu\nu mn})$ is just the magnetic description of the $B_{\mu mn}$ vectors, and there are no vectors in $d = 3$ because we have converted them all to scalars by Poincaré duality. The results are summarized in the second column of the table. The gauge group is $U(1)^{n_V}$.

Third, there is no potential — the scalars are moduli — and the moduli space metric is completely determined by symmetry. The moduli spaces are cosets G/H, as listed in the table. The structure is the same as in the toroidal example in section 8.4, a coset of a noncompact group by a compact group. In the string case there was a further identification by the discrete T-duality group. This discrete identification does not affect the

local structure of moduli space, and in particular not the effective action, and so it is not determined at this point by supersymmetry. Rather, it is determined by short-distance physics as is described in chapter 14. In the bosonic case the dilaton was decoupled, giving a separate space $SO(1,1,\mathbf{R}) = \mathbf{R}$. Below $d = 9$ in table B.3 it combines with other moduli into a larger homogeneous space.

In each case, the noncompact group in the numerator is a global symmetry of the supergravity theory, and the compact group in the denominator is the unbroken symmetry at any point in moduli space. The notation $E_{n(n)}(\mathbf{R})$ refers to an exceptional group with some sign changes in the algebra to make it noncompact, just as $SO(n,m,\mathbf{R})$ is related to $SO(n+m,\mathbf{R})$ and $SL(n,\mathbf{R})$ to $SU(n)$. Also, we adhere to what is currently the common usage in defining $Sp(k,\mathbf{R})$ to be the group of $2k$-dimensional symplectic matrices. The details of table B.3 are not at this point important, but it is interesting to see in chapter 14 how the structure fits into string theory.

For $d = 4$, the count in the table agrees with the $N = 8$ multiplet. To dimensionally reduce the supersymmetry algebra, separate the 11-dimensional 32-valued spinor index into a 4-valued $SO(3,1)$ index α and an 8-valued $SO(7)$ index A. The 11-dimensional algebra (B.5.1) becomes

$$\{Q^A_\alpha, \bar{Q}^B_\beta\} = -2P_\mu \delta^{AB}\Gamma^\mu_{\alpha\beta} - 2P_m \Gamma^{mAB}_{(7)}\Gamma_{(4)\alpha\beta} \,. \tag{B.5.9}$$

Here Γ^m factors into $\Gamma^{mAB}_{(7)}\Gamma_{(4)\alpha\beta}$ with $\Gamma^{mAB}_{(7)}$ being $SO(7)$ Γ matrices. The factor of

$$\Gamma_{(4)\alpha\beta} = i(\Gamma^0\Gamma^1\Gamma^2\Gamma^3)_{\alpha\beta} \tag{B.5.10}$$

must appear because Γ^m anticommutes with Γ^μ. Again, a central charge has arisen from the compact momenta. Since

$$(P_m\Gamma^m P_n\Gamma^n)^{AB} = \delta^{AB}P_m P^m \,, \tag{B.5.11}$$

the eigenvalues q_i of the central charge are all equal and any BPS multiplet will be ultrashort, with the same 256 states as a massless multiplet. In this case there is a simple explanation. The BPS condition is $-P_\mu P^\mu = P_m P^m$, so a BPS multiplet is actually a massless multiplet from the higher-dimensional point of view.

The $d = 4$, $N = 8$ theory has 28 gauge bosons, but only the 7 Kaluza–Klein charges appear in the dimensionally reduced algebra (B.5.9). In fact, the full algebra contains all 28 gauge charges, the remainder arising from the extended-object charges in higher dimensions. The antisymmetric matrix Z^{AB} has precisely 28 components, so in general all are nonzero. The gauge charges can be organized into an antisymmetric matrix P^{AB} so that the algebra is

$$\{Q^A_\alpha, \bar{Q}^B_\beta\} = -2P_\mu \delta^{AB}\Gamma^\mu_{\alpha\beta} - 2P^{AB}\delta_{\alpha\beta} \,. \tag{B.5.12}$$

It should be noted that the compact momenta depend on the moduli, for example $p = n/R$ for compactification on a single circle. When the central charge is written in terms of integer charges such as n, it has explicit dependence on the moduli.

B.6 Sixteen supersymmetries

$d = 10$, $N = 1$ *(type I) supergravity*

This algebra has a single Majorana–Weyl **16** supercharge. The massless vector representation has 16 states, $\mathbf{8}_v + \mathbf{8}'$ under the $SO(8)$ little group. The supergravity multiplet is $\mathbf{8}_v \times (\mathbf{8}_v + \mathbf{8}')$ as found in the bosonic and type I strings. The bosonic content is then a graviton, an antisymmetric tensor, and a dilaton from the supergravity multiplet plus $\dim g$ vectors from the gauge multiplets, g being the gauge group. The bosonic action is given in section 12.1. The action is classically invariant for any g, but as discussed in section 12.2 there are anomalies unless $g = SO(32)$ or $E_8 \times E_8$.

$d < 10$ *supergravity*

Toroidal compactification of k dimensions gives supergravity with 16 supersymmetries in $d = 10 - k$. There are a total of $k(k+r)+1$ moduli, r being the rank of the ten-dimensional gauge group g. The metric gives rise to $\frac{1}{2}k(k+1)$ moduli, the antisymmetric tensor to $\frac{1}{2}k(k-1)$, the Wilson lines to kr, and the original ten-dimensional dilaton to the final one. Of course g is $SO(32)$ or $E_8 \times E_8$, both having $r = 16$, in a consistent ten-dimensional theory, but here we are just using this as a trick to generate theories in lower dimensions. The reduced theories are parity-symmetric and have no anomalies, and so can have any g. In fact, various $r < 16$ theories can be obtained in string theory by slightly more complicated compactifications. The moduli space is as given explicitly by the Narain compactification of the heterotic string,

$$\frac{SO(1,1,\mathbf{R}) \times SO(k+r,k,\mathbf{R})}{SO(k+r,\mathbf{R}) \times SO(k,\mathbf{R})} \,, \tag{B.6.1}$$

the $SO(1,1,\mathbf{R})$ being from the dilaton. In $d = 4$ the antisymmetric tensor gives another scalar, the axion, via Poincaré duality; this combines with the dilaton to form $SL(2,\mathbf{R})/SO(2,\mathbf{R})$. In $d = 3$ ($k = 7$), the Poincaré duals of the $14 + r$ vectors combine with the dilaton and the other moduli to enlarge the moduli space (B.6.1) to

$$\frac{SO(8+r,8,\mathbf{R})}{SO(8+r,\mathbf{R}) \times SO(8,\mathbf{R})} \,. \tag{B.6.2}$$

Toroidal compactification gives gauge group $U(1)^{2k+r}$ at generic points in moduli space, the original gauge group being broken to $U(1)^r$ by Wilson lines. At special points it will be enhanced to various non-Abelian groups; the rank remains $2k + r$.

$d = 6$, $N = 2$ *supersymmetry*

Under $SO(9,1) \to SO(5,1) \times SO(4)$, the ten-dimensional $N = 1$ supersymmetry decomposes

$$\mathbf{16} \to (\mathbf{4},\mathbf{2}) + (\mathbf{4}',\mathbf{2}') . \tag{B.6.3}$$

The $(\mathbf{4},\mathbf{2})$ has eight real components, forming a single complex **4**. The **4** cannot have a Majorana condition imposed so the complex **4** is the smallest algebra in $d = 6$. The dimensionally reduced algebra is $d = 6$ $(1,1)$ supersymmetry, one supercharge in the **4** and one in the **4′**. The only representation with spins ≤ 1 is the vector, which is the dimensional reduction of the $d = 10$ vector and so consists of one vector and four scalars.

Decomposing $SO(5,1)$ into the $SO(1,1)$ of the (0,1)-plane and the transverse $SO(4)$, the Weyl spinor supercharges decompose

$$\mathbf{4} \to (+\tfrac{1}{2},\mathbf{2}) + (-\tfrac{1}{2},\mathbf{2}') = (+\tfrac{1}{2},\tfrac{1}{2},0) + (-\tfrac{1}{2},0,\tfrac{1}{2}) , \tag{B.6.4a}$$

$$\mathbf{4}' \to (+\tfrac{1}{2},\mathbf{2}') + (-\tfrac{1}{2},\mathbf{2}) = (+\tfrac{1}{2},0,\tfrac{1}{2}) + (-\tfrac{1}{2},\tfrac{1}{2},0) . \tag{B.6.4b}$$

These are complex representations, so their adjoints are independent operators. The representations of $SO(1,1)$ are all one-dimensional and are labeled by the helicity S_0. In the second equality of each line we have used the relation $SO(4) = SU(2) \times SU(2)$ and labeled the $SU(2)$ representations by their spin j, so the notation is (s_0, j_1, j_2). As in section B.1, the $s_0 = -\frac{1}{2}$ generators annihilate the massless states. The latter then form a representation of the generators with $s_0 = \frac{1}{2}$, these being Q_α and Q'_β in the $(\frac{1}{2},0)$ and $(0,\frac{1}{2})$ of $SU(2) \times SU(2)$. Treating these as lowering operators and their adjoints as raising operators, by taking all combinations of raising operators one obtains the representations

$$r = (\tfrac{1}{2},\tfrac{1}{2}) + (\tfrac{1}{2},0)^2 + (0,\tfrac{1}{2})^2 + (0,0)^4 . \tag{B.6.5}$$

Starting from an $SU(2) \times SU(2)$ multiplet $|j_1, j_2\rangle$ annihilated by the lowering operators, the raising operators generate the representations

$$r \times |j_1, j_2\rangle . \tag{B.6.6}$$

The supergravity multiplet is built on $|\frac{1}{2},\frac{1}{2}\rangle$, giving the states

$$\begin{aligned} &|1,1\rangle + |1,0\rangle + |0,1\rangle + |0,0\rangle + |\tfrac{1}{2},\tfrac{1}{2}\rangle^4 \\ &+ |1,\tfrac{1}{2}\rangle^2 + |\tfrac{1}{2},1\rangle^2 + |0,\tfrac{1}{2}\rangle^2 + |\tfrac{1}{2},0\rangle^2 . \end{aligned} \tag{B.6.7}$$

The bosonic content, in the first line, is a graviton, an antisymmetric tensor, a scalar, and four vectors. The vector multiplet is built on $|0,0\rangle$, giving

$$|\tfrac{1}{2},\tfrac{1}{2}\rangle + |0,0\rangle^4 + |\tfrac{1}{2},0\rangle^2 + |0,\tfrac{1}{2}\rangle^2 . \tag{B.6.8}$$

There is a second $d = 6$ algebra with 16 supercharges, the $(2,0)$ algebra with two complex **4** supercharges. The raising operators now form the representations

$$r' = (1,0) + (\tfrac{1}{2},0)^4 + (0,0)^5 . \tag{B.6.9}$$

Acting on $|0,1\rangle$, these produce the supergravity multiplet

$$|1,1\rangle + |\tfrac{1}{2},1\rangle^4 + |0,1\rangle^5 , \tag{B.6.10}$$

whose bosonic content is a graviton and five anti-self-dual antisymmetric tensors. Acting on $|0,0\rangle$ they produce the tensor multiplet

$$|1,0\rangle + |\tfrac{1}{2},0\rangle^4 + |0,0\rangle^5 , \tag{B.6.11}$$

with self-dual antisymmetric tensor and five scalars.

$d = 4$, $N = 4$ *gauge theory*

The four-dimensional $N = 4$ algebra is

$$\{Q^A_\alpha, \bar{Q}^B_\beta\} = -2P_\mu \delta^{AB} \Gamma^\mu_{\alpha\beta} - 2P_{Rm} \Gamma^{mAB} \delta_{\alpha\beta} . \tag{B.6.12}$$

In this case only six of the gauge charges appear; in the heterotic string these are the ones coming from right-moving currents.

We now consider the effective renormalizable theory near a point of non-Abelian symmetry h. It will be useful to derive the full action and SUSY transformation by dimensional reduction of ten-dimensional supersymmetric Yang–Mills theory, whose Lagrangian density is

$$-\frac{1}{4g^2}\mathrm{Tr}(F_{MN}F^{MN}) - \frac{i}{2g^2}\mathrm{Tr}(\bar{\lambda}\Gamma^M D_M \lambda) . \tag{B.6.13}$$

The gauge field and the gaugino λ (a Majorana–Weyl **16**) are written in matrix notation, and M, N run from 0 to 9. The supersymmetry transfor-

mation is

$$\delta A_M = -i\bar{\zeta}\Gamma_M\lambda\,, \tag{B.6.14a}$$

$$\delta\lambda = \frac{1}{2}F_{MN}\Gamma^{MN}\zeta\,. \tag{B.6.14b}$$

Reducing $M \to \mu, m$, the Lagrangian density becomes

$$-\frac{1}{4g^2}\mathrm{Tr}\left(F_{\mu\nu}F^{\mu\nu} + 2D_\mu A_m D^\mu A_m - [A_m, A_n]^2\right)$$

$$-\frac{i}{2g^2}\mathrm{Tr}(\bar{\lambda}\Gamma^\mu D_\mu\lambda + i\bar{\lambda}\Gamma_m[A_m,\lambda])\,. \tag{B.6.15}$$

The six compact components of the gauge field become the six scalars A_m of the $N = 4$ vector multiplet. The **16** index separates into $(\mathbf{2},\mathbf{4}) + (\bar{\mathbf{2}},\bar{\mathbf{4}})$ under $SO(3,1) \times SO(6)$, so the ten-dimensional spinor becomes four Weyl spinors. Similarly the transformation laws reduce to

$$\delta A_\mu = -i\bar{\zeta}\Gamma_\mu\lambda \tag{B.6.16a}$$

$$\delta A_m = -i\bar{\zeta}\Gamma_m\lambda \tag{B.6.16b}$$

$$\delta\lambda = \left(\frac{1}{2}F_{\mu\nu}\Gamma^{\mu\nu} + D_\mu\phi_n\Gamma^{\mu n} + \frac{i}{2}[A_m, A_n]\Gamma^{mn}\right)\zeta\,. \tag{B.6.16c}$$

The potential

$$V = -\frac{1}{4g^2}\mathrm{Tr}\left([A_m, A_n]^2\right) \tag{B.6.17}$$

is nonnegative and vanishes only if $[A_m, A_n] = 0$ for all m, n. Thus, in the flat directions the A_m can be taken simultaneously diagonal, and the moduli are just the $6\,\mathrm{rank}(h)$ eigenvalues. At generic points the group is broken to $U(1)^{\mathrm{rank}(h)}$. We have seen this potential before, in eq. (8.7.11) for the D-brane moduli. This is no accident, as the T-duality that produces the D-brane has the effect of dimensionally reducing the open string Yang–Mills action.

Eq. (B.6.15) is the most general renormalizable action consistent with $N = 4$ global supersymmetry. It remains the most general action if we adopt the looser moduli space power counting described in section B.2, which would have allowed field-dependent kinetic terms. In other words, $N = 4$ global supersymmetry requires the moduli space to be flat. This is no contradiction with the curved moduli space (B.6.1) found in supergravity. The only scale there is the Planck scale, so the dimensionless variable is κA and the nonlinearities vanish in the limit $\kappa \to 0$ where we ignore gravity.

The $N = 4$ Yang–Mills theory has a number of interesting properties, the first being that its beta function vanishes identically — the coupling does not run. Unlike most gauge theories, different values of g really give different theories, rather than being transmuted to a change of scale. It is easy to prove this statement, not just to all orders of perturbation theory but exactly. Consider for simplicity $h = SU(2)$. Generically, $A_m \propto \sigma^3$ breaks $SU(2)$ to $U(1)$ at a scale

$$v = (A_m A_m)^{1/2} \equiv A\,, \tag{B.6.18}$$

so the massless theory contains only an Abelian vector multiplet. Consider the gauge field kinetic term in the effective action. Its coefficient is $-1/4g^2$, but if the coupling runs in $SU(2)$ we must figure out at what scale to evaluate g. The answer is v, because this is where $SU(2)$ breaks and the coupling stops running. The scale v depends on the massless moduli, so what we really have is an effective Lagrangian density

$$-\frac{1}{4g^2(A)}F_{\mu\nu}F^{\mu\nu}\,. \tag{B.6.19}$$

However, this is a field-dependent kinetic term, which we have just stated is inconsistent with $N = 4$ supersymmetry — unless in fact the coupling is independent of scale as claimed. This argument is typical of the recent analysis of supersymmetric gauge theories, but is particularly simple because of the large amount of supersymmetry.

B.7 Eight supersymmetries

$d = 6$, $N = 1$ supersymmetry

We start in six dimensions, the maximum in which a spinor with eight components is allowed according to table B.1. We obtain the massless representations as in eq. (B.6.6), where now

$$r'' = (\tfrac{1}{2}, 0) + (0,0)^2\,. \tag{B.7.1}$$

The supergravity multiplet, built on $|\frac{1}{2},1\rangle$, is

$$|1,1\rangle + |\tfrac{1}{2},1\rangle^2 + |0,1\rangle \tag{B.7.2}$$

containing the graviton, gravitino (which requires two copies of $|\frac{1}{2},1\rangle$), and the $(0,1)$ which is an anti-self-dual 2-form. The other relevant multiplets are built on $|0,0\rangle$, $|0,\frac{1}{2}\rangle$, and $|\frac{1}{2},0\rangle$, giving

half-hypermultiplet:	$\|\frac{1}{2},0\rangle\,,\ \|0,0\rangle^2\,,$
vector multiplet:	$\|\frac{1}{2},\frac{1}{2}\rangle\,,\ \|0,\frac{1}{2}\rangle^2\,,$
tensor multiplet:	$\|1,0\rangle\,,\ \|\frac{1}{2},0\rangle^2\,,\ \|0,0\rangle\,.$

The respective bosonic content is: two scalars; a vector; a self-dual tensor plus scalar.

A general theory will have some number of vector, hyper-, and tensor multiplets. We describe the general bosonic Lagrangian, first in the most restrictive form of keeping only terms that would be renormalizable when reduced to four dimensions. The Lagrangian consists of gauge-invariant kinetic terms for the various fields and a potential for the hypermultiplets. The tensors must be neutral under the gauge group, so in this limit the tensor representation is decoupled from all other fields. To write the potential, we collect two half-hypermultiplets into a complex doublet of scalars Φ^i_α, with α the doublet index and i labeling the hypermultiplets.[4] The $N = 2$ D-term is

$$D^{Aa} = \frac{g^2}{2}\Phi^{i*}_\alpha \sigma^A_{\alpha\beta} t^a_{ij} \Phi^j_\beta \,, \tag{B.7.3}$$

with $\sigma^A_{\alpha\beta}$ the Pauli matrices ($A = 1,2,3$) and t^a_{ij} the group representation. The potential, determined entirely by the gauge symmetry, is

$$\frac{1}{2g^2} D^{Aa} D^{Aa} \,. \tag{B.7.4}$$

The interactions, incidentally, are nonrenormalizable in six dimensions.

Now consider the less restrictive moduli space action, where field-dependent kinetic terms are included, but with gravity still decoupled. Supersymmetry does not allow the gauge field kinetic term to depend on the hypermultiplet moduli, and it is allowed to depend on the scalar t in the tensor multiplets only in the precise form

$$t\mathrm{Tr}(F_{\mu\nu}F^{\mu\nu}) \,. \tag{B.7.5}$$

The linear dependence is fixed because this term is related by supersymmetry to a coupling of the self-dual tensor,

$$B_2 \mathrm{Tr}(F_2 \wedge F_2) \,, \tag{B.7.6}$$

where the tensor gauge invariance allows only the linear coupling. The term (B.7.6) is needed to cancel anomalies in a six-dimensional version of the Green–Schwarz mechanism, so these terms can arise only at exactly one loop, with a coefficient that is determined by the gauge quantum numbers of the hyper- and vector multiplets.

The hypermultiplet kinetic term may depend on the hypermultiplet moduli but not the tensor moduli. Representing the moduli by real fields ϕ^r, one has

$$G_{rs}(\phi)\partial_\mu \phi^r \partial^\mu \phi^s \,. \tag{B.7.7}$$

[4] If the scalars are in a pseudoreal representation of the gauge group, meaning that the conjugation matrix C_{ij} is antisymmetric, one can reduce to a half-hypermultiplet by the reality condition $\Phi^{i*}_\alpha = \epsilon_{\alpha\beta} C_{ij} \Phi^j_\beta$.

For the case of $d = 4$, $N = 1$ supersymmetry, we have explained in section B.2 that the moduli space must be Kähler, meaning that the $2n$ real moduli can be grouped into n complex fields ϕ^i with the metric $G_{\bar{i}j} = K_{\bar{i}ij}$. Here the supersymmetry is doubled and the metric correspondingly more restricted: it must be *hyper-Kähler*. This means that there are three different complex Kähler structures, three different ways to group the real moduli into complex fields, each giving a Kähler metric. Further there is a relation between the three complex structures. Given any complex structure, a set of complex coordinates ϕ^i and ϕ^{i*}, we can define the tensor J by

$$J^i_j = i\delta^i_j, \quad J^{\bar{i}}_{\bar{j}} = -i\delta^{\bar{i}}_{\bar{j}}, \quad J^i_{\bar{j}} = J^{\bar{i}}_j = 0 \,. \tag{B.7.8}$$

This tensor can be defined for any complex manifold and is also known as the *complex structure*. We have defined it in a particular coordinate system but now can translate it to arbitrary coordinates. It satisfies $J^2 = -1$, a frame-independent statement. The three complex structures of the hyper-Kähler space are required to satisfy

$$J^A J^B = -\delta^{AB} + \epsilon^{ABC} J^C \,. \tag{B.7.9}$$

These properties require the number of real moduli to be a multiple of 4.

An alternative characterization is as follows. There are $4m$ moduli, so a general metric (B.7.7) would have holonomy $SO(4m)$. That is, parallel transport of a vector around a loop in moduli space brings it back to itself rotated by a general element of $SO(4m)$. These ideas are familiar from general relativity, in the context of the spacetime manifold, but we emphasize that the manifold in question here is field space. Now consider the following $SU(2)$ subgroup of $SO(4m)$. We know that $SO(4) = SU(2) \times SU(2)$. Take the first $SU(2)$ and replace the elements with $m \times m$ identity matrices to make a subgroup of $SO(4m)$. The subgroup of $SO(4m)$ that commutes with this $SU(2)$ is $Sp(m)$. Then a hyper-Kähler manifold is one for which the holonomy lies in this $Sp(m)$ subgroup, the Js being the $SU(2)$ generators.

$d = 4$, $N = 2$ *supersymmetry*

Let us first consider the reduction of the self-dual tensor multiplet from $d = 6$ to $d = 5$. The components $B_{\mu 5}$ become a vector. The dual $*H_{\sigma\rho\omega}$ would give the field strength of a second vector if the tensor were unconstrained, but due to the self-duality this is the same as $H_{\mu\nu 5}$. Thus one has in all a vector and a scalar. This is the same as the content of the vector multiplet, where the scalar comes from the reduction of A_5, so these multiplets are identical in $d = 5$ and consequently in $d = 4$.

Thus in $d = 4$ we need consider only the hypermultiplet with four

scalars, and the vector multiplet with two scalars A_4, A_5 from the six-dimensional vector. First in the renormalizable limit, the action includes the bosonic terms discussed in $d = 6$, the gauge-invariant kinetic terms and the potential (B.7.4). The potential has additional terms

$$-\frac{1}{4g^2}\mathrm{Tr}([A_4, A_5]^2) \tag{B.7.10}$$

from reduction of the field strength and

$$\Phi_\alpha^{i\dagger}(M_4^2 + M_5^2)_{ij}\Phi_\alpha^j , \tag{B.7.11}$$

where

$$M_{mij} = A_m^a t_{ij}^a + q_{mij} . \tag{B.7.12}$$

The first term in M is from reduction of the covariant derivative; the parameters q_{mij} are allowed by supersymmetry and can be thought of as arising from a 'dummy' gauge field.

