STRINGS AND GEOMETRY

Clay Mathematics Proceedings
Volume 3

STRINGS AND GEOMETRY

Proceedings of the
Clay Mathematics Institute
2002 Summer School
on Strings and Geometry

Isaac Newton Institute
Cambridge, United Kingdom

March 24–April 20, 2002

Michael Douglas
Jerome Gauntlett
Mark Gross
Editors

American Mathematical Society
Clay Mathematics Institute

2000 *Mathematics Subject Classification*. Primary