The potential has various flat directions. We discuss first the moduli space approximation with gravity still decoupled. Supersymmetry requires the kinetic term for the vector multiplet to depend only on the vector multiplet moduli and the kinetic term for the hypermultiplet to depend only on the hypermultiplet moduli. The latter is required to be a hyper-Kähler space just as in $d = 6$. The vector moduli space is also a Kähler metric with extra conditions. Namely, forming the complex scalars $A^a = A_4^a + iA_5^a$ with a indexing the gauge generators, the Kähler potential must be of the form

$$K(A, A^*) = \mathrm{Im}\left(\sum_a A^{a*}\partial_a F(A)\right) \tag{B.7.13}$$

for some holomorphic *prepotential* $F(A)$. The metric on moduli space is then

$$G_{a\bar{b}} = \mathrm{Im}(\partial_a \partial_b F) . \tag{B.7.14}$$

This is known as a *rigid special Kähler metric.*

Turning on $N = 2$ supergravity, the moduli space acquires additional curvature as it did for $N = 4$, but it remains a direct product of hypermultiplet and vector multiplet moduli spaces. The hyper-Kähler metrics are replaced by *quaternionic* metrics, where the $SU(2)$ holonomy is no longer zero but has a definite curvature of order κ^2. The vector moduli space becomes a *special Kähler space.* These spaces are also relevant to $N = 1$ compactifications of the heterotic string, so we describe them in some detail. For n hypermultiplets, it is useful to begin with $n+1$ complex

coordinates X^I with the projective identification

$$(X^0, X^1, \ldots, X^n) \cong (\lambda X^0, \lambda X^1, \ldots, \lambda X^n) \tag{B.7.15}$$

for any nonzero complex λ. One can also introduce invariant coordinates; for example, away from the subspace $X^0 = 0$ the set

$$T^A = \frac{X^A}{X^0} , \quad A = 1, \ldots, n . \tag{B.7.16}$$

The low energy action is determined by a single complex function $F(X)$, which must be homogeneous of degree 2 under the identification (B.7.15),

$$F(\lambda X) = \lambda^2 F(X) . \tag{B.7.17}$$

The Kähler potential is then

$$K = -\ln \mathrm{Im}\left(\sum_I X^{I*}\partial_I F(X)\right) . \tag{B.7.18}$$

Under a projective transformation (B.7.15),

$$K \to K - \ln\lambda - \ln\lambda^* . \tag{B.7.19}$$

This is a Kähler transformation (B.2.32), so the metric is well defined on the projective space produced by the identification. The number of vectors is $n+1$, including the one from the supergravity multiplet, so the fields A_μ^I for $I = 0, \ldots, n$ are independent. Their kinetic term is again determined by F and depends only on the vector multiplet moduli; its explicit form is left to the references.

The forms (B.7.15) and (B.7.18) are not invariant under arbitrary changes of coordinates: the coordinates X^I are known as *special coordinates.* The forms are clearly invariant under linear redefinitions of the special coordinates, but there is in fact a larger set of transformations that preserves the form, namely

$$\begin{bmatrix} X'^I \\ \partial_{I'}F' \end{bmatrix} = S \begin{bmatrix} X^I \\ \partial_I F \end{bmatrix} \tag{B.7.20}$$

for S a $2(n+1) \times 2(n+1)$ real symplectic matrix. As a final comment, in recent literature it has been noted that in special cases the symplectic transformation (B.7.20) gives a would-be gradient $\partial_{I'}F'$ whose curl is actually nonvanishing. For these the definition of special geometry needs to be generalized.

Exercises

B.1 In section B.1 we defined B and C in a particular basis. The properties (B.1.17) and (B.1.24) define them in general. Under a change of basis

$$\Gamma^\mu \to U\Gamma^\mu U^{-1}$$

for unitary U, find the transformations of B and C. Show that the properties (B.1.18), (B.1.19), (B.1.21), (B.1.25), and (B.1.27) are independent of such a change of basis. Determine the relations between B and B^T and between C and C^T, and show that these are independent of basis.

B.2 Extend the decomposition (B.1.44) to the general $SO(d-1,1) \to SO(d'-1,1) \times SO(d-d')$, where some of the dimensions are odd.

B.3 Work out the details of the reduction of the $d = 4$, $N = 1$ supersymmetry algebra to the $d = 2$ (2,2) algebra. Identify the central charges.

B.4 Verify eq. (B.4.7) for $*^2$ and derive the corresponding result for Euclidean space.

B.5 List the helicities (s_1, s_2, s_3, s_4) for the massless $\mathbf{8}_v + \mathbf{8}$ open string states and show that these constitute a representation of the type I supersymmetry algebra.

References

The literature on the subjects covered in these two volumes easily exceeds 5000 papers, and probably approaches twice that number. To represent accurately the contributions of all who have worked in these many areas is an undertaking beyond the scope of this book. Rather, I have tried to assemble a list that is short enough to be useful in developing specific subjects and in giving points of entry to the literature. The resulting list includes many review articles, plus papers from the original literature where I felt they were needed to supplement the treatment in the text.

Papers listed as 'e-print: hep-th/yymmnnn' are available electronically at the Los Alamos physics e-print archive,

http://xxx.lanl.gov/abs/hep-th/yymmnnn .

A list of corrections to the text is maintained at

http://www.itp.ucsb.edu/~joep/errata.html .

General references

Other books and lectures covering material in volume two include Green, Schwarz, & Witten (1987) (henceforth denoted GSW), Peskin (1987), Lüst & Theisen (1989), Alvarez-Gaumé & Vazquez-Mozo (1995), D'Hoker (1993), Ooguri & Yin (1997), and Kiritsis (1997). A number of the review articles cited are collected in Efthimiou & Greene (1997). Many of the key papers up to 1985 can be found in Schwarz (1985), and in the reference section of GSW.

Chapter 10

Many of the topics of this chapter are covered in the general references. In addition, Friedan, Martinec, & Shenker (1986) cover superconformal field

theory, bosonization, and vertex operators. The original article Gliozzi, Scherk, & Olive (1977) is quite readable. The review Schwarz (1982) covers many of the topics in this chapter. The cancellation of divergences in the $SO(32)$ type I theory is discussed in Green & Schwarz (1985); whereas we consider the vacuum amplitude, they study the *planar amplitude*, where there are vertex operators but on one boundary only.

Chapter 11

The classification of superconformal algebras is from Ademollo *et al.* (1976) and Sevrin, Troost, & Van Proeyen (1988). Strings based on higher spin (W) algebras are reviewed in Pope (1995). Strings based on fractional spin algebras are reviewed in Tye (1995). Strings based on the $N = 2$ superconformal algebra are reviewed and a spacetime interpretation given in Ooguri & Vafa (1991). The covariant manifestly spacetime-supersymmetric string is introduced in Green & Schwarz (1984a). Topological string theory is introduced in Witten (1988).

The ten-dimensional supersymmetric heterotic string is introduced in Gross, Harvey, Martinec, & Rohm (1985). Our treatment of the nonsupersymmetric theories is similar to that in Kawai, Lewellen, & Tye (1986b), which also has references to earlier constructions of the various models. Level matching and discrete torsion are discussed in Vafa (1986).

Wybourne (1974) and Georgi (1982) are introductions to Lie algebras and groups. A recent reference covering both Lie and current algebras is Fuchs & Schweigert (1997). Many useful facts are tabulated in Slansky (1981). Goddard & Olive (1986) is a thorough review of current algebra. For more on the Sugawara construction see Goddard & Olive (1985); further current algebra references are given for chapter 15. The algebra of Chan–Paton factors is analyzed in Marcus & Sagnotti (1982).

For more on toroidal compactification of the heterotic string, see Narain (1986) and Narain, Sarmadi, & Witten (1987). The heterotic string as a BPS state is discussed in Dabholkar, Gibbons, Harvey, & Ruiz Ruiz (1990). Gauntlett, Harvey, & Liu (1993) is a reference/review on magnetic monopoles in toroidally compactified heterotic string theory.

Chapter 12

For further discussion and references on supergravity actions see chapter 13 of GSW and Townsend (1996). Gravitational anomalies are discussed in detail in Alvarez-Gaumé & Witten (1983). Anomaly cancellation is discussed in Green & Schwarz (1984b) and chapters 10 and 13 of GSW.

Superspace and picture-changing are discussed in Friedan, Martinec, & Shenker (1986). The superspace formalism for nonlinear sigma models with various supersymmetries is reviewed in Roček (1993). The connection between the world-sheet anomaly and the spacetime Chern–Simons action is discussed in Hull & Witten (1985).

Our treatment of superstring perturbation theory is similar to that in Friedan, Martinec, & Shenker (1986), Verlinde & Verlinde (1987), Martinec (1987), Alvarez-Gaumé *et al.* (1988), and Giddings (1992). See also Atick, Moore, & Sen (1988), La & Nelson (1989), and Aoki, D'Hoker, & Phong (1990). Light-cone methods are developed in chapters 7–11 of GSW.

The various tree amplitudes are treated in Schwarz (1982), Gross, Harvey, Martinec, & Rohm (1986), Kawai, Lewellen, & Tye (1986a), and chapter 7 of GSW. The extraction of higher dimension corrections to the low energy action is discussed in Grisaru, van de Ven, & Zanon (1986) and Gross & Sloan (1987). One-loop calculations in the light-cone gauge are in chapters 8–10 of GSW. The explicit one-loop calculations in section 12.6 are based on Lerche, Nilsson, Schellekens, & Warner (1988) and Abe, Kubota, & Sakai (1988). Some related higher-loop amplitudes that can be evaluated in closed form are discussed in Bershadsky, Cecotti, Ooguri, & Vafa (1994) and Antoniadis, Gava, Narain, & Taylor (1994). Nonrenormalization theorems are derived by world-sheet contour arguments in Martinec (1986).

Chapter 13

Most of the subjects in this chapter are covered in the review by Polchinski (1997). For more on T-duality see the review by Giveon, Porrati, & Rabinovici (1994). Much of the discussion in the first three sections follows Polchinski (1995). The $e^{-O(1/g)}$ effects are discussed in Shenker (1991). Their connection with D-instantons is discussed in Polchinski (1994); Green & Gutperle (1997) give a detailed treatment of D-instanton effects. A recent discussion of the Born–Infeld action appears in Tseytlin (1997).

The discussion of branes at angles is similar to that in Berkooz, Douglas, & Leigh (1996). The quartic identity of Riemann and other theta function identities are in Mumford (1983). For more on D-brane scattering see Bachas (1996), Lifschytz (1996), and Douglas, Kabat, Pouliot, & Shenker (1997) and references therein.

Non-Abelian D-brane dynamics and F–D bound states are discussed in Witten (1996a). For more on D0–D0 bound states see Sen (1996a) and Sethi & Stern (1997). For more on D0–D4 bound states see Sen (1996b)

and Vafa (1996a). For more on the connection between D-branes and instantons see Witten (1996b), Vafa (1996b), and Douglas (1996).

Chapter 14

A number of string duality conjectures have been put forward over the years, but the coherent picture presented in this chapter took shape with the work of Hull & Townsend (1995), Townsend (1995), & Witten (1995). Some reviews are Townsend (1996), Sen (1997,1998), and Schwarz (1997). For more on $SO(32)$ type I/heterotic duality see Polchinski & Witten (1996). For more on the strongly coupled $E_8 \times E_8$ theory see Hořava & Witten (1996). Some tests of string dualities based on perturbative and nonperturbative string amplitudes are discussed in Tseytlin (1995) and Green (1997).

Callan, Harvey, & Strominger (1992) review extended objects with NS–NS charges. For black p-branes see Horowitz & Strominger (1991). Townsend (1996), Duff (1997), and Stelle (1998) review the various extended objects that play a role in string duality and the connection between extended objects in M-theory and in IIA string theory. Strominger (1996) discusses extended objects ending on other objects. Harvey (1997) reviews magnetic monopoles and Montonen–Olive duality. The discussion of type I D5-branes follows Gimon & Polchinski (1996) and Witten (1996b).

Matrix theory is introduced in Banks, Fischler, Shenker, & Susskind (1997). Banks (1997) and Bigatti & Susskind (1997) are reviews.

For background on black hole thermodynamics see Carter (1979) and Wald (1997). The entropy calculation in the chapter is similar to that in Strominger & Vafa (1996). Horowitz (1997), Peet (1997), and Maldacena (1998) review D-brane calculations of black hole entropies and other properties. The correspondence principle is discussed in Horowitz & Polchinski (1997). Page (1994) gives a review of the black hole information problem. Susskind (1995) discusses the possible breakdown of locality in string theory.

The connection between branes and gauge theory dynamics is reviewed in Giveon & Kutasov (1998). For very recent progress see Maldacena (1997).

Chapter 15

Much of the first three sections is based on Belavin, Polyakov, & Zamolodchikov (1984). The review by Ginsparg (1990) covers many of the subjects in this chapter. Many of the relevant papers are collected in Goddard & Olive (1988) or in Itzykson, Saleur, & Zuber (1988).

The unitary representations of the Virasora algebra are discussed by Friedan, Qiu, & Shenker (1984). Thorn (1984) gives a stringy derivation of the Kac formula. Cardy (1986) discusses general aspects of modular invariance, and Cappelli, Itzykson, & Zuber (1987) give the modular-invariant partition functions for the minimal models and $SU(2)$ current algebras. For the solution of minimal models using the Feigin–Fuchs representation see Dotsenko & Fateev (1984, 1985).

The exact solution of current algebra CFTs is described in Knizhnik & Zamolodchikov (1984) and Gepner & Witten (1986). The nonlinear sigma model interpretation is in Witten (1984). Free-field representations of current algebras are obtained in Bershadsky & Ooguri (1989). The coset construction is developed in Goddard, Kent, & Olive (1986). Parafermionic theories are described in Zamolodchikov & Fateev (1985). W-algebras are reviewed in Bouwknegt & Schoutens (1993) and de Boer, Harmsze, & Tjin (1996). Our discussion of rational CFT is largely based on Vafa (1988). Moore & Seiberg (1989) give a systematic treatment of the monodromy and other constraints. Irrational CFT is reviewed in Halpern, Kiritsis, Obers, & Clubok (1996).

Many of the subjects in the final two sections are developed in the review by Cardy (1990). For more on the c-theorem see Zamolodchikov (1986b), and for more on Landau–Ginzburg models see Zamolodchikov (1986a).

Chapter 16

Dixon, Harvey, Vafa, & Witten (1985, 1986) develop the general framework for strings on orbifolds. Modular invariance is discussed in Vafa (1986). Orbifold vertex operators and interactions are treated in Dixon, Friedan, Martinec, & Shenker (1987), Hamidi & Vafa (1987), and the review by Dixon (1988). These papers also discuss the blowing up of the fixed points; our discussion is similar to that in Hamidi & Vafa.

Asymmetric orbifolds are developed in Narain, Sarmadi, & Vafa (1987). Antoniadis, Bachas, & Kounnas (1987) and Kawai, Lewellen, & Tye (1987) develop general free-fermion models. A generalized free-boson construction appears in Lerche, Schellekens, & Warner (1989).

Two (of the many) discussions of the motivation for spacetime supersymmetry and of general aspects of supersymmetric model building are Witten (1981) and Dine (1997). Ross (1984) is an introduction to grand unification.

Font, Ibàñez, Quevedo, & Sierra (1990) is a review of three generation orbifold models; the model (16.3.32) appears in section 4.2. A much-vamped three-generation free-fermion model appears in Antoniadis,

Ellis, Hagelin, & Nanopoulos (1989). Kakushadze, Shiu, Tye, & Vtorov-Karevsky (1997) is a recent review of free-field models with particular attention to higher level three-generation models, which can have ordinary grand unified symmetry breaking.

The discussion of the action for untwisted moduli is patterned on Witten (1985). The general expression for the one-loop threshold correction is obtained in Kaplunovsky (1988); the lectures by Kiritsis (1997) give a thorough treatment. The evaluation of Δ_a for orbifold models is in Dixon, Kaplunovsky, & Louis (1991). The paper by Ibàñez & Lüst (1992) reviews many aspects of the low energy physics of orbifolds, especially those connected with T-duality and with threshold corrections. Quevedo (1996) is a review of low energy string physics.

Chapter 17

The necessary geometric background is given in more detail in chapter 15 of GSW and in Candelas (1988). Hübsch (1992) is a full length treatment at a more advanced level. Calabi–Yau compactification is developed in Candelas, Horowitz, Strominger, & Witten (1985) and in chapter 16 of GSW. Strominger & Witten (1985) discuss various aspects of the low energy physics. For more on the low energy action see Candelas & de la Ossa (1991). The nonrenormalization theorem is from Witten (1986), who also discusses (0,2) compactifications. World-sheet instantons are discussed in Dine, Seiberg, Wen, & Witten (1986, 1987). An analysis of the field equations without the vanishing torsion assumption is in Strominger (1986).

Chapter 18

Continuous symmetries are discussed in Banks & Dixon (1988). Dine (1995) discusses discrete symmetries and the strong CP problem in string theory.

Closed string gauge couplings are discussed in Ginsparg (1987). Constraints on right-moving and type II gauge symmetries are in Dixon, Kaplunovsky, & Vafa (1987). Dienes (1997) is an extensive review of coupling constant unification in string theory. The argument in figure 18.1 for the proximity of the compactification and string scales is based on Kaplunovsky (1985). The discussion of the effect of an extra dimension in figure 18.2 is based on Witten (1996c). The derivation of the moduli independence of $\sin^2\theta_w$ follows Banks, Dixon, Friedan, & Martinec (1988). The unification of the couplings in supersymmetric theories is reviewed in Dimopoulos, Raby, & Wilczek (1991). The discussion of fractional

charges is taken from Schellekens (1990). For more on proton stability in supersymmetric and string theories see Ibàñez & Ross (1992), Pati (1996).

The general argument that spacetime supersymmetry requires $N = 2$ world-sheet supersymmetry is from Banks, Dixon, Friedan, & Martinec (1988). The analysis for extended supersymmetry is in Banks & Dixon (1988). The world-sheet argument that supersymmetry breaking cannot be turned on continuously is also in that paper; the spacetime derivation of the same result is in Dine & Seiberg (1988). The use of PQ symmetry and the scaling of S to derive nonrenormalization theorems is in Dine & Seiberg (1986). Derivation of nonrenormalization theorems from the structure of string perturbation theory is in Martinec (1986). The reader will note that the spacetime derivations are generally shorter and less intricate, and can in some cases give nonperturbative information as well. Generation of D-terms by string loops is discussed in Dine, Seiberg, & Witten (1987). The reviews by Quevedo (1996) and Dine (1997) discuss nonperturbative supersymmetry breaking in more detail, with extensive references. The cosmological constant problem is reviewed in Weinberg (1989).

Chapter 19

Many of the subjects in this chapter are covered in the review by Greene (1997).

For more on chiral rings see Lerche, Vafa, & Warner (1989). For type II strings on Calabi–Yau manifolds and their low energy actions, see Cecotti, Ferrara, & Girardello (1989). The world-sheet argument for the vanishing of the potential for the moduli is given in more detail in Dixon (1988). For a systematic derivation of the constraints from (2,2) superconformal symmetry, derived from analysis of string scattering amplitudes, see Dixon, Kaplunovsky, & Louis (1990). For arguments using the relation between type II and heterotic compactification see Dine & Seiberg (1988).

For more on $N = 2$ minimal models see Boucher, Friedan, & Kent (1986); for more on their connection with $SU(2)$ current algebra see Zamolodchikov & Fateev (1986) and Qiu (1987). For more on $N = 2$ Landau–Ginzburg models and singularity theory see Martinec (1989) and Vafa & Warner (1989). Gepner models are constructed in Gepner (1988). Our discussion is based on Vafa (1989); our discussion of the connection to Calabi–Yau compactification is based on Witten (1993).

Numerical evidence for mirror symmetry appears in Candelas, Lynker, & Schimmrigk (1990). The construction via twisted Gepner models is in Greene & Plesser (1990). Strominger, Yau, & Zaslow (1996) obtain the connection to T-duality. Toric geometry and other advanced ideas are

covered in the review by Greene (1997). The use of mirror symmetry to obtain the exact low energy action is in Candelas, de la Ossa, Green, & Parkes (1991). The flop transition is described in Aspinwall, Greene, & Morrison (1993) and Witten (1993). Cox & Katz (1998) is a recent treatment of mathematics and mirror symmetry.

The interpretation of the conifold singularity in terms of a light black hole/D-brane is in Strominger (1995). Shenker (1995) discusses the short-distance cutoff on the loop graph. Greene, Morrison, & Strominger (1995) show that condensation of these states leads to topology change, providing a physical interpretation for the geometric observations of Candelas, Green, & Hübsch (1989).

The basic features of string theories on K3 are described in Seiberg (1988) (the discussion of Calabi–Yau moduli space in that paper has been superceded by later references). Aspinwall (1997) gives an extended review of this subject. The lectures by Sagnotti (1997) and Schwarz (1997) also cover various six-dimensional string theories, discussing in particular anomaly cancellation. The tensionless string phase transition is described in Seiberg & Witten (1996) and Ganor & Hanany (1996).

For more on the duals of toroidally compactified heterotic strings see Hull & Townsend (1995) and Witten (1995). Our discussion is similar to that in Sen (1997). Sen (1994) is a review of $SL(2,\mathbf{Z})$ duality of the heterotic string on T^6. F-theory is introduced in Vafa (1996c). Kachru & Silverstein (1997) apply F-theory to find heterotic phase transitions that change generation number, and give further references. There is a growing literature on duals of theories with $N = 1$ and $N = 2$ supersymmetry. Vafa & Witten (1995) and Ferrara, Harvey, Strominger, & Vafa (1995) give some relatively simple examples.

Appendix

Our treatment of the spinor representations of $SO(D-1,1)$ and $SO(n)$ follows the treatment for $SO(n)$ in Georgi (1982). Sohnius (1985) also discusses spinors in general dimensions.

Two references on $d = 4$, $N = 1$ supersymmetry are Sohnius (1985) and Wess & Bagger (1992); the former also has some discussion of extended supersymmetry and higher-dimensional theories. The general $d = 4$, $N = 1$ supergravity action is given in Cremmer, Ferrara, Girardello, & Van Proeyen (1983). The significance of the BPS property is developed in Witten & Olive (1978).

The $d = 11$ supergravity theory appears in Cremmer, Julia, & Scherk (1978). Table B.3 (with a misprint corrected) is taken from Hull & Townsend (1995), who give original references. Chapter 13 of GSW and Townsend (1996) have more on supergravity actions in $d = 11$ and $d = 10$; Townsend also discusses the central charges in the supersymmetry algebra. Salam & Sezgin (1989) is a collection of many relevant papers.

The general $d = 4$, $N = 4$ supergravity theory is obtained in de Roo (1985). The general $d = 4$, $N = 2$ supergravity theory is obtained in Andrianopoli *et al.* (1996). The hypermultiplet moduli space is described in Bagger & Witten (1983) and Hitchin, Karlhede, Lindstrom, & Roček (1987). The vector multiplet moduli space is described in de Wit, Lauwers, & Van Proeyen (1985). Seiberg & Witten (1994) give a review of the global supersymmetry limit.

References

Abe, M., Kubota, H., & Sakai, N., (1988). Loop corrections to the $E_8 \times E_8$ heterotic string effective Lagrangian. *Nuclear Physics*, **B306**, 405.

Ademollo, M., Brink, L., D'Adda, A., D'Auria, R., Napolitano, E., Sciuto, S., Del Guidice, E., Di Vecchia, P., Ferrara, S., Gliozzi, F., Musto, R., Pettorino, R., & Schwarz, J. H. (1976). Dual string with $U(1)$ color symmetry. *Nuclear Physics*, **B111**, 77.

Alvarez-Gaumé, L., Nelson, P., Gomez, C., Sierra, G., & Vafa, C. (1988). Fermionic strings in the operator formalism. *Nuclear Physics*, **B311**, 333.

Alvarez-Gaumé, L., & Vazquez-Mozo M. A. (1995). Topics in string theory and quantum gravity. In *Les Houches Summer School on Gravitation and Quantizations, 1992* eds. J. Zinn-Justin & B. Julia, pp. 481–636. Amsterdam: North-Holland. E-print hep-th/9212006.

Alvarez-Gaumé, L., & Witten, E. (1983). Gravitational anomalies. *Nuclear Physics*, **B234**, 269.

Andrianopoli, L., Bertolini, M., Ceresole, A., D'Auria, R., Ferrara, S., & Fré, P. (1996). General matter coupled $N = 2$ supergravity. *Nuclear Physics*, **B476**, 397. E-print hep-th/9603004.

Antoniadis, I., Bachas, C. P., & Kounnas, C. (1987). Four-dimensional superstrings. *Nuclear Physics*, **B289**, 87.

Antoniadis, I., Ellis, J., Hagelin, J. S., & Nanopoulos, D. V. (1989). The flipped $SU(5) \times U(1)$ string model revamped. *Physics Letters*, **B231**, 65.

Antoniadis, I., Gava, E., Narain, K. S., & Taylor, T. R. (1994). Topological amplitudes in string theory. *Nuclear Physics*, **B413**, 162. E-print hep-th/9307158.

Aoki, K., D'Hoker, E., & Phong, D. H. (1990). Unitarity of closed superstring perturbation theory. *Nuclear Physics*, **B342**, 149.

Aspinwall, P. S. (1997). K3 surfaces and string duality. In *Fields, Strings, and Duality, TASI 1996*, eds. C. Efthimiou & B. Greene, pp. 421–540. Singapore: World Scientific. E-print hep-th/9611137.

Aspinwall, P. S., Greene, B. R., & Morrison, D. R. (1993). Multiple mirror manifolds and topology change in string theory. *Physics Letters*, **B303**, 249. E-print hep-th/9301043.

Atick, J. J., Moore, G., & Sen A. (1988). Catoptric tadpoles. *Nuclear Physics*, **B307**, 221.

Bachas, C. (1996). D-brane dynamics. *Physics Letters*, **B374**, 37. E-print hep-th/9511043.

Bagger, J., & Witten, E. (1983). Matter couplings in $N = 2$ supergravity. *Nuclear Physics*, **B222**, 1.

Banks, T. (1997). Matrix theory. E-print hep-th/9710231.

Banks, T., & Dixon, L. J. (1988). Constraints on string vacua with spacetime supersymmetry. *Nuclear Physics*, **B307**, 93.

Banks, T., Dixon, L. J., Friedan, D., & Martinec, E. (1988). Phenomenology and conformal field theory or can string theory predict the weak mixing angle? *Nuclear Physics*, **B299**, 613.

Banks, T., Fischler, W., Shenker, S. H., & Susskind, L. (1997). M theory as a matrix model: a conjecture. *Physical Review*, **D55**, 5112. E-print hep-th/9610043.

Belavin, A. A., Polyakov, A. M., & Zamolodchikov, A. B. (1984). Infinite conformal symmetry in two-dimensional quantum field theory. *Nuclear Physics*, **B241**, 333. Also in Goddard & Olive (1988), Itzykson *et al.* (1988).

Berkooz, M., Douglas, M. R., & Leigh, R. G. (1996). Branes intersecting at angles. *Nuclear Physics*, **B480**, 265. E-print hep-th/9606139.

Bershadsky, M., Cecotti, S., Ooguri, H., & Vafa, C. (1994). Kodaira–Spencer theory of gravity and exact results for quantum string amplitudes. *Communications in Mathematical Physics*, **165**, 311. E-print hep-th/9309140.

Bershadsky, M., & Ooguri, H. (1989). Hidden $SL(n)$ symmetry in conformal field theories. *Communications in Mathematical Physics*, **126**, 49.

Bigatti, D., & Susskind, L. (1997). Review of matrix theory. E-print hep-th/9712072.

Boucher, W., Friedan, D., & Kent, A. (1986). Determinant formulae and unitarity for the $N = 2$ superconformal algebras in two dimensions or exact results on string compactification. *Physics Letters*, **B172**, 316.

Bouwknegt, P., & Schoutens, K. (1993). W-symmetry in conformal field theory. *Physics Reports*, **223**, 183. E-print hep-th/9210010.

Callan, C. G., Harvey, J. A., & Strominger, A. (1992). Supersymmetric string solitons. In *String Theory and Quantum Gravity, Trieste 1991*, eds. J. A. Harvey *et al.*, pp. 208–244. Singapore: World Scientific. E-print hep-th/9112030.

Candelas, P. (1988). Lectures on complex manifolds. In *Superstrings '87*, eds. L. Alvarez-Gaumé *et al.*, pp. 1–88. Singapore: World Scientific.

Candelas, P., & de la Ossa, X. C. (1991). Moduli space of Calabi–Yau manifolds. *Nuclear Physics*, **B355**, 455.

Candelas, P., de la Ossa, X. C., Green, P. S., & Parkes, L. (1991). A pair of Calabi–Yau manifolds as an exactly soluble superconformal theory. *Nuclear Physics*, **B359**, 21.

Candelas, P., Green, P. S., & Hübsch, T. (1989). Rolling among Calabi–Yau vacua. *Nuclear Physics*, **B330**, 49.

Candelas, P., Horowitz, G. T., Strominger, A., & Witten, E. (1985). Vacuum configurations for superstrings. *Nuclear Physics*, **B258**, 46.

Candelas, P., Lynker, M., & Schimmrigk, R. (1990). Calabi–Yau manifolds in weighted P_4. *Nuclear Physics*, **B341**, 383.

Cappelli, A., Itzykson, C., & Zuber, J.-B. (1987). Modular-invariant partition functions in two dimensions. *Nuclear Physics*, **B280**, 445. Also in Itzykson *et al.* (1988).

Cardy, J. L. (1986). Operator content of two-dimensional conformally invariant theories. *Nuclear Physics*, **B270**, 186. Also in Goddard & Olive (1988), Itzykson *et al.* (1988).

Cardy, J. L. (1990). Conformal invariance and statistical mechanics. In *Fields, Strings, and Critical Phenomena, Les Houches 1988*, eds. E. Brezin & J. Zinn-Justin, pp. 169–246. Amsterdam: North-Holland.

Carter, B. (1979). The general theory of the mechanical, electromagnetic, and thermodynamic properties of black holes. In *General Relativity. An Einstein Centennial Survey*, eds. S. W. Hawking & W. Israel. Cambridge: Cambridge University Press.

Cecotti, S., Ferrara, S., & Girardello, L. (1989). Geometry of type II superstrings and the moduli of superconformal field theories. *International Journal of Modern Physics*, **A4**, 2475.

Cox, D. A., & Katz, S. (1998). *Mirror Symmetry and Algebraic Geometry*, American Mathematical Society, in press.

Cremmer, E., Ferrara, S., Girardello, L., & Van Proeyen, A. (1983). Yang–Mills theories with local supersymmetry: Lagrangian, transformation laws, and super-Higgs effect. *Nuclear Physics*, **B212**, 413.

Cremmer, E., Julia, B., & Scherk, J. (1978). Supergravity theory in eleven dimensions. *Physics Letters*, **B76**, 409.

Dabholkar, A., Gibbons, G., Harvey, J. A., & Ruiz Ruiz, F. (1990). Superstrings and solitons. *Nuclear Physics*, **B340**, 33.

de Boer, J., Harmsze, F., & Tjin, T. (1996). Non-linear finite W-symmetries and applications in elementary systems. *Physics Reports*, **272**, 139. E-print hep-th/9503161.

de Roo, M. (1985). Matter coupling in $N = 4$ supergravity. *Nuclear Physics*, **B255**, 515.

de Wit, B., Lauwers, P. G., & Van Proeyen, A. (1985). Lagrangians of $N = 2$ supergravity-matter systems. *Nuclear Physics*, **B255**, 569.

D'Hoker, E. (1993). TASI Lectures on critical string theory. In *Recent Directions in Particle Theory, TASI 1992*, eds. J. Harvey & J. Polchinski, pp. 1–100. Singapore: World Scientific, 1993.

Dienes, K. R. (1997). String theory and the path to unification: a review of recent developments. *Physics Reports*, **287**, 447. E-print hep-th/9602045 .

Dimopoulos, S., Raby, S. A., & Wilczek, F. (1991) Unification of couplings. *Physics Today*, **44**, No. 10, 25.

Dine, M. (1995). Topics in string phenomenology. In *Proceedings of Strings '93*, eds. M. B. Halpern, G. Rivlis, & A. Sevrin, pp. 268–284. Singapore: World Scientific. E-print hep-ph/9309319.

Dine, M. (1997). Supersymmetry phenomenology (with a broad brush). In *Fields, Strings, and Duality, TASI 1996*, eds. C. Efthimiou & B. Greene, pp. 813–882. Singapore: World Scientific. E-print hep-ph/9612389.

Dine, M., & Seiberg, N. (1986). Nonrenormalization theorems in superstring theory. *Physical Review Letters*, **57**, 2625.

Dine, M., & Seiberg, N. (1988). Microscopic knowledge from macroscopic physics in string theory. *Nuclear Physics*, **B301**, 357.

Dine, M., Seiberg, N., Wen, X.-G., & Witten, E. (1986). Nonperturbative effects on the string world sheet. *Nuclear Physics*, **B278**, 769.

Dine, M., Seiberg, N., Wen, X.-G., & Witten, E. (1987). Nonperturbative effects on the string world sheet. II. *Nuclear Physics*, **B289**, 319.

Dine, M., Seiberg, N., & Witten, E. (1987). Fayet–Iliopoulos terms in string theory. *Nuclear Physics*, **B289**, 589.

Dixon, L. J. (1988). Some world sheet properties of superstring compactifications, on orbifolds and otherwise. In *Superstrings, Unified Theories, and Cosmology, 1987*, eds. G. Furlan *et al.*, pp. 67–126. Singapore: World Scientific.

Dixon, L. J., Friedan, D., Martinec, E., & Shenker, S. (1987). The conformal field theory of orbifolds. *Nuclear Physics*, **B282**, 13.

Dixon, L. J., Harvey, J. A., Vafa, C., & Witten, E. (1985). Strings on orbifolds. *Nuclear Physics*, **B261**, 678.

Dixon, L. J., Harvey, J. A., Vafa, C., & Witten, E. (1986). Strings on orbifolds. II. *Nuclear Physics*, **B274**, 285.

Dixon, L. J., Kaplunovsky, V. S., & Louis, J. (1990). On effective field theories describing (2,2) vacua of the heterotic string. *Nuclear Physics*, **B329**, 27.

Dixon, L. J., Kaplunovsky, V. S., & Louis, J. (1991). Moduli dependence of string loop corrections to gauge coupling constants. *Nuclear Physics*, **B355**, 649.

Dixon, L. J., Kaplunovsky, V. S., & Vafa, C. (1987). On four-dimensional gauge theories from type II superstrings. *Nuclear Physics*, **B294**, 43.

Dotsenko, V. S., & Fateev, V. A. (1984). Conformal algebra and multipoint correlation functions in two-dimensional statistical models. *Nuclear Physics*, **B240**, 312. Also in Itzykson *et al.* (1988).

Dotsenko, V. S., & Fateev, V. A. (1985). Four-point correlation functions and the operator algebra in two-dimensional conformal-invariant theories with central charge $c \leq 1$. *Nuclear Physics*, **B251**, 691.

Douglas, M. R. (1996). Gauge fields and D-branes. E-print hep-th/9604198.

Douglas, M. R., Kabat, D., Pouliot, P., & Shenker, S. H. (1997). D-branes and short distances in string theory. *Nuclear Physics*, **B485**, 85. E-print hep-th/9608024.

Duff, M. (1997). Supermembranes. In *Fields, Strings, and Duality, TASI 1996*, eds. C. Efthimiou & B. Greene, pp. 219–290. Singapore: World Scientific. E-print hep-th/9611203.

Efthimiou, C., & Greene, B., eds. (1997). *Fields, Strings, and Duality, TASI 1996*, Singapore: World Scientific.

Ferrara, S., Harvey, J. A., Strominger, A., & Vafa, C. (1995). Second-quantized mirror symmetry. *Physics Letters*, **B361**, 59. E-print hep-th/9505162.

Font, A., Ibáñez, L. E., Quevedo, F., & Sierra, A. (1990). The construction of 'realistic' four-dimensional strings through orbifolds. *Nuclear Physics*, **B331**, 421.

Friedan, D., Martinec, E., & Shenker, S. (1986). Conformal invariance, supersymmetry, and string theory. *Nuclear Physics*, **B271**, 93.

Friedan, D., Qiu, Z., & Shenker, S. (1984). Conformal invariance, unitarity and critical exponents in two dimensions. *Physical Review Letters*, **52**, 1575. Also in Goddard & Olive (1988), Itzykson *et al.* (1988).

Fuchs, J., & Schweigert, C. (1997). *Symmetries, Lie Algebras and Representations*. Cambridge: Cambridge University Press.

Ganor, O. J., & Hanany, A. (1996). Small E_8 instantons and tensionless noncritical strings. *Nuclear Physics*, **B474**, 122. E-print hep-th/9602120.

Gauntlett, J. P., Harvey, J. A., & Liu, J. T. (1993). Magnetic monopoles in string theory. *Nuclear Physics*, **B409**, 363. E-print hep-th/9211056.

Georgi, H. (1982). *Lie Algebras in Particle Physics*. Reading: Benjamin-Cummings.

Gepner, D. (1988). Spacetime supersymmetry in compactified string theory and superconformal models. *Nuclear Physics*, **B296**, 757.

Gepner, D., & Witten, E. (1986). String theory on group manifolds. *Nuclear Physics*, **B278**, 493.

Giddings, S. B. (1992). Punctures on super-Riemann surfaces. *Communications in Mathematical Physics*, **143**, 355.

Gimon, E. G., & Polchinski, J. (1996). Consistency conditions for orientifolds and D-manifolds. *Physical Review*, **D54**, 1667. E-print hep-th/9601038.

Ginsparg, P. (1987). Gauge and gravitational couplings in four-dimensional string theories. *Physics Letters*, **197**, 139.

Ginsparg, P. (1990). Applied conformal field theory. In *Fields, Strings, and Critical Phenomena, Les Houches 1988*, eds. E. Brezin & J. Zinn-Justin, pp. 1–168. Amsterdam: North-Holland.

Giveon, A., Porrati, M., & Rabinovici, E. (1994). Target space duality in string theory. *Physics Reports*, **244**, 77. E-print hep-th/9401139.

Giveon, A., & Kutasov, D. (1998). Brane dynamics and gauge theory. E-print hep-th/9802067.

Gliozzi, F., Scherk, J., & Olive, D. (1977). Supersymmetry, supergravity theories, and the dual spinor model. *Nuclear Physics*, **B122**, 253.

Goddard, P., Kent, A., & Olive, D. (1986). Unitary representations of the Virasoro and super-Virasoro algebras. *Communications in Mathematical Physics*, **103**, 105. Also in Goddard & Olive (1988), Itzykson *et al.* (1988).

Goddard, P., & Olive, D. (1985). Kac–Moody algebras, conformal symmetry, and critical exponents. *Nuclear Physics*, **B257**, 226.

Goddard, P., & Olive, D. (1986). Kac–Moody and Virasoro algebras in relation to quantum physics. *International Journal of Modern Physics*, **A1**, 303. Also in Goddard & Olive (1988).

Goddard, P., & Olive, D., eds. (1988). *Kac–Moody and Virasoro Algebras*. Singapore: World Scientific.

Green, M. B. (1997). Connections between M-theory and superstrings. E-print hep-th/9712195.

Green, M. B., & Gutperle, M. (1997). Effects of D-instantons. *Nuclear Physics*, **B498**, 195. E-print hep-th/9701093.

Green, M. B., & Schwarz, J. H. (1984a). Covariant description of superstrings. *Physics Letters*, **B136**, 367.

Green, M. B., & Schwarz, J. H. (1984b). Anomaly cancellations in supersymmetric $d = 10$ gauge theory and superstring theory. *Physics Letters*, **B149**, 117.

Green, M. B., & Schwarz, J. H. (1985). Infinity cancellations in $SO(32)$ superstring theory. *Physics Letters*, **B151**, 21.

Green, M. B., Schwarz, J. H., & Witten, E. (1987). *Superstring Theory*, in two volumes. Cambridge: Cambridge University Press.

Greene, B. R. (1997). String theory on Calabi–Yau manifolds. In *Fields, Strings, and Duality, TASI 1996*, eds. C. Efthimiou & B. Greene, pp. 543–726. Singapore: World Scientific. E-print hep-th/9702155.

Greene, B. R., & Plesser, M. R. (1990). Duality in Calabi–Yau moduli space. *Nuclear Physics*, **B338**, 15.

Greene, B. R., Morrison, D. R., & Strominger, A. (1995). Black hole condensation and the unification of string vacua. *Nuclear Physics*, **B451**, 109. E-print hep-th/9504145.

Grisaru, M. T., van de Ven, A. E. M., & Zanon, D. (1986). Two-dimensional supersymmetric sigma models on Ricci-flat Kähler manifolds are not finite. *Nuclear Physics*, **B277**, 388.

Gross, D. J., Harvey, J. A., Martinec, E., & Rohm, R. (1985). Heterotic string theory (I). The free heterotic string. *Nuclear Physics*, **B256**, 253.

Gross, D. J., Harvey, J. A., Martinec, E., & Rohm, R. (1986). Heterotic string theory (II). The interacting heterotic string. *Nuclear Physics*, **B267**, 75.

Gross, D. J., & Sloan, J. H. (1987). The quartic effective action for the heterotic string. *Nuclear Physics*, **B291**, 41.

Halpern, M. B., Kiritsis, E., Obers, N., & Clubok, K. (1996). Irrational conformal field theory. *Physics Reports*, **265**, 1. E-print hep-th/9501144.

Hamidi, S., & Vafa, C. (1987). Interactions on orbifolds. *Nuclear Physics*, **B279**, 465.

Harvey, J. A. (1997). Magnetic monopoles, duality, and supersymmetry. In *Fields, Strings, and Duality, TASI 1996*, eds. C. Efthimiou & B. Greene, pp. 157–216. Singapore: World Scientific. E-print hep-th/9603086.

Hitchin, N. J., Karlhede, A., Lindström, U., & Roček, M. (1987). Hyperkähler metrics and supersymmetry. *Communications in Mathematical Physics*, **108**, 535.

Hořava, P., & Witten, E. (1996). Heterotic and type I string dynamics from eleven dimensions. *Nuclear Physics*, **B460**, 506. E-print hep-th/9510209.

Horowitz, G. T. (1997). Quantum states of black holes. E-print gr-qc/9704072.

Horowitz, G. T., & Polchinski, J. (1997). Correspondence principle for black holes and strings. *Physical Review*, **D55**, 6189. E-print hep-th/9612146.

Horowitz, G. T., & Strominger, A. (1991). Black strings and p-branes. *Nuclear Physics*, **B360**, 197.

Hübsch, T. (1992). *Calabi–Yau Manifolds a Bestiary for Physicists*. Singapore: World Scientific.

Hull, C. M., & Townsend, P. K. (1995). Unity of superstring dualities. *Nuclear Physics*, **B438**, 109. E-print hep-th/9410167.

Hull, C. M., & Witten, E. (1985). Supersymmetric sigma models and the heterotic string. *Physics Letters*, **160B**, 398.

Ibáñez, L. E., & Lüst, D. (1992). Duality-anomaly cancellation, minimal string unification, and the effective low-energy Lagrangian of 4-D strings. *Nuclear Physics*, **B382**, 305. E-print hep-th/9202046.

Ibáñez, L. E., & Ross, G. G. (1992). Discrete gauge symmetries and the origin of baryon and lepton number conservation in supersymmetric versions of the standard model. *Nuclear Physics*, **B368**, 3.

Itzykson, C., Saleur, H., & Zuber, J.-B., eds. (1988). *Conformal Invariance and Applications to Statistical Mechanics.* Singapore: World Scientific.

Kachru, S., & Silverstein, E. (1997). Chirality changing phase transitions in 4d string vacua. *Nuclear Physics,* **B504**, 272. E-print hep-th/9704185.

Kakushadze, Z., Shiu, G., Tye, S. H. H., & Vtorov-Karevsky, Y. (1997). A review of three family grand unified string models. E-print hep-th/9710149.

Kaplunovsky, V. S. (1985). Mass scales of string unification. *Physical Review Letters,* **55**, 1036.

Kaplunovsky, V. S. (1988). One-loop threshold effects in string unification. *Nuclear Physics,* **B307**, 145; erratum *ibid.* **B382**, 436. E-print hep-th/9205068.

Kawai, H., Lewellen D. C., & Tye, S.-H. H. (1986a). A relation between tree amplitudes of closed and open strings. *Nuclear Physics,* **B269**, 1.

Kawai, H., Lewellen D. C., & Tye, S.-H. H. (1986b). Classification of closed-fermionic-string models. *Physical Review,* **D34**, 3794.

Kawai, H., Lewellen D. C., & Tye, S.-H. H. (1987). Construction of fermionic string models in four dimensions. *Nuclear Physics,* **B288**, 1.

Kiritsis, E. (1997). *Introduction to Superstring Theory.* Leuven: Leuven University, in press. E-print hep-th/9709062.

Knizhnik, V. G., & Zamolodchikov A. B. (1984). *Nuclear Physics,* **B247**, 83. Also in Goddard & Olive (1988), Itzykson *et al.* (1988).

La, H., & Nelson, P. (1989). Unambiguous fermionic string amplitudes. *Physical Review Letters,* **63**, 24.

Lerche, W., Nilsson, B. E. W., Schellekens, A. N., & Warner, N. P. (1988). Anomaly cancelling terms from the elliptic genus. *Nuclear Physics,* **B299**, 91.

Lerche, W., Schellekens, A. N., & Warner, N. P. (1989). Lattices and strings. *Physics Reports,* **177**, 1.

Lerche, W., Vafa, C., & Warner, N. P. (1989). Chiral rings in $N = 2$ superconformal theories. *Nuclear Physics,* **B324**, 427.

Lifschytz, G. (1996). Comparing D-branes to black-branes. *Physics Letters,* **B388**, 720. E-print hep-th/9604156.

Lüst, D., & Theisen, S. (1989). *Lectures on String Theory.* Berlin: Springer-Verlag.

Maldacena, J. M. (1997). The large-N limit of superconformal field theories and supergravity. E-print hep-th/9711200.

Maldacena, J. M. (1998). Black holes and D-branes. *Nuclear Physics Proceedings Supplement,* **61A**, 111. E-print hep-th/9705078.

Marcus, N., & Sagnotti, A. (1982). Tree-level constraints on gauge groups for type I superstrings. *Physics Letters,* **119B**, 97.

Martinec, E. (1986). Nonrenormalization theorems and fermionic string finiteness. *Physics Letters,* **B171**, 189.

Martinec, E. (1987). Conformal field theory on a (super-)Riemann surface. *Nuclear Physics,* **B281**, 157.

Martinec, E. (1989). Algebraic geometry and effective Lagrangians. *Physics Letters,* **B217**, 431.

Misner, C. W., Thorne, K. S., & Wheeler, J. A. (1973). *Gravitation.* San Francisco: Freeman.

Moore, G., & Seiberg, N. (1989). Classical and quantum conformal field theory. *Communications in Mathematical Physics,* **123**, 177.

Mumford, D. (1983). *Tata Lectures on Theta,* volume one. Boston: Birkhauser.

Narain, K. S. (1986). New heterotic string theories in uncompactified dimensions < 10. *Physics Letters,* **169B**, 41.

Narain, K. S., Sarmadi, M. H., & Vafa, C. (1987). Asymmetric orbifolds. *Nuclear Physics,* **B288**, 551.

Narain, K. S., Sarmadi, M. H., & Witten, E. (1987). A note on toroidal compactification of heterotic string theory. *Nuclear Physics,* **B279**, 369.

Ooguri, H., & Vafa, C. (1991). Geometry of $N = 2$ strings. *Nuclear Physics,* **B361**, 469.

Ooguri, H., & Yin, Z. (1997). Lectures on perturbative string theories. In *Fields, Strings, and Duality, TASI 1996,* eds. C. Efthimiou & B. Greene, pp. 5–82. Singapore: World Scientific. E-print hep-th/9612254.

Page, D. N. (1994). Black hole information. In *Proceedings of the 5th Canadian Conference on General Relativity and Relativistic Astrophysics,* eds. R. B. Mann & R. G. McLenaghan, pp. 1–41. Singapore: World Scientific. E-print hep-th/9305040.

Pati, J. C. (1996). The essential role of string-derived symmetries in ensuring proton-stability and light neutrino masses. *Physics Letters,* **B388**, 532.

Peet, A. W. (1997). The Bekenstein formula and string theory. E-print hep-th/9712253.

Peskin, M. (1987). Introduction to string and superstring theory. II. In *From the Planck Scale to the Weak Scale, TASI 1986,* ed. H. Haber, pp. 277–408. Singapore: World Scientific.

Polchinski, J. (1994). Combinatorics of boundaries in string theory. *Physical Review,* **D50**, 6041. E-print hep-th/9407031.

Polchinski, J. (1995). Dirichlet branes and Ramond–Ramond charges. *Physical Review Letters,* **75**, 4724. E-print hep-th/9510017.

Polchinski, J. (1997). TASI lectures on D-branes. In *Fields, Strings, and Duality, TASI 1996,* eds. C. Efthimiou & B. Greene, pp. 293–356. Singapore: World Scientific. E-print hep-th/9611050.

Polchinski, J., & Witten, E. (1996). Evidence for heterotic–type I string duality. *Nuclear Physics*, **B460**, 525. E-print hep-th/9510169.

Pope, C. N. (1995). *W* strings '93. In *Proceedings of Strings '93*, eds. M. B. Halpern, G. Rivlis, & A. Sevrin, pp. 439–450. Singapore: World Scientific. E-print hep-th/9309125.

Qiu, Z. (1987). Nonlocal current algebra and $N = 2$ superconformal field theory in two dimensions. *Physics Letters*, **B188**, 207.

Quevedo, F. (1996). Lectures on superstring phenomenology. In *Workshops on Particles and Fields and Phenomenology of Fundamental Interactions, Puebla 1995*, eds. J. C. D'Olivo, A. Fernánadez, & M. A. Pérez, pp. 202–242. Woodbury, NY: American Institute of Physics. E-print hep-th/9603074.

Roček, M. (1993). Introduction to supersymmetry. In *Recent Directions in Particle Theory, TASI 1992*, eds. J. Harvey & J. Polchinski, pp. 101–140. Singapore: World Scientific, 1993.

Ross, G. G. (1984). *Grand Unified Theories*. Menlo Park: Benjamin/Cummings.

Sagnotti, A. (1997). Surprises in open string perturbation theory. *Nuclear Physics Proceedings Supplement*, **56B**, 332. E-print hep-th/9702093.

Salam, A., & Sezgin, E., eds. (1989). *Supergravities in Diverse Dimensions*. New York, NY: Elsevier.

Schellekens, A. N. (1990). Electric charge quantization in string theory. *Physics Letters*, **B237**, 363.

Schwarz, J. H. (1982). Superstring theory. *Physics Reports*, **89**, 223.

Schwarz, J. H., ed. (1985). *Superstrings. The First 15 Years of Superstring Theory*, in two volumes. Singapore: World Scientific.

Schwarz, J. H. (1997). Lectures on superstring and M theory dualities. In *Fields, Strings, and Duality, TASI 1996*, eds. C. Efthimiou & B. Greene, pp. 359–418. Singapore: World Scientific. E-print hep-th/9607201.

Seiberg, N. (1988). Observations on the moduli space of superconformal field theories. *Nuclear Physics*, **B303**, 286.

Seiberg, N., & Witten, E. (1994). Electric–magnetic duality, monopole condensation, and confinement in $N = 2$ supersymmetric Yang–Mills theory. *Nuclear Physics*, **B426**, 19. E-print hep-th/9407087.

Seiberg, N., & Witten, E. (1996). Comments on string dynamics in six dimensions. *Nuclear Physics*, **B471**, 121. E-print hep-th/9603003.

Sen, A. (1994). Strong–weak coupling duality in four-dimensional string theory. *International Journal of Modern Physics*, **A9**, 3707. E-print hep-th/9402002.

Sen, A. (1996a). A note on marginally stable bound states in type II string theory. *Physical Review*, **D54**, 2964. E-print hep-th/9510229.

Sen, A. (1996b). *U*-duality and intersecting D-branes. *Physical Review*, **D53**, 2874. E-print hep-th/9511026.

Sen, A. (1997). Unification of string dualities. *Nuclear Physics Proceedings Supplement*, **58**, 5. E-print hep-th/9609176.

Sen, A. (1998). An introduction to nonperturbative string theory. E-print hep-th/9802051.

Sethi, S., & Stern, M. (1997). D-brane bound states redux. E-print hep-th/9705046.

Sevrin, A., Troost, W., & Van Proeyen, A. (1988). Superconformal algebras in two dimensions with $N = 4$. *Physics Letters*, **208B**, 447.

Shenker, S. (1991). The strength of nonperturbative effects in string theory. In *Random Surfaces and Quantum Gravity*, eds. O. Alvarez, E. Marinari, & P. Windey, pp. 191–200. New York: Plenum.

Shenker, S. (1995). Another length scale in string theory? E-print hep-th/9509132.

Slansky, R. (1981). Group theory for unified model building. *Physics Reports*, **79**, 1.

Sohnius, M. F. (1985). Introducing supersymmetry. *Physics Reports*, **128**, 39.

Stelle, K. S. (1998). BPS branes in supergravity. E-print hep-th/9803116.

Strominger, A. (1986). Superstrings with torsion. *Nuclear Physics*, **B274**, 253.

Strominger, A. (1995). Massless black holes and conifolds in string theory. *Nuclear Physics*, **B451**, 96. E-print hep-th/9504090.

Strominger, A. (1996). Open *p*-branes. *Physics Letters*, **B383**, 44. E-print hep-th/9512059.

Strominger, A., & Vafa, C. (1996). Microscopic origin of the Bekenstein–Hawking entropy. *Physics Letters*, **B379**, 99. E-print hep-th/9601029.

Strominger, A., & Witten, E. (1985). New manifolds for superstring compactification. *Communications in Mathematical Physics*, **101**, 341.

Strominger, A., Yau, S.-T., & Zaslow, E. (1996). Mirror symmetry is *T*-duality. *Nuclear Physics*, **B479**, 243. E-print hep-th/9606040.

Susskind, L. (1995). The world as a hologram. *Journal of Mathematical Physics*, **36**, 6377. E-print hep-th/9409089.

Thorn, C. B. (1984). Computing the Kac determinant using dual model techniques and more about the no-ghost theorem. *Nuclear Physics*, **B248**, 551.

Townsend, P. K. (1995). The eleven-dimensional membrane revisited. *Physics Letters*, **B350**, 184. E-print hep-th/9501068.

Townsend, P. K. (1996). Four lectures on M-theory. In *High-Energy Physics and Cosmology, Trieste 1996*, eds. E. Gava *et al.*, pp. 385–438. Singapore: World Scientific. E-print hep-th/9612121.

Tseytlin, A. A. (1996). On $SO(32)$ heterotic–type I superstring duality in ten dimensions. *Physics Letters*, **B367**, 84. E-print hep-th/9510173.

Tseytlin, A. A. (1997). On non-Abelian generalisation of Born–Infeld action in string theory. *Nuclear Physics*, **B501**, 41. E-orint hep-th/9701125.

Tye, S. H. H. (1995). Status of fractional superstrings. In *Proceedings of Strings '93*, eds. M. B. Halpern, G. Rivlis, & A. Sevrin, pp. 364–378. Singapore: World Scientific. E-print hep-th/9311021.

Vafa, C. (1986). Modular invariance and discrete torsion on orbifolds. *Nuclear Physics*, **B273**, 592.

Vafa, C. (1988). Toward classification of conformal theories. *Physics Letters*, **206B**, 421.

Vafa, C. (1989). String vacua and orbifoldized LG models. *Modern Physics Letters*, **A4**, 1169.

Vafa, C. (1996a). Gas of D-branes and Hagedorn density of BPS states. *Nuclear Physics*, **B463**, 415. E-print hep-th/9511088.

Vafa, C. (1996b). Instantons on D-branes. *Nuclear Physics*, **B463**, 435. E-print hep-th/9512078.

Vafa, C. (1996c). Evidence for F-theory. *Nuclear Physics*, **469**, 403. E-print hep-th/9602022.

Vafa, C., & Warner, N. (1989). Catastrophes and the classification of conformal theories. *Physics Letters*, **B218**, 51.

Vafa, C., & Witten, E. (1995). Dual string pairs with $N = 1$ and $N = 2$ supersymmetry in four dimensions. E-print hep-th/9507050.

Verlinde, E., & Verlinde, H. (1987). Multiloop calculations in covariant superstring theory. *Physics Letters*, **B192**, 95.

Wald, R. M. (1997). Black holes and thermodynamics. E-print gr-qc/9702022.

Weinberg, S. (1989). The cosmological constant problem. *Reviews of Modern Physics*, **61**, 1.

Wess, J., & Bagger, J. (1992). *Supersymmetry and Supergravity*. Princeton: Princeton University Press.

Witten, E. (1981). Dynamical breaking of supersymmetry. *Nuclear Physics*, **B188**, 513.

Witten, E. (1984). Non-Abelian bosonization in two dimensions. *Communications in Mathematical Physics*, **92**, 455. Also in Goddard & Olive (1988).

Witten, E. (1985). Dimensional reduction of superstring models. *Physics Letters*, **155B**, 151.

Witten, E. (1986). New issues in manifolds of $SU(3)$ holonomy. *Nuclear Physics*, **B268**, 79.

Witten, E. (1988). Topological sigma models. *Communications in Mathematical Physics*, **118**, 411.

Witten, E. (1993). Phases of $N = 2$ theories in two dimensions. *Nuclear Physics*, **B403**, 159. E-print hep-th/9301042.

Witten, E. (1995). String theory dynamics in various dimensions. *Nuclear Physics*, **B443**, 85. E-print hep-th/9503124.

Witten, E. (1996a). Bound states of strings and p-branes. *Nuclear Physics*, **B460**, 335. E-print hep-th/9510135.

Witten, E. (1996b). Small instantons in string theory. *Nuclear Physics*, **B460**, 541. E-print hep-th/9511030.

Witten, E. (1996c). Strong coupling expansion of Calabi–Yau compactification. *Nuclear Physics*, **B471**, 135. E-print hep-th/9602070.

Witten, E., & Olive, D. (1978). Supersymmetry algebras that include topological charges. *Physics Letters*, **B78**, 97.

Wybourne, B. G. (1974). *Classical Groups for Physicists*. New York: Wiley.

Zamolodchikov, A. B. (1986a). Conformal symmetry and multicritical points in two-dimensional quantum field theory. *Soviet Journal of Nuclear Physics*, **44**, 529. Also in Itzykson *et al.* (1988).

Zamolodchikov, A. B. (1986b). 'Irreversibility' of the flux of the renormalization group in a 2D field theory. *JETP Letters*, **43**, 730. Also in Itzykson *et al.* (1988).

Zamolodchikov, A. B., & Fateev, V. A. (1985). Nonlocal (parafermion) currents in two-dimensional conformal quantum field theory and self-dual critical points in $\mathbf{Z}_N$-symmetric statistical mechanics. *Soviet Physics JETP*, **62**, 215. Also in Itzykson *et al.* (1988).

Zamolodchikov, A. B., & Fateev, V. A. (1986). Disorder fields in two-dimensional conformal quantum field theory and $N = 2$ extended supersymmetry. *Soviet Physics JETP*, **63**, 913.

Glossary

(0, 2) compactification a generic heterotic string vacuum having $d = 4$, $N = 1$ supersymmetry. The world-sheet CFT has $N = 2$ right-moving superconformal invariance.

(2, 0) theory in current usage, this refers to a family of nontrivial fixed point theories with tensionless strings and $d = 6$ (2, 0) supersymmetry. These arise on coincident M5-branes and IIA NS5-branes, and on the IIB theory at an A–D–E singularity.

(2, 2) compactification one of a special subset of $d = 4$, $N = 1$ heterotic string vacua, which includes the Calabi–Yau compactifications. The world-sheet CFT has both right-moving and left-moving $N = 2$ superconformal invariance. In the type II string theories, these CFTs give vacua with $d = 4$, $N = 2$ supersymmetry.

A–D–E singularity a singularity of a four-(real)-dimensional complex manifold, resulting from the collapse of one or more two-spheres to zero volume. The terminology A–D–E refers to the Dynkin diagrams of the simply-laced Lie algebras, which describe the intersection numbers of the collapsed spheres.

Abelian differential a globally defined holomorphic (1,0)-form on a Riemann surface.

abstruse identity one of a set of quartic theta function identities due to Jacobi, it implies the degeneracy of bosons and fermions in GSO-projected string theories as required by supersymmetry.

affine Lie algebra see *current algebra.*

anomaly the violation of a classical symmetry by quantum effects. A *gravitational anomaly* is an anomaly in coordinate invariance. A *global anomaly* is an anomaly in a large symmetry transformation (one not continuously connected to the identity).

anomaly polynomial a formal $(d+2)$-form in d-dimensions, which encodes the gauge and gravitational anomalies.

asymptotically locally Euclidean (ALE) space a space which at long distance approaches flat Euclidean space identified under a discrete group. This is the geometry in the neighborhood of an orbifold fixed point (or blown-up fixed point).

Atiyah–Drinfeld–Hitchin–Manin (ADHM) construction a method for the construction of all Yang–Mills field configurations having self-dual field strength.

auxiliary field a nonpropagating field, one whose field equation is algebraic rather than differential. In many supersymmetric theories, the transformations can be simplified by introducing such fields.

axion a Goldstone boson associated with spontaneously broken PQ symmetry. The *model-independent axion* appears in every perturbative string theory, and is closely related to the graviton and dilaton.

***bc* CFT** a free CFT of anticommuting fields with an action of first order in derivatives. There is a family of such CFTs, parameterized by the weight $h_b = 1 - h_c$. For $h_b = 2$ this CFT describes the Faddeev–Popov ghosts associated with conformal invariance.

***βγ* CFT** a free CFT of commuting fields with an action of first order in derivatives. There is a family of such CFTs, parameterized by the weight $h_\beta = 1 - h_\gamma$. For $h_\beta = \frac{3}{2}$ this CFT describes the Faddeev–Popov ghosts associated with superconformal invariance.

Batalin–Vilkovisky formalism an extension of the BRST formalism, for quantizing more general theories with constraints. This has been useful in string field theory.

Becchi–Rouet–Stora–Tyutin (BRST) invariance a nilpotent symmetry of Faddeev–Popov gauge-fixed theories, which encodes the information contained in the original gauge symmetry.

Beltrami differential the derivative with respect to the moduli of the complex structure of a Riemann surface.

Berezin integration a linear operation taking functions of Grassmann variables to complex numbers, with many of the key properties of ordinary integration.

beta function 1. in quantum field theory, the derivative of the effective strength of an interaction with respect to length scale; 2. a special function involving a ratio of gamma functions, which appears in the Veneziano amplitude.

Betti numbers the number of nontrivial p-forms in de Rham cohomology, denoted B_p.

black hole entropy a quantity S proportional to the area of the horizon of a black hole, $S = 2\pi A/\kappa^2$. This has the properties of a thermodynamic entropy: it is nondecreasing in classical general relativity, and the sum of the black hole entropy and the ordinary entropy is nondecreasing even with the inclusion of Hawking radiation. To find a statistical mechanical

derivation of this entropy has been a major goal, partly realized in recent work.

black hole evaporation the emission of thermal (Hawking) radiation by a black hole, due to pair production near the horizon.

black hole information paradox a conflict between quantum mechanics and general relativity. Information falling into a black hole is lost and does not reappear when the black hole evaporates; this is inconsistent with ordinary quantum mechanical evolution. It apparently requires either a significant modification of quantum mechanics, or a significant breakdown of the usual understanding of locality.

black *p*-brane a p-dimensional extended object with an event horizon: a space that is translationally invariant in p directions and has a black hole geometry in the remaining directions.

blow up to deform a singular manifold into a smooth manifold.

Bogomolnyi–Prasad–Sommerfield (BPS) state a state that is invariant under a nontrivial subalgebra of the full supersymmetry algebra. Such states always carry conserved charges, and the supersymmetry algebra determines the mass of the state exactly in terms of its charges. BPS states lie in smaller supersymmetry representations than non-BPS states, so-called *short representations*. When there are short representations of different sizes, one also distinguishes *ultrashort representations*, which are the smallest possible (generally their dimension is the square root of the non-BPS dimension).

Borel summation a method of defining the sum of a divergent series. This has been used as a means of studying nonperturbative effects in field and string theories, but it should be understood that most nonperturbative effects are not usefully studied in terms of the perturbation series.

Born–Infeld action a generalization of the usual gauge field action which is nonpolynomial in the gauge field strength. This was originally proposed as a possible short-distance modification of electromagnetism. It arises as the low energy effective action of the gauge fields on D-branes.

bosonization the exact equivalence of a theory of fermionic fields and a theory of bosonic fields, possible in two dimensions. The boson is a fermion–antifermion pair; the fermion is a coherent state of bosons.

***c*-map** a method for constructing the hypermultiplet moduli space of a type II string theory compactified on a Calabi–Yau three-fold from the vector multiplet moduli space of the other type II theory on the same three-fold.

***c*-theorem** the existence, in unitary CFTs in two dimensions, of a positive quantity c that is monotonically nonincreasing with increasing length scale, and which at fixed points is stationary and equal to the central charge. This is a strong constraint on the global form of the renormalization group flow; no simple analog seems to exist in higher dimensions. Also known as the *Zamolodchikov c-theorem*.

***CPT* symmetry** the combined operation of parity-reversal, time-reversal, and charge conjugation, which is a symmetry of all Lorentz-invariant local quantum field theories.

Calabi–Yau manifold a Kähler manifold with vanishing first Chern class. A *Calabi–Yau n-fold* has $2n$ real $= n$ complex coordinates. Yau's theorem guarantees the existence of a Ricci-flat metric of $SU(n)$ holonomy.

canceled propagator argument a general principle implying, under broad conditions, the vanishing of surface terms on the moduli space of Riemann surfaces and therefore the decoupling of unphysical states in string amplitudes. Such amplitudes are defined by analytic continuation from a regime where the integrand falls rapidly at the boundary and the surface term is identically zero; its continuation is therefore also identically zero.

Casimir energy a shift in the ground state energy of a quantum field theory due to boundary conditions on the fields.

center-of-mass mode the zeroth spatial Fourier component of a quantum field.

central charge an operator (which might be a constant) that appears on the right-hand side of a Lie algebra and commutes with all operators in the algebra. Prominent examples include the constant term in the Virasoro algebra and the charges appearing on the right-hand sides of many supersymmetry algebras.

Chan–Paton degrees of freedom degrees of freedom localized at the endpoints of open strings. These are now interpreted as designating the D-brane on which the string ends.

Chan–Paton factor the vertex operator factor for the state of the Chan–Paton degrees of freedom.

Chern–Simons term a term in the action which involves p-form potentials as well as field strengths. Such a term is gauge-invariant as a consequence of the Bianchi identity and/or the modification of the p-form gauge transformation. These terms usually have a close connection to topology and to anomalies.

chiral 1. acting in a parity asymmetric fashion; see *chiral multiplet, chiral symmetry, chiral theory, chirality, extended chiral algebra*; 2. invariant under part of the supersymmetry algebra; see *chiral field, chiral multiplet, chiral primary, chiral ring*.

chiral field in supersymmetry, a local operator that is invariant under part of the algebra: the operator analog of a *BPS state*.

chiral multiplet the multiplet of $d = 4$, $N = 1$ supersymmetry with two real scalars. The quarks and leptons are contained in such multiplets. This multiplet is connected to both senses of *chiral:* it contains a fermion with chiral couplings, and the integral of the associated superfield is invariant under half of the supersymmetry algebra.

chiral primary in an $N = 2$ SCFT, a primary field that is also annihilated by one of the rigid supersymmetries $G^{\pm}_{-1/2}$.

chiral ring the closed OPE algebra of chiral fields.

chiral symmetry a symmetry whose action on spinor fields is not parity-symmetric.

chiral theory a gauge theory in which the gauge couplings are not parity-symmetric.

chirality in $d = 2k$ dimensions, the eigenvalue of the operator Γ^d which anti-commutes with the Γ^μ. This eigenvalue distinguishes the two Weyl representations, which are related to one another by parity.

Christoffel connection in general relativity, the connection that is constructed from the metric.

critical behavior the behavior of a quantum field theory at an IR fixed point with massless fields, and the approach to this behavior.

closed string a string with the topology of a circle.

cocycle in a vertex operator, an operator-valued phase factor which multiplies the creation–annihilation normal ordered exponential. This in needed in some cases in order to give the operator the correct commuting or anticommuting property.

coefficient functions the position-dependent coefficients of local operators appearing in the expansion of an operator product.

cohomology in any vector space with a nilpotent operator Q (one such that $Q^2 = 0$), the kernel of Q modulo the image of Q. That is, the space of *closed* states (those annihilated by Q) with the *exact* states (those of the form $Q\psi$) defined to be equivalent to zero. *De Rham cohomology* is the cohomology of the exterior derivative d acting on differential forms. On a complex manifold, *Dolbeault cohomology* is the cohomology of ∂ and $\bar{\partial}$ (the $(1,0)$ and $(0,1)$ parts of d) on (p,q)-forms. *Homology* is the cohomology of the boundary operator. *BRST cohomology* is the cohomology of the BRST operator, and defines the physical space of a gauge-invariant theory.

Coleman–Weinberg formula the expression for the vacuum energy density of a free quantum field, from the renormalized sum of the zero-point energies of its modes.

collapsing cycle a cycle whose volume vanishes in a limit, usually giving rise to a singular manifold.

collective coordinate in quantizing a soliton or other extended object, the degrees of freedom corresponding to its position or configuration.

compact CFT a CFT in which the number of states with energy less than any given value is finite. This is defined by analogy with the spectrum of a differential operator on a compact space.

compactification scale the characteristic mass scale of states whose wavefunctions have a nontrivial dependence on the compact dimensions.

compactify to consider a field theory or string theory in a spacetime, some of whose spatial dimensions are compact.

complex manifold a manifold with an assigned system of complex coordinates, modulo holomorphic reparameterizations of these coordinates.

complex structure an equivalence class of complex coordinates. A given differentiable manifold may have many inequivalent complex structures.

complex structure moduli the moduli that parameterize the inequivalent complex structures on a manifold. In compactification on a Calabi–Yau 3-fold, these are associated with $(2,1)$-forms.

conformal block in CFT, the contribution of a single conformal family to a sum over states.

conformal bootstrap the partially successful program to construct all CFTs by using only symmetry and consistency conditions.

conformal family the set of states obtained by acting on a highest weight state with Virasoro raising generators in all inequivalent ways; or, the corresponding set of local operators. A *degenerate conformal family* contains null states, which are orthogonal to all states in the family.

conformal field theory (CFT) a conformally invariant quantum field theory.

conformal gauge a choice of coordinates in two dimensions, such that the metric is proportional to the unit metric.

conformal Killing vector a globally defined infinitesimal diff×Weyl transformation that leaves the metric invariant.

conformal transformation a mapping of Euclidean or Minkowski space to itself that leaves the flat metric invariant up to a position-dependent rescaling; equivalently, the subgroup of diff × Weyl that leaves invariant the flat metric. In $d \geq 3$ dimensions this has $\frac{1}{2}(d+1)(d+2)$ parameters. In two dimensions it is the set of all holomorphic maps. Finite transformations require the inclusion of points at infinity, as in the case of the Möbius transformations of the sphere.
This usage has become standard in string theory and quantum field theory, but in general relativity *conformal transformation* is defined to be any position-dependent rescaling of the metric, now called a *Weyl transformation* in string theory.

conifold a Calabi–Yau manifold with a singular complex structure, corresponding to the collapse of a three-cycle. The string theory on this space is singular; this is now understood to be due to the quantum effects of a massless 3-brane wrapped on the cycle.

conifold transition a change of topology due to condensation of massless 3-brane fields. Under appropriate conditions, the potential for the massless 3-brane fields on a conifold with multiple collapsed cycles has a flat direction for these fields; this corresponds to a change of topology, blowing up a 2-cycle rather than the collapsed 3-cycle.

constraint a symmetry generator whose matrix elements are required to vanish in physical states, either in the BRST or OCQ sense. These can usually

be understood as arising from a gauge symmetry (for so-called *first class constraints*, which are all that we consider), and consist of those gauge symmetry generators that do not vanish by the equations of motion.

coset CFT a CFT constructed as one of the factors of a known CFT, when the energy-momentum tensor of the latter can be written as a sum of commuting pieces. In the classic example the full CFT G is a current algebra, as is one of the factors (H). This can also be thought of as gauging the symmetry H.

cosmological constant the energy density of the vacuum. In a nonsupersymmetric quantum theory (including one with spontaneously broken supersymmetry), there are many effects that give rise to such an energy density. The *cosmological constant problem* is the problem that the cosmological constant in nature is many orders of magnitude smaller than known nonzero effects.

critical dimension the dimension in which a perturbative string theory is consistent in flat Minkowski spacetime; the value is 26 for the bosonic string and 10 for the supersymmetric string theories.

current algebra in quantum field theory, the algebra of the currents associated with a continuous symmetry group g (or of their Fourier modes). As used here, it is the specific algebra that occurs in two-dimensional CFTs, with the energy-momentum tensor defined to be of Sugawara form. The terms *affine Lie algebra* and *affine Kac–Moody algebra* are also used for this algebra, though like *current algebra* they both have broader definitions as well. The term *affine* refers to the c-number (Schwinger) term. An *untwisted* current algebra is the algebra of periodic currents, with integer modes. An algebra can be *twisted* by any automorphism of g.

cycle a topologically nontrivial submanifold (in the sense of homology); a *p-cycle* is p-dimensional. The *A-* and *B-cycles* are a standard basis for the nontrivial one-cycles on a Riemann surface.

D-brane in the type I, IIA, and IIB string theories, a dynamical object on which strings can end. The term is a contraction of *Dirichlet brane*. The coordinates of the attached strings satisfy Dirichlet boundary conditions in the directions normal to the brane and Neumann conditions in the directions tangent to the brane. A *Dp-brane* is p-dimensional, with p taking any even value in the IIA theory, any odd value in the IIB theory, and the values 1, 5, and 9 in the type I theory; a D9-brane fills space and so corresponds to an ordinary Neumann boundary condition. The Dp-brane is a source for the $(p+1)$-form R–R gauge field. The mass or tension of a D-brane is intermediate between that of an ordinary quantum or a fundamental string and that of a *soliton*. The low energy fluctations of D-branes are described by supersymmetric gauge theory, which is non-Abelian for coincident branes.

D-instanton an object localized in (Euclidean) time as well as space, defined by Dirichlet conditions on all coordinates of attached strings. This is similar to a field-theoretic instanton, corresponding to a tunneling process that changes the value of an R–R field strength. More generally, the $(p+1)$-dimensional world-volume of a Dp-brane, when localized in time and wrapped on a $(p+1)$-cycle in space, has similar effects.

D-string a D1-brane, in the type I and IIB string theories.

***D*-term** 1. in gauge theories with four or eight supersymmetries, the auxiliary field in the gauge multiplet; 2 the potential term proportional to the square of this auxiliary field, which depends only on the gauge couplings and, in the $U(1)$ case, the value of the Fayet–Iliopoulos parameter.

Del Guidice–Di Vecchia–Fubini (DDF) operators operators satisfying an oscillator algebra, which create a complete set of physical states in OCQ.

descendant a state obtained by acting on a highest weight state with Virasoro raising generators.

diagonal modular invariant a modular-invariant CFT formed by imposing common boundary conditions on the left- and right-moving fields.

diff invariance general coordinate (reparameterization) invariance, usually applied to the world-sheet coordinates.

dilaton the massless scalar with gravitational-strength couplings, found in all perturbative string theories. An exactly massless dilaton would violate limits on nongravitational interactions, but a mass for the dilaton is not forbidden by any symmetry and so dynamical effects will generate one in vacua with broken supersymmetry (the same holds for other moduli). The string coupling constant is determined by the value of the dilaton field.

dimensional reduction in the simplest cases, toroidal compactification retaining only the states of zero compact momentum. More generally (and less physically) the construction of a lower-dimensional field theory by requiring all fields to be invariant under a set of symmetries; this may have no interpretation in terms of compactification.

Dirac quantization condition for an electric and a magnetic charge, the condition that the product be quantized, $\mu_e \mu'_m = 2\pi n$. For two dyons, which have both charges, the condition is $\mu_e \mu'_m - \mu_m \mu'_e = 2\pi n$. These conditions generalize to objects of dimension p and $d-p-4$ in d dimensions, where one is the source of a $(p+2)$-form field strength and the other of the Poincaré dual $(d-p+2)$-form field strength.

Dirac spinor the unique irreducible representation of the algebra of Dirac matrices, which is also known as a Clifford algebra. This is also a representation of the Lorentz group; in even dimensions it is reducible to two Weyl representations of the Lorentz group. The Dirac spinor is complex; in certain dimensions a further Majorana (reality) condition is compatible with Lorentz invariance.

Dirichlet boundary condition the condition that the value of a field be fixed at a boundary. This is the relevant usage in string theory, but in other contexts only the tangent derivative need be fixed.

discrete torsion in forming a *twisted CFT*, a change in the phases of the path integral sectors and therefore in the projection on the Hilbert space.

doubling trick the representation of holomorphic and antiholomorphic fields on a manifold with boundary by holomorphic fields alone, on the doubled copy of the manifold obtained by reflecting through the boundary.

dual resonance model a phenomenological model of the strong interaction, which developed into string theory. The *dual* refers here to *world-sheet duality*.

duality the equivalence of seemingly distinct physical systems. Such an equivalence often arises when a single quantum theory has distinct classical limits. One classic example is *particle–wave duality*, wherein a quantum field theory has one limit described by classical field theory and another described by classical particle mechanics. Another is the high-temperature–low-temperature duality of the Ising model. Here, low temperature is the statistical mechanical analog of the classical limit, the Boltzmann sum being dominated by the configurations of lowest energy. See *Montonen–Olive duality*, *S-duality*, *string–string duality*, *T-duality*, *U-duality*, *world-sheet duality*.

effective field theory the description of a physical system below a given energy scale (or equivalently, above a given length scale).

Einstein metric the metric whose leading low energy action is the *Hilbert action*

$$\frac{1}{2\kappa^2}\int d^dx(-G)^{1/2}R\ ;$$

this is independent of other fields. Here κ is the gravitational coupling, related to the Planck length by $\kappa = (8\pi)^{1/2}L_P$. This metric is related to other metrics such as the *sigma-model metric* by a field-dependent Weyl transformation. The existence of distinct metrics would appear to violate the equivalence principle, but when the dilaton and other moduli are massive the distinction disappears.

electroweak scale the mass scale of electroweak symmetry breaking, roughly 10^2 GeV.

enhanced gauge symmetry a gauge symmetry appearing at special points in moduli space, which is not evident in the original formulation of a theory. The classic examples are the gauge symmetries that arise at special radii of toroidal compactification, whose gauge bosons are winding states. Many other mechanisms are now known: D-branes and black p-branes wrapped on collapsing cycles, F-strings or D-strings stretched between various branes in the limit that the latter become coincident, and the gauge symmetry appearing on a zero size $SO(32)$ instanton.

Euclidean having a metric of strictly positive signature. The original connotation that the metric be flat is somewhat disregarded; thus one refers to *Euclidean quantum gravity*, a conjectured analytic continuation of the Minkowskian theory. For the metric itself, the term *Riemannian* for a curved metric of Euclidean signature is more precise.

Euclidean adjoint in a Euclidean quantum theory, the Hermitean adjoint combined with time-reversal. The latter operation undoes the time-reversing effect of the adjoint, so that the combined operation is local.

Euler number the topological invariant

$$\chi = \sum_{p=0}^{d}(-1)^p B_p\ ,$$

where B_p is the pth Betti number. It is equal to $2(1-g)$ for a Riemann surface of genus g. More properly, the *Euler characteristic*.

expectation value a path integral with specified insertions. This is the term that we have chosen to use, but *correlation function* and *correlator* are also in common usage.

extended chiral algebra the full set of holomorphic operators in a CFT.

extended supersymmetry a supersymmetry algebra in which the supercharges comprise more than one copy of the smallest spinor representation for the given spacetime dimension.

F-term 1. the auxiliary field in the chiral multiplet of $d = 4$, $N = 1$ supersymmetry; 2. the potential term proportional to the square of this field.

F theory 1. a systematic description of IIB superstring states with nontrivial dilaton and R–R scalar backgrounds, which relates these fields to the modulus τ of an auxiliary two-torus; 2. a conjectured twelve-dimensional quantum theory underlying the IIB string. The name is inspired by M-theory, with F for father.

Faddeev–Popov determinant the Jacobian determinant arising from the reduction of a gauge-invariant functional integral to an integral over a gauge slice.

Faddeev–Popov ghosts the wrong-statistics quantum fields used to give a functional integral representation of the Faddeev–Popov determinant.

Fayet–Iliopoulos term in $U(1)$ gauge theories with four or eight supersymmetries, a term in the action which is linear in the auxiliary D-term field.

Feigin–Fuchs representation a representation of minimal model expectation values in terms of free fields.

fibration a space which is locally the product of a fiber F and a base B. The geometry of the fiber varies as one moves over the base, and may become singular. Typical fibers are tori and the $K3$ manifold.

first Chern class on a complex manifold, the Dolbeault cohomology class of the Ricci form $R_{i\bar{j}}dz^i d\bar{z}^j$.

first-quantized description the representation of a quantized particle theory as a sum over particle paths, or of a string theory as a sum over world-sheets. *Second-quantized* refers to the representation in terms of a functional integral over ordinary or string fields. The term second-quantized implies

the reinterpretation of the first-quantized *wavefunction* as a *field operator*. This terminology is in common usage, but it has been argued that it is unsatisfactory, in that it implies a deep principle where none may exist. Since the sum over world-sheets is itself a quantum field theory, one can equally well call it second-quantized, in which case string field theory is *third-quantized*. Third quantization of an ordinary field theory would describe operators that create and destroy universes, a concept which may or may not be useful.

Fischler–Susskind mechanism the cancellation of divergences and anomalies in the world-sheet quantum field theory against divergences and anomalies from integration over small topological features at higher orders of string perturbation theory. This is needed for the consistency of string perturbation theory in a quantum-corrected background.

fixed point 1. (*in geometry*) a point left invariant by a given symmetry transformation. This becomes a boundary point or a singularity if the space is identified under the transformation; 2. (*in quantum field theory*) a quantum theory whose physics is independent of length scale (scale-invariant). Usually such a theory is conformally invariant as well. A *UV fixed point* is the theory governing the short-distance physics of a quantum field theory; an *IR fixed point* is the theory governing the long-distance physics of a quantum field theory. A *trivial IR fixed point* has no massless fields. A *nontrivial IR fixed point* has massless fields with nonvanishing interactions. A theory whose IR limit is a massless free field theory is therefore described by neither of these terms; it is a *noninteracting IR fixed point*.

flat direction in scalar field space, a line of degenerate local minima. The field corresponding to this direction is a *modulus*.

flop a change of topology which can occur in weakly coupled string theory, where a two-cycle collapses and then a different two-cycle blows up.

fractional charge an unconfined particle whose electric charge is not a multiple of that of the electron. These exist in most $d = 4$ string theories, although in many cases all are superheavy.

fractional string theory a proposed generalization of string theory having constraints whose spin is not a multiple of $\frac{1}{2}$. No complete construction exists.

Fuchsian group a discrete subgroup Γ of the $SL(2,\mathbf{R})$ Möbius transformations of the complex upper-half-plane H, with additional conditions such that H/Γ is a manifold and in particular a Riemann surface.

functional integral in our usage, synonymous with *path integral*.

fundamental region in relation to a coset space M/Γ where Γ is a discrete group, a region F such that every point in M is identified with exactly one point in the interior of F or with one or more points on the boundary of F.

fundamental string (F-string) the original string whose quantization defines a weakly coupled string theory, as distinguished from D-strings and solitonic strings.

fusion rule the specification of which conformal families appear in the operator product of any two primary fields in a given CFT.

gauge-fixing the reduction of a redundant (gauge-invariant) description of a quantum theory to a description with a single representative from each equivalence class.

gaugino a spin-$\frac{1}{2}$ fermion in the same supersymmetry multiplet as a gauge boson.

gaugino condensation a strong coupling effect where a product of gaugino fields acquires a vacuum expectation value. This generally breaks a chiral symmetry but does not directly break supersymmetry; however, in combination with other fields it often induces supersymmetry breaking.

generation a family of quarks and leptons, described by spinor fields in a chiral but anomaly-free set of $SU(3) \times SU(2) \times U(1)$ representations. In $SU(5)$ grand unification these become a $\mathbf{5} + \overline{\mathbf{10}}$, in $SO(10)$ they are contained in a $\mathbf{16}$, and in E_6 they are contained in a $\mathbf{27}$. An *antigeneration* is the conjugate representation; the distinction between generation and antigeneration is a matter of convention.

genus the number g of handles on a closed oriented Riemann surface: $g = 0$ is a sphere, $g = 1$ is a torus, and so on.

Gepner model a string model based on $N = 2$ *minimal model* CFTs.

ghosts see *Faddeev–Popov ghosts*.

Gliozzi–Scherk–Olive (GSO) projection a construction of modular-invariant string theories by summing over R and NS boundary conditions on the fermion fields and projecting onto states of definite world-sheet fermion number. In supersymmetric string theories there are independent GSO projections on the left-movers and right-movers. The diagonal projection, which acts simultaneously on both sides, produces a nonsupersymmetric theory.

goldstino the massless spin-$\frac{1}{2}$ Goldstone fermion associated with spontaneously broken supersymmetry. In supergravity it combines with the gravitino to form a massive fermion.

Goldstone boson the massless scalar corresponding to fluctuations of the direction of spontaneous symmetry breaking.

grand unification the unification of the $SU(3) \times SU(2) \times U(1)$ gauge symmetries in a simple group.

grand unification scale the mass scale of spontaneous breaking of the grand unified group. Proton stability and the unification of the couplings require that it be within two or three orders of magnitude of the gravitational scale.

Grassmann variable the elements θ_i of an algebra with the relation $\theta_i\theta_j = -\theta_j\theta_i$. These are used to give a path integral representation of fermionic fields, and to define *superspace*. They are also called *anticommuting c-numbers*.

gravitational scale the mass scale at which the dimensionless gravitational coupling becomes of order 1, $m_{\text{grav}} = \kappa^{-1} = 2.4 \times 10^{18}$ GeV; this is $(8\pi)^{-1/2}$ times the Planck mass.

gravitino a spin-$\frac{3}{2}$ fermion in the same supersymmetry multiplet as the graviton.

Green–Schwarz mechanism the cancellation of an anomaly by the modified transformation law of a p-form potential in a *Chern–Simons term.*

Green–Schwarz superstring a manifestly supersymmetric formulation of the supersymmetric string theories, with a spacetime-fermionic gauge invariance known as κ *symmetry*. There is no simple covariant gauge fixing.

***H*-monopole** a monopole carrying the magnetic charge of the antisymmetric tensor gauge field $B_{\mu n}$.

heterotic 5-brane the 5-brane carrying the magnetic charge of the massless heterotic string 2-form potential. It is obtained as the limit of a zero size instanton in the heterotic string gauge fields. The instanton configuration is localized in four spatial dimensions, and is therefore a 5-brane in nine spatial dimensions.

heterotic string a string with different constraint algebras acting on the left- and right-moving fields. The case of phenomenological interest has a $(0,1)$ superconformal constraint algebra, with spacetime supersymmetry acting only on the right-movers and with gauge group $E_8 \times E_8$ or $SO(32)$.

Hagedorn temperature the temperature at which the thermal partition function of free strings diverges, due to the exponential growth of the density of states of highly excited strings.

hidden sector the fields that couple to the Standard Model only through gravitational-strength interactions. In *hidden sector models*, these include the fields responsible for supersymmetry breaking.

highest weight state in CFT, a state annihilated by all Virasoro lowering operators, or more generally by all lowering operators in a given algebra.

Hodge number the number of nontrivial (p,q)-forms in Dolbeault cohomology, denoted $h^{p,q}$.

holographic principle the conjecture that the states of quantum gravity in d dimensions have a natural description in terms of a $(d-1)$-dimensional theory. This radical departure from local field theory was motivated by the black hole information problem, and has played a role in attempts to formulate M-theory.

holomorphic analytic, as used in the theory of complex variables. The Minkowskian continuation is *left-moving*. An *antiholomorphic* field is analytic in the conjugate variable, and its continuation is *right-moving*.

holomorphic quadratic differential a globally defined holomorphic $(2,0)$-form on a Riemann surface.

holomorphic vector field a globally defined holomorphic $(-1,0)$-form on a Riemann surface.

holonomy consider the parallel transport of a vector around a closed loop on a d-dimensional manifold: it returns to an $O(n)$ rotation of its original value. The set of all rotations that are obtained in this way for a given manifold is a subgroup of $O(n)$; this is the *holonomy group.*

homology see *cohomology.*

hyperelliptic surface a Riemann surface with a $\mathbf{Z}_2$ symmetry. This can be represented as a two-sheeted cover of the sphere with branch cuts.

hyper-Kähler manifold a $4k$-dimensional manifold of holonomy $Sp(k) \subset SO(4k)$. This is the geometry of the moduli space of hypermultiplets in $d = 6$, $N = 1$ or $d = 4$, $N = 2$ supersymmetry, in the limit in which gravity decouples.

hypermultiplet in $d = 6$, $N = 1$ or $d = 4$, $N = 2$ supersymmetry, the multiplet whose bosonic content is four real massless scalars.

identify to define two points (or other objects) to be equivalent, thus producing a coset space.

infrared (IR) divergence a divergence arising from long distances in spacetime, usually signifying that one has calculated the wrong thing.

inheritance principle in twisted (orbifold) theories, the principle that the tree-level amplitudes of untwisted states are the same as in the untwisted theory.

insertion the integrand of a path integral, excluding the weight $\exp(iS)$ or $\exp(-S)$.

instanton in a Euclidean path integral, a nonconstant configuration that is a local but not a global minimum of the action. Such configurations are usually localized in spacetime, are usually topologically nontrivial, and are of interest when they give rise to effects such as tunneling that are not obtained from small fluctuations around a constant configuration. *Spacetime instantons* are instantons in the effective field theory in spacetime. *World-sheet instantons* are instantons in the world-sheet quantum field theory, and correspond to world-sheets wrapping around nontrivial two-cycles of spacetime.

intersection number the number of points at which a set of surfaces intersect, weighted by the orientation of the intersection.

irrelevant interaction an interaction whose dimensionless strength decreases with increasing length scale. In perturbation theory, this is equivalent to a *non-renormalizable* interaction.

K3 manifold the unique nontrivial Calabi–Yau manifold of four (real) dimensions. To be precise, it is topologically unique, but possesses complex structure and Kähler moduli. Its holonomy is $SU(2)$, so that half of the supersymmetries of a theory are broken upon compactification on K3.

Kac determinant the determinant of the matrix of inner products of states at a given L_0 level of a Verma module.

Kac–Moody algebra see *current algebra.*

Kähler form the $(1,1)$-form $G_{i\bar{j}}dz^i d\bar{z}^{\bar{j}}$, formed from a Hermitean metric on a complex manifold.

Kähler manifold a complex manifold of $U(n)$ holonomy in n complex dimensions.

Kähler moduli the moduli parameterizing the Kähler form.

Kähler potential the potential $K(z,\bar{z})$, in terms of which the metric of a Kähler manifold is determined, $G_{i\bar{j}} = \partial_i \partial_{\bar{j}} K$. This is not globally defined, being determined only up to a *Kähler transformation* $K(z,\bar{z}) \to K(z,\bar{z}) + f(z) + f(z)^*$.

Kaluza–Klein gauge field in a compactified theory, a gauge field originating from the metric of the higher-dimensional theory. The gauge group is the isometry group of the compact space.

Kaluza–Klein monopole a monopole carrying the magnetic charge of a $U(1)$ Kaluza–Klein gauge symmetry. The monopole configuration is the smooth *Taub–NUT* spacetime. It is localized in three spatial dimensions, and is therefore a 6-brane in nine spatial dimensions.

Kaluza–Klein states states with nonzero momentum in a compact spatial direction.

Knizhnik–Zamolodchikov (KZ) equation the differential equation determining the expectation values of the primary fields of a current algebra.

Landau–Ginzburg model a scalar field theory which has a nontrivial IR fixed point when the potential is appropriately tuned. In particular, this gives a Lagrangian representation of the minimal model CFTs.

large coordinate transformation a coordinate transformation that is not continuously connected to the identity.

lattice the set Γ of integer linear combinations of n linearly independent basis vectors in n dimensions. Given a Euclidean or Lorentzian metric, an *even lattice* is one whose points have even length-squared. The dual lattice Γ^* is the set of points v such that $v \cdot w \in \mathbf{Z}$ for all $w \in \Gamma$. The *root lattice* is the set of integer linear combinations of the roots of a Lie algebra.

level 1. the quantized c-number term in a current algebra, also known as the *Schwinger term*; 2. the total oscillator excitation number in free field theory; 3. in a conformal family, the difference between the L_0 eigenvalue of a given state and that of the highest weight state.

level-matching the modular invariance condition that $L_0 - \tilde{L}_0 \in \mathbf{Z}$.

Lie algebra an algebra with an antisymmetric product that satisfies the Jacobi identity. A *simple* Lie algebra has no subalgebra that commutes with its complement. A *simply-laced* Lie algebra has all roots of equal length. A *graded Lie algebra* has odd and even elements, with a symmetric product between odd elements.

light-cone gauge in string theory, the choice of world-sheet time coordinate to coincide with a particular spacetime null coordinate. In theories with local symmetries, a gauge choice such that the connection in a given null direction vanishes.

linear dilaton theory a scalar CFT in which the energy-momentum tensor includes a term proportional to the second derivative of the scalar. This arises in string theory when the dilaton is a linear function of position.

linear sigma model a scalar field theory whose kinetic term is field independent, but whose long-distance physics is governed by a nonlinear sigma model.

Liouville field theory the CFT of a scalar field with an exponential interaction. This arises in various situations, including the *noncritical string*. It corresponds to bosonic string theory in a linear dilaton plus exponential tachyon background.

little string theory one of several interacting string theories without gravity, notably found on NS5-branes in the limit of zero string coupling.

loop expansion in quantum field theory, the Feynman graph expansion, which is equivalent to the expansion in powers of $\hbar$. The *string loop expansion* is the sum over Riemann surfaces, with dimensionless string coupling g. The *world-sheet loop expansion* is the nonlinear sigma model perturbation expansion, in powers of α'/R_c^2 with R_c the compactification radius.

Lorentzian having a mixed signature $(-,\ldots,-,+,\ldots,+)$.

lowering operator operators that reduce the energy of a given state. In CFT, operators that reduce the Virasoro generator L_0 (or $\tilde{L}_0$) by n units carry a grading (subscript) n.

M-theory 1. (*narrow*) the limit of strongly coupled IIA theory with eleven-dimensional Poincaré invariance; 2. (*broad; most common usage*) the entire quantum theory whose limits include the various weakly coupled string theories as well as M-theory in the narrow sense. The name is deliberately ambiguous, reflecting the unknown nature of the theory; M has variously been suggested to stand for membrane, matrix, mother, and mystery.

M2-brane the 2-brane of M-theory, which couples to the potential A_3 of eleven-dimensional supergravity.

M5-brane the 5-brane of M-theory, which carries the magnetic charge of the potential A_3 of eleven-dimensional supergravity.

macroscopic string a string whose length is much greater than the characteristic string length scale. In particular, it is sometimes useful to consider an infinite string stretching across spacetime.

Majorana condition a Lorentz-invariant reality condition on a spinor field. This can be imposed only if the spacetime dimension is 1, 2, 3, 4, or 8 (mod 8).

marginal interaction an interaction whose dimensionless strength is independent of the length scale. In general this might hold only to first order in the coupling of the interaction; a *truly marginal* interaction is one that remains marginal even with finite coupling.

matrix models quantum mechanical systems with matrix degrees of freedom, with critical points governed by noncritical string theories.

matrix theory a quantum mechanical system with matrix degrees of freedom and 32 supercharges, obtained by dimensional reduction of $d = 10$ supersymmetric $U(n)$ Yang–Mills theory. In the large-n limit this is conjectured to define M-theory (in the broad sense).

minimal models several families of solvable CFTs, in which every conformal family is degenerate. There are infinite series of unitary minimal models having $N = 0$, $N = 1$, and $N = 2$ superconformal symmetries, which converge from below on the central charges 1, $\frac{3}{2}$, and 3 respectively.

Minkowskian having a signature $(-,+,+,\ldots,+)$.

mirror symmetry an equivalence between string theories compactified on distinct manifolds. The equivalence reverses the sign of the $U(1)$ charge of one $N = 2$ superconformal algebra, and therefore changes the sign of the Euler number of the manifold.

Möbius group the globally defined $SL(2,\mathbf{C})$ conformal symmetry of the sphere; or, the globally defined $SL(2,\mathbf{R})$ conformal symmetry of the disk.

mode operators the spatial Fourier components of a quantum field.

model see *vacuum.*

modular group the group of large coordinate transformations (often applied to a Riemann surface but also applicable to spacetime).

modular invariance the invariance of the string path integral under large coordinate transformations.

moduli 1. the parameters labeling the geometry of a manifold. Notable examples are the parameters for the complex structure of the string world-sheet, and the parameters for the geometry of compactification; 2. the parameters labeling a space of degenerate (and, usually, physically inequivalent) vacua in quantum field theory. This is closely related to the compactification example: in expanding around the classical limit, each compact solution of the field equations gives a vacuum of the quantum theory, to leading order; 3. the massless fields corresponding to position dependence of these parameters. Contrast *Goldstone boson.*

moduli space the space of geometries or vacua, whose coordinates are the moduli.

monodromy for a quantity which is locally single-valued, the multi-valuedness around nontrivial closed paths.

Montonen–Olive duality the weak–strong duality of $d = 4$, $N = 4$ Yang–Mills theory.

Nambu–Goto action a string action, which is proportional to the invariant area of the world-sheet in spacetime.

Narain compactification the abstract description of toroidal compactification in terms of the lattice of left- and right-moving momenta.

naturalness problem the problem of explaining why a constant of nature takes a value much smaller than estimated nonzero contributions. Examples are the Higgs scalar mass, the cosmological constant, and the QCD θ-angle.

Neumann boundary condition the condition that the normal derivative of a field vanish at a boundary; the value of the field is free to fluctuate.

Neveu–Schwarz algebra the world-sheet algebra of the Fourier modes of the supercurrent and energy-momentum tensor, in a sector where the supercurrent is antiperiodic and its moding therefore half-integer-valued.

Neveu–Schwarz (NS) boundary condition the condition that a fermionic field on the world-sheet be antiperiodic, in the closed string or in the double of the open string (see *doubling trick*). Its Fourier moding is then half-integer-valued.

Neveu–Schwarz 5-brane in the type I and type II superstring theories, the 5-brane that carries the magnetic charge of the NS–NS 2-form potential.

Neveu–Schwarz–Neveu–Schwarz (NS–NS) states in type I and type II superstring theories, the bosonic closed string states whose left- and right-moving parts are bosonic. These include the graviton and dilaton, and in the type II case a 2-form potential.

no-ghost theorem 1. the theorem that the OCQ or BRST Hilbert space has a positive inner product; 2. the further theorem that the string amplitudes are well defined and unitary in this space.

no-scale model a field theory that has, to some approximation, a line of degenerate vacua with broken supersymmetry.

Noether's theorem the theorem that an invariance of the Lagrangian implies a conserved quantity.

noncommutative geometry a generalization of ordinary geometry, focusing on the algebra of functions on a space. The noncommutative collective coordinates of D-branes suggest the need for such a generalization.

noncritical string theory 1. a Weyl-noninvariant string theory — one with a measure of world-sheet distance that is independent of the embedding in spacetime. These include strings with an independent world-sheet metric field, and strings with a short-distance cutoff; 2. more recently, the term has been applied to any string theory that does not have a weakly coupled limit with a Weyl-invariant world-sheet theory. In this form it includes various theories with stringlike excitations in which the coupling is fixed to be of order 1, such as the *(2,0) theory* and the *little string theories.* Such a theory does not have a well-defined world-sheet, because processes that change the world-sheet topology cannot be turned off.

nonlinear sigma model a scalar field theory in which the kinetic term has a field-dependent coefficient. This has a natural interpretation in terms of a curved field space, and corresponds to string theory in curved spacetime, or more generally one with position-dependent background fields.

nonrenormalization theorem a theorem restricting the form of quantum corrections to a given amplitude, or to the effective action. It may require that these corrections vanish, that they arise only at specific orders of perturbation theory, or that they arise only nonperturbatively.

normal ordering a prescription for defining products of free fields by specific subtractions of divergent terms. *Conformal normal ordering*, denoted $: \; :$, produces operators with simple conformal properties. *Creation–annihilation normal ordering*, denoted $: \; :$, where lowering operators are put to the right of raising operators, produces operators with simple matrix elements. *Boundary normal ordering*, denoted $: \; :$, is conformal normal ordering with an additional image charge subtraction to produce operators that are finite as they approach a boundary.

null state a physical state that is orthogonal to all physical states including itself. Or, a descendant in a conformal family which is orthogonal to all states in the family.

old covariant quantization (OCQ) a method of quantizing string theory, similar to the Gupta–Bleuler quantization of electrodynamics. It is equivalent to the light-cone and BRST quantizations.

one-loop the leading quantum correction, coming from surfaces of Euler number zero in string perturbation theory.

open string a string that is topologically a line segment.

operator equation in quantum theory, an equality between operators that holds in arbitrary matrix elements; equivalently, an equality that holds when inserted into a functional integral with arbitrary boundary conditions.

operator product expansion (OPE) the expansion of a product of operators as a sum of local operators. This provides an asymptotic expansion, as the separation of the operators vanishes, for an arbitrary expectation value containing the product. In CFT the expansion is convergent.

orbifold 1. (*noun*) a coset space M/H, where H is a group of discrete symmetries of a manifold M. The coset is singular at the fixed points of H; 2. (*noun*) the CFT or string theory produced by the gauging of a discrete world-sheet symmetry group H. If the elements of H are spacetime symmetries, the result is a theory of strings propagating on the coset space M/H. A *non-Abelian orbifold* is one whose point group is non-Abelian. An *asymmetric orbifold* is one where H does not have a spacetime interpretation and which in general acts differently on the right-movers and left-movers of the string; 3. (*verb*) to produce such a CFT or string theory by gauging H; this is synonymous with the second definition of *twist*.

oriented string theory a string theory in which the world-sheet has a definite orientation; world-sheet parity-reversal is not treated as a gauge symmetry.

orientifold a string theory produced by the gauging of a world-sheet symmetry group H, where H includes elements that combine the world-sheet parity-reversal Ω with other symmetries.

orientifold plane a plane (of any dimension p) consisting of fixed points of the orientifold group H (specifically, of an element of H that includes Ω).

***p*-brane** a p-dimensional spatially extended object. Examples are *black p-branes*, *Dp-branes*, *M2-* and *M5-branes*, *NS5-branes*, *heterotic 5-branes*, and (in $d > 4$) *Kaluza–Klein monopoles*.

***p*-form** a fully antisymmetric p-index tensor, usually written in an index-free notation.

***p*-form gauge field** a generalization of Abelian gauge theory, with a p-form potential A, a $(p-1)$-form gauge parameter λ, and a $(p+1)$-form field strength F. For $p = 0$ this is an ordinary massless scalar; for $p = 1$ it is an Abelian gauge field. A $(p+1)$-form potential couples naturally to a p-brane, through the integral of the form over the world-volume. A *self-dual p-form theory* is one where $*F = F$; this requires the spacetime dimension to be $d = 2p + 2$, and further d must be 2 mod 4 (in the Minkowskian case) in order that $** = 1$.

***p*-*p*′ string** an open string with one endpoint on a Dp-brane and the other on a Dp'-brane.

(p, q)-form on a complex manifold, a tensor that is completely antisymmetric in p holomorphic indices and q antiholomorphic indices.

(p, q) string a bound state of p F-strings and q D-strings in the IIB theory.

parafermion CFT a family of coset CFTs with $\mathbf{Z}_n$ symmetry, generalizing the $\mathbf{Z}_2$-invariant free fermion theory. These describe the generic critical behavior of a system with $\mathbf{Z}_n$ symmetry.

parity transformation an operation that reflects one spatial dimension, or any odd number. One distinugishes *spacetime parity*, P (or β) and *world-sheet parity* Ω.

partition function a sum over the spectrum of a quantum system, weighted by $e^{-H/T}$ where H is the Hamiltonian and T the temperature. Often additional charges are included in the exponent. This is the basic object in equilibrium statistical mechanics. In string theory it is given by a path integral on the torus or the cylinder, and so arises in one-loop amplitudes.

path integral a representation of the transition amplitudes of a quantum system as a coherent sum over all possible histories. In quantum mechanics the history is a particle path; in quantum field theory it is a path in field space; in first-quantized string theory it is the embedding of the string world-sheet in spacetime.

Pauli–Villars regulator a means of regulating quantum field theories by introducing a very massive wrong-statistics field.

Peccei–Quinn (PQ) symmetry an approximate symmetry, violated only by anomalies.

period matrix a $g \times g$ matrix characterizing the complex structure of a genus-g Riemann surface.

perturbation theory the expansion of the amplitudes of a quantum system in powers of the coupling.

physical state in a quantum system with constraints, a state annihilated by the constraints. In OCQ this is a state annihilated by the Virasoro lowering generators and having a specified L_0 eigenvalue.. In BRST quantization it is a state annihilated by the BRST operator. In both these cases, the true physical spectrum is the space of physical states with an additional equivalence relation, physical states differing by a null state being identified.

picture in the RNS superstring, one of several isomorphic representations of the vertex operators. The q-picture consists of vertex operators of $\beta\gamma$ ghost charge q. The natural pictures are $q = -1$ and $-\frac{1}{2}$, with higher pictures including partial integrations over supermoduli space. The *picture changing operator* increases q by one.

Planck length the natural length scale of quantum gravity, $L_P = M_P^{-1} = 1.6 \times 10^{-33}$ cm, constructed from $\hbar$, c, and G_N.

Planck mass the natural mass scale of quantum gravity, $M_P = 1.22 \times 10^{19}$ GeV, constructed from $\hbar$, c, and G_N.

plumbing fixture a procedure for constructing higher genus Riemann surfaces from lower ones by sewing in a handle. The construction includes a parameter q, such that when q goes to 0 the handle degenerates or *pinches*. This gives a canonical representation of the boundary of the moduli space of Riemann surfaces.

Poincaré dual a map from p-forms to $(d-p)$-forms, given by contraction with the completely antisymmetric tensor.

Poincaré invariance the invariance group of the flat metric, consisting of translations and Lorentz transformations.

point group the orbifold group H, with translations ignored (applicable only for orbifolds having a spacetime interpretation).

Polyakov path integral a representation of first-quantized string theory as a path integral with an independent world-sheet metric. A local Weyl symmetry guarantees that the classical degrees of freedom are the same as those of the Nambu–Goto theory.

primary field in CFT, a local operator annihilated by all of the lowering generators of a given algebra, such as the Virasoro algebra. The corresponding state is a *highest weight state*.

projective space a compact n-dimensional space constructed from a linear $(n+1)$-dimensional space by identifying points under the overall rescaling $(x_1, \ldots, x_{n+1}) \cong (\lambda x_1, \ldots, \lambda x_{n+1})$. For x_i and λ real this produces RP^n, and for x_i and λ complex it produces CP^n (which has n complex $= 2n$ real dimensions).

pseudospin in a current algebra, an $SU(2)$ subalgebra not contained in the Lie algebra of the center-of-mass modes.

puncture a marked point on a Riemann surface, the position of a vertex operator.

QCD string a reformulation of non-Abelian gauge theory as a string theory, conjectured to exist at least in the limit of a large number of colors.

quaternionic manifold a $4k$-dimensional manifold with the holonomy group $Sp(k) \times SU(2) \subset SO(4k)$, with specific $SU(2)$ curvature. This is the geometry of the moduli space of hypermultiplets in $d = 6$, $N = 1$ or $d = 4$, $N = 2$ supergravity.

***R* symmetry** a symmetry that acts nontrivially on the supercurrent(s).

raising operator operators that reduce the energy of a given state. In CFT, operators that increase the Virasoro generator L_0 (or $\tilde{L}_0$) by n units carry a grading (subscript) $-n$.

Ramond algebra the world-sheet algebra of the Fourier modes of the supercurrent and energy-momentum tensor, in a sector where the supercurrent is periodic and its moding therefore integer-valued.

Ramond (R) boundary condition the condition that a fermionic field on the world-sheet be periodic, in the closed string or in the double of the open string (see *doubling trick*). Its Fourier moding is then integer-valued.

Ramond–Neveu–Schwarz (RNS) superstring the formulation of type I and II superstrings that has superconformal invariance but not manifest spacetime supersymmetry. The latter emerges after imposing the GSO projection on the string Hilbert space.

Ramond–Ramond (R–R) states in type I and type II superstring theories, the bosonic closed string states whose left- and right-moving parts are fermionic. These include p-form potentials C_p, with p taking all odd values in the IIA string and all even values in the IIB string.

rank 1. the maximal number of commuting generators of a Lie algebra; 2. the number of indices on a tensor.

rational CFT a CFT with a finite number of primary fields under an extended chiral algebra, a generalization of the *minimal models*. Such CFTs are highly constrained.

refermionization after bosonization, the construction of new spin-$\frac{1}{2}$ fields from linear combinations of the bosonic fields. These fermions are nonlocal, nonlinear functions of the original fermions.

Regge behavior in scattering at large center-of-mass energy-squared s and fixed momentum-transfer-squared $-t$, the scaling of the amplitude as $s^{\alpha(t)}$. The values of t where $\alpha(t) = j$ is a nonnegative integer correspond to exchange of a particle of spin j and mass-squared $-t$.

Regge slope denoted α', the square of the characteristic length scale of perturbative string theory. The tension of the fundamental string is $1/2\pi\alpha'$.

relevant interaction an interaction whose dimensionless strength increases with distance. In perturbation theory, this is equivalent to a *superrenormalizable* interaction.

renormalization group equation the differential equation governing the change of physics with length scale.

renormalization theory the calculus of path integrals.

Riemann–Roch theorem the theorem that the number of metric moduli minus the number of conformal Killing vectors on a Riemann surface is -3χ, with χ being the Euler number of the surface; and, generalizations of this result.

Riemann surface a two-(real)-dimensionsional complex manifold, equivalent to a Weyl equivalence class of Riemannian manifolds.

root a vector of the eigenvalues of the maximal set of commuting generators of a Lie algebra, associated with a state in the adjoint representation.

***S*-duality** a duality under which the coupling constant of a quantum theory changes nontrivially, including the case of weak–strong duality. Important examples are the $SL(2,\mathbf{Z})$ self-dualities of IIB string theory and of $d = 4$, $N = 4$ supersymmetric Yang–Mills theory. More loosely, it is used for weak–strong dualities between different theories, such as IIA–M-theory (on a circle) duality, $SO(32)$ heterotic–type I duality, and E_8 heterotic–M-theory (on an interval) duality. In compactified theories, the term *S*-duality is limited to those dualities that leave the radii invariant, up to an overall coupling-dependent rescaling; contrast *T-duality* and *U-duality*.

S-matrix the overlap amplitude between states in the infinite past and states in the infinite future; the scattering amplitude. In coordinate-invariant quantum theories this is generally the simplest invariant. The term usually implies a basis of free particle states; this is problematic in theories with massless particles due to IR divergences, and meaningless in theories at nontrivial IR fixed points.

scale transformation a rigid rescaling of spacetime, or of the world-sheet.

Scherk–Schwarz mechanism the breaking of supersymmetry by dimensional reduction that includes a spacetime rotation.

Schottky group a discrete subgroup Γ of the $SL(2,\mathbf{C})$ Möbius transformations of the sphere S_2, with additional conditions such that S_2/Γ is a manifold and in particular a Riemann surface.

Schwarzian the combination of derivatives appearing in the finite conformal transformation of the energy-momentum tensor.

Schwinger–Dyson equation the operator equations of a quantum theory, expressed as equations for the expectation values.

short multiplet see *BPS state*.

sigma model metric the metric appearing in the string world-sheet action. Also known as the *string metric*, this differs from the *Einstein metric* by a dilaton-dependent Weyl transformation.

simple current an operator J such that, for any primary field $\mathcal{O}$, the operator product $J\mathcal{O}$ contains only a single conformal family.

soliton a state whose classical limit is a smooth, localized, and (usually) topologically nontrivial classical field configuration; this includes particle states, which are localized in all directions, as well as extended objects. By contrast, a state of ordinary *quanta* is represented near the classical limit by small fluctuations around a constant configuration. In a theory with multiple classical limits (*dualities*), solitons and quanta may exchange roles.

special Kähler geometry the geometry of the moduli space of vector multiplets in $d = 4$, $N = 2$ supergravity. It is most simply defined (section B.7) in terms of *special coordinates*, which are fixed up to a symplectic transformation. *Rigid special geometry* is obtained in the limit where gravity decouples and the supersymmetry becomes global.

spectral flow the adiabatic change in the spectrum produced by a continuous change in the boundary conditions.

spin the behavior of a field or a state under rotations. In a CFT this is given in terms of the conformal weights by $h - \tilde{h}$.

spin field the vertex operator for a Ramond ground state, which produces a branch cut in the spinor fields.

spin structure one of a set of inequivalent ways of defining a spinor field globally on a manifold. Roughly speaking, it corresponds to a choice of signs in the square roots of the transition functions.

spurious state in OCQ, a state produced by Virasoro raising operators.

state–operator isomorphism in CFT, a one-to-one isomorphism between states of the theory quantized on a circle and local operators. Also, a one-to-one isomorphism between states of the theory quantized on an interval and local operators on a boundary. In d dimensions the circle becomes a $(d-1)$-sphere and the interval a $(d-1)$-hemisphere.

string coupling the dimensionless parameter g governing the weights of different Riemann surfaces in string perturbation theory, the contribution from surfaces of Euler number χ being weighted by $g^{-\chi}$. The string coupling is related to the dilaton by $g = e^{\Phi}$. This definition corresponds to the amplitude to emit a closed string; the amplitude to emit an open string is proportional to $g^{1/2}$.

string field theory the representation of string theory as theory of fields, the fields being maps from a circle (or interval) into spacetime. This corresponds

to an infinite number of ordinary quantum fields. This formalism can reproduce string perturbation theory, but it is unclear whether it can be defined beyond perturbation theory.

string metric see *sigma model metric.*

string scale the mass scale $\alpha'^{-1/2}$ characterizing the tower of string excitations.

string–string duality a term sometimes used to denote a weak–strong duality between different string theories, in particular between the heterotic string compactified on T^4 and the IIA string compactified on K3.

string tension the mass per unit length of a string at rest, related to the Regge slope by $1/2\pi\alpha'$.

Sugawara construction in current algebra, the construction of the energy-momentum tensor as a product of two currents. Originally proposed as a phenomenological model in four dimensions, this was later found to be an exact result in two dimensions.

superconformal algebra an extension of the conformal (Virasoro) algebra to include anticommuting spinor generators. The $(N, \tilde{N})$ superconformal algebra has N left-moving and $\tilde{N}$ right-moving supercurrents.

superconformal current algebra an extension of the conformal transformations to include both spin-$\frac{3}{2}$ and spin-1 currents.

superconformal field theory (SCFT) a quantum field theory that is invariant under superconformal transformations.

supercurrent a conserved spinor current. This includes the world-sheet current T_F associated with superconformal transformations, and the spacetime current associated with spacetime supersymmetry.

superfield a field on superspace, with specific transformation properties under a change of coordinates.

supergravity the union of general relativity and supersymmetry, implying also the promotion of supersymmetry to a local symmetry.

supermanifold (or *superspace*) a formal extension of the concept of manifold to include both commuting and anticommuting (Grassmann) coordinates.

supermoduli the anticommuting parameters characterizing a super-Riemann surface.

superpartner scale the mass scale of the superpartners of the Standard Model particles. This is expected to be between 10^2 and 10^3 GeV if supersymmetry solves the naturalness problem of the Higgs scalar mass.

superpotential in $d = 4$, $N = 1$ supersymmetry, the holomorphic function of the superfields that determines the nongauge interactions.

super-Riemann surface a supermanifold defined in terms of superconformal transition functions between patches.

supersymmetry a symmetry whose charge transforms as a spinor, which relates the masses and couplings of fermions and bosons.

supersymmetry breaking scale the mass scale of the expectation value that breaks supersymmetry. The superpartner scale is the supersymmetry breaking scale times the strength of the coupling of the Standard Model fields to the supersymmetry breaking fields.

***T*-duality** a duality in string theory, usually in a toroidally compactified theory, that leaves the coupling constant invariant up to a radius-dependent rescaling and therefore holds at each order of string perturbation theory. Most notable is $R \to \alpha'/R$ duality, which relates string theories compactified on large and small tori by interchanging winding and Kaluza–Klein states. More generally it includes shifts of antisymmetric tensor backgrounds and large coordinate transformations in spacetime. Contrast *S-duality* and *U-duality.*

't Hooft–Polyakov monopole a classical solution with magnetic charge, which exists whenever a simple group is spontaneously broken to a group with a $U(1)$ factor.

tachyon a particle (almost always a scalar) with a negative mass-squared, signifying an instability of the vacuum.

tadpole an amplitude for creation of a single particle from the vacuum, induced by quantum effects.

target space the space in which a function takes its values. This is usually applied to the nonlinear sigma model on the string world-sheet, where the target space is itself spacetime.

Teichmüller parameters the moduli for the complex structure of a Riemann surface (strictly speaking, points in *Teichmüller space* are not identified under the modular group).

tensionless string theory an interacting theory with tensionless strings. These can arise as p-branes with $p-1$ directions wrapped on a collapsing cycle, as various 2-branes with one direction stretched between higher-dimensional branes when the latter become coincident, and on zero-size $E_8 \times E_8$ instantons. In general the coupling is fixed to be of order one, so there is no perturbation expansion. *Tensionless* implies that the string tension in units of the gravitational scale goes to zero; it is not applied to the fundamental string, which becomes noninteracting in that limit.

tensor multiplet 1. the multiplet of $d = 6$, $(1,0)$ supersymmetry whose bosonic content is one self-dual tensor and one scalar. This reduces to a vector multiplet of $d = 4$, $N = 2$ supersymmetry; 2. the multiplet of $d = 6$, $(2,0)$ supersymmetry whose bosonic content is one self-dual tensor and five scalars. This reduces to a vector multiplet of $d = 4$, $N = 4$ supersymmetry.

tensor operator in CFT, a local operator whose conformal or superconformal transformation involves only the first derivative of the transformation; synonymous with a *primary field* of the conformal or superconformal symmetry. Such an operator is mapped to a *highest weight state* by the state–operator isomorphism.

tetrad a basis of d orthonormal vector fields in d dimensions. The term tetrad originates in $d = 4$ (as does the equivalent term *vierbein*) but there is no other convenient term for general d.

the theory formerly known as strings see *M-theory*, second definition.

theory see *vacuum*.

theta functions the family of holomorphic functions having simple periodicity properties on a torus.

Thirring model the solvable quantum field theory of a single Dirac fermion with a quartic interaction in $1 + 1$ dimensions. This is equivalent under bosonization to a free scalar field at a general real radius.

threshold correction a correction to the low energy effective action, and in particular to the gauge coupling, due to virtual massive particles.

topological string theory a modification of string theory without local dynamics; all observables are topological.

toric geometry a generalization of the idea of projective space. Roughly speaking, this corresponds to the most general linear sigma model.

toroidal compactification the periodic identification of one or more flat dimensions.

torsion a term applied to various 3-form field strengths, so called because they appear in covariant derivatives in combination with the Christoffel connection.

tree-level the Feynman graphs which become disconnected if one propagator is cut, or the analogous string amplitudes, the sphere and disk. These correspond to classical terms in the effective action.

twist 1. (*verb*) to define a field in a periodic space to be aperiodic by a symmetry transformation h; 2. (*noun*) the aperiodicity h; 3. (*verb*) given a CFT or string theory, to construct a new theory using a symmetry group H. One adds closed strings twisted by any of the elements $h \in H$, and requires all states to be invariant under the transformations in H. This is equivalent to treating H as a world-sheet gauge symmetry. The term *orbifold* is also used as a synonym; 4. (*noun, archaic*) world-sheet parity.

twisted state a closed string with twisted periodicity.

type ... supergravity the low energy supergravity theory of the corresponding string theory. The Roman numeral signifies the number of $d = 10$ supersymmetries, and IIA and IIB distinguish whether the two supersymmetries have opposite or identical chiralities respectively.

type I superstring the theory of open and closed unoriented superstrings, which is consistent only for the gauge group $SO(32)$. The right-movers and left-movers, being related by the open string boundary condition, transform under the same spacetime supersymmetry.

type IIA superstring a theory of closed oriented superstrings. The right-movers and left-movers transform under separate spacetime supersymmetries, which have opposite chiralities.

type IIB superstring a theory of closed oriented superstrings. The right-movers and left-movers transform under separate spacetime supersymmetries, which have the same chirality.

type 0 string a pair of nonsupersymmetric string theories, which have tachyons and no spacetime fermions, with the same world-sheet action as the type II theories but with different projections on the Hilbert space.

***U*-duality** any of the dualities of a string theory, usually of a toroidally compactified type II theory. This includes the *S-dualities* and *T-dualities*, but in contrast to these includes also transformations that mix the radii and couplings.

ultrashort multiplet see *BPS state*.

unit gauge a choice of coordinate and Weyl gauges in two dimensions, such that the metric is the identity.

unitary as applied to a quantum system, the property of having a conserved inner product in a positive-norm Hilbert space.

unoriented string theory a string theory in which world-sheet parity-reversal Ω is a discrete gauge symmetry. The perturbation theory includes unoriented world-sheets, and the spectrum is restricted to states with $\Omega = +1$.

UV divergence a divergence arising from short distances in spacetime, usually signifying a limit to the validity of a theory.

vacuum a stable Poincaré-invariant state. The novel feature of systems with unbroken supersymmetries is the frequent appearance of degenerate but physically inequivalent vacua.
While it is now clear that the different string theories are actually different *vacua* in a single theory, it is still common to use the term *theory* for each. Also, different CFTs within a single string theory are sometimes referred to as different *theories* rather than *vacua*. The term *model* is used to refer to string vacua whose low energy physics resembles the Standard Model.

vector multiplet in $d = 4$, $N = 1$ or $d = 6$, $N = 1$ supersymmetry, the multiplet whose bosonic content is a massless vector field. The $d = 6$ multiplet reduces to a vector field plus two real scalars in $d = 4$, $N = 2$ supersymmetry.

Veneziano amplitude the bosonic string tree-level amplitude for four open string tachyons.

Verma module the set of states obtained by acting on a highest weight state with Virasoro raising generators in all inequivalent ways, with the requirement that all such states be linearly independent (the generators act freely). This gives a representation of the Virasoro algebra that depends on the central charge c and on the weight h of the highest weight state. This is

very similar to a *conformal family*, except that in the case of a degenerate representation some states in the latter may vanish; the Verma module would in this case give a reducible representation.

vertex operator a local operator on the string world-sheet, corresponding to a string in the initial or final state.

Virasoro algebra in a CFT, the infinite-dimensional Lie algebra of the Fourier modes of the energy-momentum tensor.

Virasoro–Shapiro amplitude the bosonic string tree-level amplitude for four closed string tachyons.

***W* algebras** a family of extended chiral algebras with currents of spin greater than 2.

***W* string** a proposed generalization of string theory with constraints of spin greater than 2. No complete construction exists.

Ward identity a relation between the divergence of the expectation value of a conserved current and the same expectation value without the current.

weight 1. in CFT, the L_0 or $\tilde{L}_0$ eigenvalue, which determines the behavior of an operator under scale transformations and rotations (= *conformal weight*); 2. in Lie algebra, a vector of the eigenvalues of the maximal set of commuting generators of a Lie algebra, in a state of a given representation.

Wess–Zumino consistency condition the condition that the second derivative of a functional integral with respect to the background fields be symmetric. This strongly constrains the form of possible anomalies in a theory.

Wess–Zumino–Novikov–Witten (WZNW) model a conformally invariant nonlinear sigma model on a group manifold with an antisymmetric tensor background.

Weyl anomaly an anomaly in the Weyl transformation, this determines the critical dimension in the Polyakov formalism. This is sometimes called the *trace anomaly* or the *conformal anomaly*.

Weyl condition in even dimensions, the condition that a fermion have definite chirality. This defines a spinor representation of the Lorentz group, which contains half of the components of the Dirac representation of the gamma matrix algebra.

Weyl transformation a position-dependent rescaling of the metric.

Wilson line 1. a gauge field with vanishing field strength but with nontrivial parallel transport (holonomy) around nontrivial paths in spacetime. 2. the gauge-invariant operator that measures such a field: the trace of the path-ordered product of the line integral of the vector potential.

Wilsonian action the action in a low energy effective field theory, which incorporates the effects of higher energy virtual states.

winding state a closed string whose configuration is a nontrivial path in a non-simply-connected spacetime.

world-sheet 1. the two-dimensional surface in spacetime swept out by the motion of a string. 2. the abstract two-dimensional parameter space used to describe the motion of a string.

world-sheet duality the equivalence between Hamiltonian descriptions obtained by cutting open a string world-sheet along inequivalent circles. Important examples are the associativity condition, from the four-point sphere amplitude, and the equivalence of the open string loop and closed string tree descriptions of the cylinder.

wrapped refers to a p-brane, q of whose dimensions are wound on a nontrivial compact submanifold of spacetime, leaving a $(p-q)$-dimensional extended object.

Yang–Mills field non-Abelian gauge field.

Zamolodchikov metric the expectation value of a pair of local operators on the sphere, giving a natural inner product for the corresponding string states. This is the metric that appears in the kinetic term in the spacetime action.

zero modes 1. of a differential operator: eigenfunctions with zero eigenvalue. In a Gaussian functional integral, these would give an infinite factor in the bosonic case and a zero factor in the fermionic case. In general these have a physical origin and the functional integral has an appropriate integrand to give a finite result; 2. center-of-mass modes.

zero-point energy the energy due to vacuum fluctuations of quantum fields.

zero-slope limit the limit $\alpha' \to 0$. Only massless string states, described by low energy field theory, remain.

Index

Page numbers in italics refer to exercises

(0, 2) compactification 324–6, 376, 472
(2, 0) (tensionless string) theory 204–5
(2, 2) compactification
  Calabi–Yau 311–12, 318–19, 323–5
  Gepner models 394–402
  of heterotic strings 386–90
    constraints on spacetime action 388–90, 473
  of type II strings 379–85
  orbifold 292, 297

A–D–E singularity 425
A_{n-1} *see* $SU(n)$
action (D-brane) 149–50
  D0-brane 161
  Dp-Dp' system ($\#_{ND} = 4$) 163–4, 172
action (general)
  Pontrjagin term 173, 187, 333
action (spacetime)
  anomaly-cancelling terms 99–100, 130
  $d = 4$, $N = 1$ supergravity 446, 474
  $d = 4$, $N = 1$ supersymmetry 443, 474
  $d = 4$, $N = 4$ Yang–Mills theory 443
  $d = 10$, $N = 1$ Yang–Mills theory 443
  eleven-dimensional supergravity 85, 453, 468, 474
  F^4 term 116, 132
    and Born–Infeld action 152, 469
    and duality 193–4
  from (2,2) compactification 473
    heterotic string 388–90
    type II string 381–4
  from Calabi–Yau compactification 315–20, 404–6, 472
    higher corrections 321–4
  from orbifold compactification 472
    threshold corrections 298–300
    twisted states 296–8
    untwisted states 292–6
  in $N = 1$ heterotic string vacua 359–62, 473
  p-form gauge field 451–2
  R^4 term 117
  R–R tadpole term 39
  singularities in 411
  type I supergravity 92–4
    type I–heterotic relation 93, 192
  type IIA supergravity 87–8
    massive 89–90
  type IIB supergravity 90–2
    $SL(2, \mathbf{R})$ symmetry 91, 181–2
action (world-sheet)
  BC SCFT 4, 106
  heterotic string 49
  Landau–Ginzburg 270, 392
  linear dilation 5
  linear sigma model 398–9
  $N = 2$ free SCFT 48
  nonlinear sigma model 106–7
  superfield form 105–7

superstring 2
WZNW 247–50
affine Lie algebra 66
annulus, *see* cylinder
anomalies 94–103, 468
anomaly polynomial 97–101, *374*
cancellation for $SO(32)$ and $E_8 \times E_8$ 99–101
gauge 95ff
global gravitational 34
gravitational 95ff
in six dimensions 417, 420
in two dimensions 94–6, 107
mixed 95ff
see also Green–Schwarz mechanism
anti-de-Sitter space 223
antisymmetric tensor gauge field, *see* *p*-form gauge field
Atiyah–Drinfeld–Hitchin–Manin (ADHM) construction 174
auxiliary field 442–4
axion 79, 286, 315, *373–4*
field (a) 293–4
and strong-CP problem 333–5

bc CFT, bosonization 15
BC SCFT 4–5
B_k *see* $SO(2k+1)$
$\beta\gamma$ CFT 15–20
bosonization 17–20, 118–19
BPS bound 449
D0–Dp system 168
F-string–D-string system 164, *176*, 182
BPS states 449, 474
behavior at strong coupling 180–1
black p-branes 183, 220
bound states 164–75
D-branes 140–5
Kaluza–Klein states 456
macroscopic strings 81, *226*, 468
stretched strings 186–7, *226*
winding states 79–80
wrapped branes 404, 411–12
zero-force property 146
BRST charge 15–16, 24, *43*, 50
and picture changing 118–21, 125
in topological string theory 385
BRST quantization 24–5, *43*, 45, 50
beta function
and scale transformation 261, 264–6
for gauge coupling 298, 346, 367
vanishes for $d = 4$, $N = 4$ 460
nonlinear sigma model 247, 321
Betti numbers 305, 308
Bianchi identity
Chern–Simons term in 86, 88, 91
for R–R vertex operators 88–9
in Calabi–Yau compactification 304, 324, 418
related to field equation by Poincaré duality 141, 451–2
black hole
entropy 219–25, *227*, 470
information paradox 225–6
produced in graviton–graviton scattering 209
black p-brane 183–5, 201, 219–20, *227*, 470
blow up (a fixed point) 297, 309–11, 388, 471
bosonization 11–15, 279–80, 467
and refermionization 131–2
bc CFT 15
$\beta\gamma$ CFT 17–20, 118–19
current algebra fermions 54–5, 74–5
higher genus 13
kink operator 12
of $U(1)$ current 68
Ramond states 13–15
twisted fermions 13–14
bound states
D0–D0 168–9, 469
and M-theory 198
and matrix theory 212–14
D0–D2 169–70
U-dual to FD-string 189
D0–D4 170–5, *176*, 469–70
and instantons 172–5
U-dual to winding string 189
D0–D6 175
D1–D5 black hole 220–3

F-string–D-string 164–7, *176*, 469
$SL(2,\mathbf{Z})$ duality 182
U-dual to D0–D2 state 189

C invariance 330–1
C_k *see* $Sp(k)$
c-map 382
CP invariance 331, 333–5, 472
CPT invariance 331–2
c-theorem 262–3
and supersymmetry multiplets 441–2, 447
Calabi–Yau compactification 302–24, 472–4
and Gepner models 397–402
as special case of (2,2) compactification 311–12, 386–90
field theory approximation 315
low energy field theory 315–20
massless spectrum 312–15
of type II string 379–86
Calabi–Yau manifold 309–12
Casimir energy 145
Casimir invariant 60
central charge (conformal) 3–5, 15, *43*
and BRST nilpotence 24
and density of states 222, 236
and rank of gauge group 421
and renormalization group 262–6, 271
in heterotic string 49–50
in rational CFT 258–9
in topological string theory 385
in unitary CFT 228–33
of ghosts 46–8
specific
coset models 250–4
current algebra 70–2
$N = 0$ minimal models 231
$N = 1$ minimal models 254–5
$N = 2$ minimal models 391
superconformal current algebra 340–2
WZNW 247–8
central charge (supersymmetry)
and BPS states 449, 456
from dimensional reduction 454, 456–7, *466*
in extended supersymmetry 448–9
in heterotic string 79–82
in type I/II string 144–5, 475
central extension 67
Chan–Paton degrees of freedom 30
for type I D5-brane 195–7
reinterpreted as D-brane label 141–2
Chan–Paton factor 73, 109, 468
character 235–6, 245–6
charge conjugation matrix C 434
Chern class (first) 309, 366, 402
Chern–Simons 3-form 92, 100
and modified gauge transformation 92, 107, 469
Chern–Simons terms 86, 453
and anomaly cancellation 99–100
from string perturbation theory 113–14, 127–30
in D-brane action 150
and induced R–R charge 170, 173, 203
chiral gauge couplings 282, 325, 330
constraints from 337, 340–2, 351–2, 447–8
in (2,2) models 386–8
in Calabi–Yau models 313
in orbifold models 288–9, 291
chiral multiplet 315, 441–2
chiral primary fields 378–9, 384–5
in Landau–Ginzburg models 394–5, 397
chiral ring 384–95, 473
chiral superfields 442–7
chirality
and anomalies 96–8, *101*
effect of T-duality on 137–8
IIA vs IIB 27–9, 415–16, 453–4
matrix (Γ) 10, 432
of NS5-branes 204
Christoffel connection *106*
Clifford (Dirac matrix) algebra 2, 193, 430
cocycle 19–20, *83*
cohomology

and homology 306, 317
BRST 25, 385–6
De Rham 305
Dolbeault 307
of K3 416–17
ring 385
Coleman–Weinberg formula 31, 156
collective coordinate 139–40, 149
for eleventh dimension 202, 204
commutators and anticommutators
$[\alpha_m^\mu, \alpha_n^\nu]$ 6
$\{b_m, c_n\}$ 16
$\{\beta_r, \gamma_s\}$ 16
$[E_\alpha, E_\beta]$ 61
$\{G_r, G_s\}$ 7
$[H^i, E_\alpha]$ 61
$[j_m^a, j_n^b]$ 66
$[L_m, G_r]$ 7
$[L_m, j_m^a]$ 66
$[L_m, L_n]$ 7
$\{\psi_r, \bar{\psi}_s\}$ 14
$\{\psi_r^\mu, \psi_s^\nu\}$ 6
$N = 2$ superconformal 376
$\{Q_B, b_n\}$ 24
$[Q_B, \beta_r]$ 24
$[T^a, T^b]$ 59
see also supersymmetry algebras
compactification 274
scale 344, 350
see also (0, 2), (2, 2), Calabi–Yau, Gepner, orbifold, toroidal compactifications
complex manifold 306–7
complex structure 306–7
hyper-Kähler manifold 463
large 403, 405
complex structure moduli 291, 314–15, 381
conifold singularity 409–15
effective action 319–20, 472
K3 416
moduli space 296, 383–4, 402–9
orbifold 291, 296, 299
quintic 310
conformal block 235, *272*
differential equation 237–8
in rational CFT 255–7
conformal bootstrap 233–6
conformal family 229
degenerate 232, 239
conformal field theory (CFT)
$c < 1$ 236–42
$c = 1$ 35–7
density of states 222, 236
irrational 258–9, 471
nonunitary 232
rational 255–8, 471
unitary 45, 228–32
see also bc, $\beta\gamma$, ψ, coset, Landau–Ginzburg, parafermion CFTs, current algebras, minimal models, superconformal field theory
conformal gauge 20
conformal invariance
constraints on correlators 233–5
vs scale invariance 260, 262–3
conformal perturbation theory 263–6
conifold 405, 409–10, 474
massless 3-brane 411–12
singularity in CFT 410–11
transition 412–15
constraint algebra 3, 20, 24, 45–9
coset CFT 250–4, *272*, 471
from gauged world-sheet symmetry 251
minimal models 250–1
$N = 1$ minimal models 255
parafermion CFTs 251–2
W algebras 253–4
cosmological constant 226, 372–3, 473
coupling constants
gauge 298, 361
unification 314, 345–52
relations among 107–10, 150–2, 335–6, 472
string vs M-theory 199
Yukawa 317, 331
critical dimension 4, 47–8, 198
critical phenomena 266–71, 393, 400
current algebra 66–73, 243–50, *272*, 468, 471

and coset models 250–4
and unification 337–8, 351, 353–4
bosonic representation 74–5
fermionic representation 68
gauging 94–5
KZ equation 244
level 67
modular invariants 245–6, 249–50
primary fields 72–3
restriction on representations 72–3, 289, 337–8, 349
Sugawara construction 69, *82*
superconformal 338–42
WZNW models 246–50
cycle 306, 319–20
collapsing 404, 410–13, 425, 427
D-brane wrapped on 404, 411–13, 425
intersection number 317–18
monodromy 410
world-sheet wrapped on 322–4, 406, 408
cylinder
D-brane interactions from 146–7, 155–60, *176*, 196
in type I theory 37–9, *135*

D-branes 138–77, 469–70
action 149–50
and black *p*-branes 219–23, *227*
D1–D5 black *p*-brane 220–23
as BPS states 140–1, 145, *226*
as R–R sources 141–2
charge 146–7
coincident, non-Abelian dynamics 149, 166, 211–23
D0-brane quantum mechanics 161–2
in matrix theory 211–14, 217–18
D0-branes as Kaluza–Klein states 198–9
D1-branes and IIB duality 180–2
D3-branes and Montonen–Olive duality 180–2
Dp–Dp' system 152–8, 162–4, *175–6*
BPS bound 168
effect of T-duality on 143–4, *429*
gauge coupling 151
in type I theory 190–7
interactions between *175–6*
parallel 146–7
rotated 154–8
velocity-dependent 158–61
on collapsing cycle 411–15, 425
p even, relation to M-theory 200–8
substring length scale 160–2, 412
tension 146, 150
see also bound states
D-instanton 142, 145, 334, 372, 469
D_k *see* $SO(2k)$
D-terms
four supersymmetries 364–6, 399, 443–4, 473
and orbifold blow-up modes 297
eight supersymmetries 462
in Dp-Dp' system 163, 172, 221–2
descendants 229, 233–4, *271*
diagonal modular invariant 36
current algebra 246, 249
heterotic string 55
type 0 superstring 35
differential forms 305–6, 450–2
dilaton
and string loop expansion 87–9, 360–2, 383–4
coupling to R–R fields 87–9
in chiral multiplet 294, 315
in hypermultiplet 380, 383–4, *429*
in vector multiplet 428
dilaton-mediated supersymmetry breaking 370
dimensional reduction
of action 84–6, 292–5, *374*
of couplings 206, 335
of $d = 10$ Yang–Mills 459–61
of spectrum 78–80, 84–6, 188, 453–5
of supersymmetry 79–80, 311, 454, 458, 463–4
to $d = 2$ 392, 399, 449, *466*
Dirac–Born–Infeld action 149
Dirac equation 1–2, 22, 88, 179
Dirac matrices 1–2, 8, 430

Dirac quantization condition 147–9, 183–5, 196, 247–8, 355, 366
Dirac spinor 8, 430–2
Dirichlet boundary condition 139, 142, 143, 152
discrete light-cone quantization (DLCQ) 217
discrete series, *see* minimal models
discrete symmetry (spacetime) 328–35, 472
  $\mathbf{Z}_5$ example 397–8
  see also S-, T-, U-dualities
discrete torsion 58, 468
disk amplitudes 102, 110–12, 114–16, *134*, *175*
  tadpole 38–9, 42, 147
divergences, cancellation 37–42, 102–3, *135*, 142–3, 468
doubling trick 7–8, 153
dual Coxeter number 63–4, 71
duality, *see* Montonen–Olive, S-, string–string, T-, and U-dualities

E_6 64–5, *82*, 285ff, 312, 386–8
  singlets 313, 315, 324–6, 388
  see also grand unification
$E_{6(6)}$ 188–9
E_7 *82*
E_8 54, 63–4, *82–3*, 285, 290
  and anomalies 100, 110, *134*
  root lattice 74–5
Eguchi–Hanson space 309–10
electroweak scale 343, 369
energy-momentum tensor 1–5, 50
  in bosonization 12, 15, 18
  in type I theory 37–9, *135*
  topological 385
enhanced gauge symmetries
  at Gepner point 398
  at orbifold points 309–10, 312–13
  on coincident branes 149, 185–6
  on singular manifolds 425, 427
  on small $SO(32)$ instantons 195–6
  winding state 66, 76–7
Euler number (Calabi–Yau) 305, 309, 414–15
extended chiral algebra 253–4, 256
extended supersymmetry 359, 440, 447–9, 473–4
exterior derivative 305, 307, 450

F-terms 399, 443–5
F-theory 200–1, 474
Faddeev–Popov ghosts 15–20, 46–7
Fayet–Iliopoulos term 364–5, 399, 443, 446
Feigin–Fuchs representation 241–2, *272*
fermion number
  spacetime 32
  world-sheet 8–9, 19, 22ff, 32–3
Feynman diagrams 102, 176, 316
fibrations 404, 428
fixed point
  orbifold 275, 284, 286
    blowing up 297, 309–10, 388, 471
  field theory 266
    nontrivial 260–1, 415, 421, 460–1
flipped $SU(5)$ 348
flop 406–9, 474
fractional charge 352–5
fractional string theory 46, 468
free fermion models 279–80, 471–2
free parameters, absence of 369, 373
fundamental (F-) string 145, 150
fundamental region 31, 40, 284
fusion rule 238–40, 244–5, 256–8, 270–1

gauge coupling, *see* coupling constants, gauge
gauge-fixing 20
gauginos 30, 315
  condensation 367–9
general relativity 183, 408, 429
generation (quark and lepton) 64–5, 288–92, 313–15, 324–6, 387–8, 471–2
  gen.-changing transition 414–15, 474
Gepner models 394–405, 473
  relation to Calabi–Yau 397–402
Gliozzi–Scherk–Olive (GSO) projection 27, 30, 33

and spin–statistics 29
and supersymmetry 50–1, 359
current algebra fermions 51–5, *82*, 193, 386
diagonal 28, 35–7, 268
global symmetry 45, 94, 327–8, 334–5
goldstino 140
Goldstone boson 140
grand unification 63–5, 81, 281, 288–90, 351–6
partial 348
scale 298–9, 345–50
symmetry breaking 289, 325–6, 337–8, 349
gravitational scale 180, 344
eleven-dimensional 199, 212
gravitino 28–31, 79, 286, 315, 441
world-sheet 126
Green–Schwarz mechanism 99–101, 365, 462
Green–Schwarz superstring 29, 49, 133

H-monopole 82
Hagedorn temperature *83*
Hamiltonian, D0-brane 161–2, 212
heterotic 5-brane 195
heterotic string 45–83, 468
(0,2) and (1,2) 48
amplitudes 112–13, 126–33, *134*, *373–4*
bosonic form 73–5
$E_8 \times E_8$ 53–5
at strong coupling 205–8
fermionic form 49–50
nonsupersymmetric 55–9
$SO(16) \times SO(16)$ 57–9
$SO(32)$ 51–3
at strong coupling 190–4
T-duality between $SO(32)$ and $E_8 \times E_8$ 78
toroidal compactification 76–82, 422–9
hidden sector 289, 367–71
hierarchy (Higgs naturalness) problem 281–2, 350, 369
highest weight (primary) state 229ff
current algebra 72–3, 243
$N = 2$ chiral 378–9, 384–5
superconformal 105
Hodge numbers 309–10, 315, 402–3, 414, 416
holomorphic fields 45–6, 255
holonomy 304ff, 415–17, 463–4
homology 306, 317
hyper-Kähler manifold 417, 463–4
hypermultiplet 380–4ff, 447, 461–4
from p-p' strings 163, 172–3, 196

index theorem 330
inheritance principle 290, 293, 314, 351
instanton (Pontrjagin) number 187, 333
instantons 334–5
D- 142, 145, 334, 372, 469
world-sheet 323–5, 383, 406, 408, 472
Yang–Mills 172–5, 194–6, 418–20, 470
intersection number 317–18
irrelevant interaction 192, 264
Ising model 266–70
tricritical 269

Jacobi's abstruse identity 35

K3 manifold *301*, 415–18
and heterotic string 418–21
and type II string 415–18, 425–8
Kac determinant 230–3, 242, 254, *271*, 471
Kac–Moody algebra 66
Kähler class 308
Kähler form 307–8
Kähler manifold 307–8, 446
and supersymmetry 311, 446
Ricci-flat 308–11
see also hyper-Kähler manifold, special Kähler manifold
Kähler moduli 291, 315, 381
effective action 316–19, 323, 381–4, 472

K3 416
moduli space 296, 381–3, 402–9
Kähler cone 406–8
Kähler potential 294–8, 308, 359–62, 390, 446–7
Kähler transformation 296, 308, 446–7
Kaluza–Klein gauge symmetry 76, 79, 86, 188–9
and supersymmetry algebra 454–6
vertex operators 66
Kaluza–Klein monopole 81, 205
Kaluza–Klein states 198
Klein bottle 39–41
Klein–Gordon equation 1
Knizhnik–Zamolodchikov (KZ) equation 244, *272*

Landau–Ginzburg model 270–1, 392–4, 400, 471, 473
large-N limit 223
large order behavior 145
lattice (momentum) 19, 74–8, *83*
Laurent expansion, *see* mode expansion
length, minimum 160–2, 412
level (current algebra) 67
level-matching 25, 57, 276–9, 284–5ff, 395, 468
Lie algebra 59–65, 468
and grand unification 63–5
normalization of generators 67–8, 93, 112, 151, 336
roots and weights 61–3, 74–6
simply-laced 63
light-cone quantization 212, 216
calculations 117–18, 131–3
discrete (DLCQ) 217–18
lightlike limit 217–18
linear dilaton CFT 5, 15, *43*, 230
applications 89, 185, 241–2, *272*
linear multiplet 294
linear sigma model 398–402, 408
Lorentz generators 8–9, 431
Lorentz invariance 9, 331
eleven-dimensional 212
local 96

M-theory 198–205, 208–11, 426–7, 470
on $S_1/\mathbf{Z}_2$ 207–8, 350, 419–20
M2-brane 201–2, 470
in matrix theory 214–16
stretched 204–5, 421
wrapped 427
M5-brane 201–5, 419–21, 470
macroscopic string 81
magnetic monopoles 81–2, 187, 205, 353, 468, 470
Dirac quantization condition 147–8, 355
Majorana spinor 433–4, 437
Majorana–Weyl spinor 433
manifold
Calabi–Yau 309–11
complex 306–7
group 246–50
hyper-Kähler 463
Kähler 307–8
quaternionic 464
real 305–6
$SU(3)$ holonomy 308–9
special Kähler 464–5
marginal interaction 264–5
mass-shell condition 1, 20, 23
matrix theory 211–18, 226, 470
Maurer–Cartan 1-form 248
measure, supermoduli 124–6
membrane, *see* M2-brane
Mermin–Wagner–Coleman theorem 140
minimal models 236–42
coset construction 250–1
Feigin–Fuchs representation 241–2
fusion rule 238
Ising model 267–9
Landau–Ginzburg models 270–1
mirror symmetry 402–4, 473–4
and conifold transition 401, 414
and flop transition 406–9
and Kähler moduli space 404–9, 428–9
Möbius invariance 233
Möbius strip 39, 41–2, *135*

mode (Laurent) expansion
  b, c, β, γ 16
  free scalar (X^μ) 6
  j^a 66
  $N = 2$ superconformal 376
  p-p' strings 153–5
  ψ 6–8, 14
  T_B, T_F 7
model-independent axion 333
modular invariance 33–7, 235–6, 471
  A–D–E invariants 246, 250–1
  and anomalies 103
  current algebra 245–6, 249–50
  Gepner models 395
  heterotic string 52–3, 55–8, *82*, 388–9
  minimal model 251
  orbifold 276–9
  rational CFT 268
  superstring 26–7, 33–5
modular transformations 34, 38, 156
modular weights 297
moduli, physical effects 291, 324–6, 331, 370–3, 445
  in threshold correction 298–300, 349
moduli space
  (2, 2), product form 381, 389–90
  IIB theory 91–2
  as spacetime 404, 417
  branches 172–4, 373, 412–15, 419–21
  D1–D5 system 221–2
  dilaton–axion 296, 428, 457
  for toroidal compactification 76–8, 81, 91–2, 427–8, 455–8
  K3 417–19, 427–8, 474
  metric on 362, *374*
  orbifold 296–7, 472
  with 8 supersymmetries 462–5, 475
  with 16 supersymmetries 159, 460, 475
  see also complex structure, Kähler moduli
monodromy 240, 252, 410, 471
Montonen–Olive duality 186–7, 428, 470
multicritical behavior 268–70
NS algebra 7, *43*
NS boundary condition 5–8, 15–16
  vertex operators 10–13, 16–18
NS–NS charge 81
  in supersymmetry algebra 81–2
  magnetic 81–2, 145, 183–5, 189
NS–NS fields 87
  background 106
  tadpole 37–42
NS5-brane 470
  IIA 202–4, 426
  IIB 182–6, 202–4
    multiple 186
  heterotic 195
Narain compactification 73–9, *83*, 468
naturalness problem 282, 350, 369
Neumann boundary condition 141, 143, 152
no-ghost theorem 25, 50, 232–3
no-scale models 371
noncommutative geometry 149, 161
nonlinear sigma model 106–8, 247, 311, 469, 471
nonperturbative definition (of string theory) 180–1, 208–11
nonperturbative effects 49, 134, 145, 178, 343, 366–73, 421–2, 425
  see also instantons, strong coupling
nonrenormalization theorems 133–4, 300
  α' corrections 111–17, 185, 321–5, 382–3, 472
  perturbative, violation by instantons 323–5, 367–8, 372, 472
  string corrections 469, 473
    4 supersymmetries 361–2, 364–5, 444
    8 supersymmetries 381, 383–4, 410–11, 428–9
    16 supersymmetries 133–4, 159, 194, 460–1
    32 supersymmetries 133–4, 180
normal ordering 2, 7, *43*
null state 231–3, 236–9, 243–5, 252–3
  decoupling of 98, 101–2, 119, 121

old covariant quantization (OCQ) 3, 20–3, 25
one-loop string amplitudes 31–42, 126–34, *135*, 146–7, 155–60, 298–300, *374*, 468–9, 472
operator product expansion (OPE) *43*, *82–3*, 95–6
  and bosonization 11–12, 17–19
  $B\,C$ 106, *134*
  $\beta\gamma$ CFT 17–19
  closure and locality 26, 52–5, 74
  $H\,H$ 11
  in Sugawara construction 69–71
  $j^a\,j^b$ 66, 68
  $\tilde{J}_\alpha\,\tilde{J}_\beta$ 357
  $j_B\,b$ 24
  $j_B\,\beta$ 24
  $\lambda^A\,\lambda^B$ 49
  $\lambda^a\,\lambda^b$ 73
  $N = 2$ superconformal 37, 47, 375
  $\mathcal{O}_m\,\mathcal{O}_n$ 234
  $\psi\,\mathcal{A}_v$ 14
  $\psi_l\,\psi_{l'}$ 252
  $\psi^\mu\,\psi^\nu$ 2, 49
  $\psi^\mu\,\Theta_\alpha$ 112
  superconformal current algebra 339
  $T_B\,\mathcal{O}$ 233
  $T_B\,T_B$ 3–4
  $T_B\,T_F$ 3–4
  $T_F\,\mathcal{A}$ 10
  $T_F\,T_F$ 3–4, 280
  $X^\mu\,X^\nu$ 2, 49
  $\boldsymbol{X}^\mu\,\boldsymbol{X}^\nu$ 106, *134*
  $\mathcal{V}_\alpha\,\mathcal{V}_\beta$ 114
  $W\,W$ 253
orbifolds 274–301, 331ff, 471–2
  asymmetric 279
  fixed points 311
    blowing up 296–7, 309–10, 312–13
  K3 *300–1*, 417, 425, 427
  modular invariance 276–9
  non-Abelian 275
  nonsingularity of CFT 275, 410, 425
  T-duality 295–6
  twisted sector 286–8, 296–8
  twisting construction 55–9
  $\mathbf{Z}_3$ 283–91, *300*, 309–10
  $\mathbf{Z}_4$ 291–2, *300*
orientifold 196, 422
orientifold plane 138, 142–4, 147, 151, 175, 206–8

p-brane, *see* black *p*-, D-, M2-, M5-, and NS5-branes
p-form, *see* differential forms
p-form gauge field 85, 451–2
  field strength 86
  self-dual 91, 452
  see also Chern–Simons term, R–R fields
p-p' string 153–5, 162–4
  $\#_{ND} = 4$ 162–4, 197
  $\#_{ND} = 8$ 191
(p,q)-form 307–9
(p,q) string 167, *176–7*, 182
P invariance, *see* parity
parafermion CFT 251–2, 269, 391–2, 471
parallel transport 304, 463
parity
  and anomalies 95–8, 285, 304, 329–30
  and T-duality 137
  spacetime 27–8, 127
  world-sheet 27–31, 41
partition functions 235–6, 266, *272*, 471
  and bosonization 13
  BPS states 171, 190
  GSO projected 32–7, 51–3, 55–8
  with twists 33–4, 156
Pati–Salam unification 348
Pauli–Villars regulator 95, *285*
Peccei–Quinn (PQ) symmetry 333–5, 360–2, 365–6, 371–2, 382–3, 473
physical states 1, 3, 20–3, 28, 45
picture changing operator (PCO) 118–21, *126*, 469
pictures 108–10, 118–21, 147
Planck scale, eleven-dimensional 199, 212
Poincaré duality 86, 450–2

point group 275
primary field, *see* highest weight state, tensor field
projective plane 39, 42
projective space 310–11, 320, 398, 404, 465
proton decay 355–6, 472
pseudospin 67, 244

quantum mechanics 210, 225–6, 429
quaternionic manifold 464

R symmetry 393, 444
rank (algebra) 61, 77, 326, 421, 457–8
rank (form) 10, 85, 450
Ramond algebra 7, *43*
Ramond (R) boundary condition 5–9, 15–16
  vertex operators 13–15, 19
R–R charge
  in supersymmetry algebra 144–5
  of black *p*-brane 219
  of D-brane 141–52, 425
    value 147
  of orientifold plane 142–3
  vanishes for strings 88–9, 92, 381
R–R fields 9–10, 27, 87, 334
  background 106
  coupling of dilaton 87–9
  effect of *T*-duality 137–8
  tadpole 37–42
Ramond–Neveu–Schwarz (RNS) superstring 29
rational CFT 255–9, *272–3*
refermionization 131–3, 156–7
relevant interaction 166, 264–6, 268–71
renormalization 393, 445
renormalization group 259–66, 346, 350, 369
Ricci flatness 308–9
Ricci form 308–9
Riemann–Roch theorem 121–3
Riemann surfaces 13, 122–4
Riemann tensor 96, 113
roots 61–3, 67–8, 425
  lattice 74–5

S-duality (weak–strong) 179–82, 186–7, 190–2, 198–9, 205–7, 422–9
$SL(2,\mathbf{R})$ 91–2, 182
$SL(n,\mathbf{R})$ 61, 187–8
$SL(2,\mathbf{Z})$ 182, 186–8, 200, 295–6, 427–8, 474
$SL(3,\mathbf{Z})$ 200
$SO(4) = SU(2) \times SU(2)$ 439
$[S]O(5,5,\mathbf{Z})$ 200
$SO(8)$ spin 22–3
$SO(10)$ 64–5, *82*, 289, 338, 346
$SU(1,1)$ 296, 428
$SU(5)$ 64–5, *82*, 289, 338, 345–8, 352–6
$SO(n)$ 60, 62
  traces 99
$[S]O(m,n,\mathbf{R})$ 61, 76, 187–8, 417, 427–8
$[S]O(m,n,\mathbf{Z})$ 76, 187–8, 428
$SU(n)$ 60, 63
S-matrix 124
scale transformation 259–63
Scherk–Schwarz mechanism 363–4
short multiplet 170, 449
simple current 240–1, *272*, 354
$\sin^2\theta_w$ 290, 346–54, 472
soliton 173, 182, 203–5, 426
$Sp(k)$ 60, 62
space group 275
special Kähler geometry 319–20, 381–3, 390, 405, 464–5
spectral flow 14, 379, 387
sphere amplitudes 112–18, *134*
  relation to disk 116
spin 22–3
spin connection 203, 303
  embedded in gauge connection 284, 291, 304, 324, 418–19
    and (2,2) supersymmetry 311–12, 376
spin field 15, 19, 131–2
  current algebra 54, 66
  four-dimensional 341, 356
spin–statistics relation 19, 29, 33, 35, 49, 52, 55, *82*, 332
spin structure 122–4, 127, 130–2
spinor representations 430–9
  products 435–7

$SO(N)$ 437–8
under subgroups 438–9
Standard Model 178, 274, 326, 370–3, 421–2
anomalies 94, 98
CP symmetry 331–5
chirality 98, 282, 288–9, 329–31
gauge couplings 298, 314, 343–52
gauge group 64–5, 81, 289–90
gauge quantum numbers 337–8, 340–3, 352–5
generations 64–5, 288–92, 313–15, 324–6, 387–8, 471–2
hierarchy problem 281–2, 350, 369
proton stability 355–6
supersymmetric 346–8
Yukawa couplings 314, 318
state–operator isomorphism, *see* vertex operators
statistical mechanics 219, 224–5, 260, 266–71
Stokes's theorem 450
string coupling, *see* coupling constants
string field theory 210, 259
string metric 205
string scale 343–50
string–string duality 425, *429*
string tension 150, 164–7, 180
strong coupling limit
heterotic string 190–4, 205–8, 421–9
type I superstring 190–4
type IIA superstring 198–9
type IIB superstring 179–81
strong CP problem 333–5, 472
structure constants 59
Sugawara construction 69–72, *82*, 243, 247, 250, 339, 468
superconformal algebra 3–4, 7, *272*
as constraint 20, 45
general N 45–9
$N = 2$ 37, 47–8, *82*, 375–6
primary field 105
superconformal current algebra (SCCA) 338–42
superconformal field theory (SCFT) 467–8
(1,1) 2–5, 254–5
(2,2) 292, 311–12, 380–415, 473
Landau–Ginzburg models 392–4
minimal models 390–2
(4,4) 417–18
superconformal Killing vector 123
superconformal transformation 2, 10, *43*, 103–4
supercurrent 2–5, 50
superderivative 103–4
superfield 104
supergravity 468
$d = 4$, $N = 1$ 293–5, 315, 359–62
eleven-dimensional 84–7, 198–9, 201, 453
long-distance 160, *176*, 213–14
toroidally compactified 187–9, 454–7
type I 52–3, 92–3, 98–101
same as heterotic 93, 192
type IIA 28–9, 87–90, 96
type IIB 28–9, 90–2, 98
supermanifold, *see* superspace
supermembrane, *see* M2-brane
supermoduli space 121–6
boundary of 102–3
superpartner scale 343, 350, 369
superpotential 443–7
Calabi–Yau 315–20, 325, 390
constraints from holomorphicity 322–3, 359–62, 367–8
Landau–Ginzburg/linear sigma model 392, 398, 400
nonperturbative contributions 323–4, 371–2
orbifold 294–6
super-Riemann surface 121–4
superspace 103–8, *134*
actions 105–8, 469
superstring, *see* types 0, I, IIA, IIB
supersymmetry (spacetime) 471
4 supersymmetries 440–7
8 supersymmetries 461–5
16 supersymmetries 457–61
32 supersymmetries 452–7
and GSO projection 29–30, 50–1, 359

and hierarchy problem 281–2
and $\sin^2\theta_w$ 346–8, 473
and vanishing of vacuum amplitude 35, 38, 146
as means of studying strong coupling 178–227
conditions for unbroken
D-brane systems 152–5, *175–7*
general CFT 356–9, *374*, 473
orbifolds 282–3, *300*
smooth manifolds 302–5
$d = 2$ 449
$d = 4$, $N = 1$ 440–7
$d = 4$, $N > 1$ 448–9
in specific systems
D0–Dp system 168
F-string–D-string system 164
heterotic string 79–82
type I/II superstring 144
in statistical mechanics 269
transformations 173, 203, 302, 443
vs world-sheet 29
see also BPS bound, BPS states, nonrenormalization
supersymmetry breaking 473
at tree level 362–4
by D-branes 139–40, 152–5, 168, *175–7*
dilaton-mediated 370
in loop expansion 364–6
nonperturbative 366–73
scale 369
symmetries, continuous 327–8

T-duality 452, 469
and mirror symmetry 402–4, 473
and U-duality 187–8
as a gauge symmetry 328
combined with S-duality 199, 205–7, *227*, 422–7
effect on D-branes 138–45, 163–4ff, *176*
effect on R–R fields 137, 149
heterotic theories 76–8
on orbifolds 295–6, 298, 300, 472
type I theory 138–43
type II theories 136–8
't Hooft–Polyakov monopole 82, 187
tachyon 21–3, *82*, 266, 341
in nonsupersymmetric D-brane systems 158, 169
removal by GSO projection 27, 31, 52, 54, 58–9
theories with 27, 55–8
tadpoles 37–42, 142–3
temperature *83*, 219, 266–9
tensionless string theory 205, 260–1, 421, 474
tensor multiplet
$d = 6$ (1,0) 418, 420, 461–2
$d = 6$ (2,0) 204, 417, 459
tensor superfield 104
tetrad 77, 96
theta functions 33, 156, 469
Thirring model 36
threshold correction 298–300, 334, 349, 370–1, 472
topological string theory 49, 385
topology change (spacetime)
conifold 412–15, 474
flop 406–9, 474
toric geometry 404
toroidal compactification
heterotic string 76–82, 422–9, 468, 474
type I string 143–4
type II string 187–90, 454–7
torsion 106, 203, 303, 324
torus amplitudes 31–5, 51–9, 126–33, *374*
transition functions 122, 125, 306
tree-level amplitudes 110–18, *135*
trinification 349
twist (orbifold) construction 55–8, 251, 275–6, 279
twisted boundary conditions 13–4, 32–4, *44*, 155–6ff
twisted sector 276–9, 286–8
type 0 superstring 27, 35
type I superstring 31, 37–42
amplitudes 110–16, *135*
strong coupling limit 190–7

T-duality 138, 143–4
type II superstrings 26–9, 31–7
  amplitudes 113–18, *134*
  on Calabi–Yau manifolds 379–86ff
  on K3 415–18
  strong coupling limits 179–87, 198–205
  T-duality between 136–8, 143–4
  toroidally compactified 187–90, 454–7

U-duality 187–90, 199–201, *226*, 328
$U(n)$ 30, 63, 151, 186–7
UV divergence 103
ultrashort multiplet 80, 169–71, 210, 449
unitary CFT 45–6, 228–32, 254–5, 262–3, 391, 471

vacuum amplitudes 31–42, 146–7, 155–8
vacuum selection 373
vector multiplets
  $d = 4$, $N = 1$ 441
  $d = 4$, $N = 2$ 380–4ff, 447, 463–5
  $d = 4$, $N = 4$ 79, 448
  $d = 6$ (1,0) 163, 172, 461
  $d = 6$ (1,1) 204, 459
Verlinde formula 258, *273*
Verma module 232, 236
vertex operators 10–15, 108–10, 468, 471
  and chiral primaries 384–9
  ghosts 15–20
  locality 25–6, 50–1, 256–7
  gravitino 28
  p-p' strings 174, 197
  pictures 108, 118–24
  Ramond, branch cuts 6, 10
  R–R 30, 88
Virasoro algebra 228–33ff
Virasoro master equation 258

W algebra 46, 471
W string 46, 468
WZNW model 246–50
weight (conformal) *43*, 263
  current algebra primary 72
  unitary CFTs 231, 255, 378, 391
weight (Lie algebra) 61–3
  lattice 75
weight (modular) 297
Wess–Zumino term 248
Weyl spinor 8, 432
Wilson criterion (computability) 209–11, 214
Wilson line 76–8, *83*, 206, 208, 227, *429*
  gauge symmetry breaking 289–91, 314, 325–6
Wilsonian action 300, 361
winding string 77–81, 189–90

Yang–Mills theory, $d = 4$, $N = 4$ 186–7, 428, 460–1

Zamolodchikov metric 362, 374, 389
zero modes 119
  instanton 324–5
  sphere 108
  torus 121, 123, 127, 130, 132–3
zero-norm states, *see* null states
zero-point energy 7, 14, 34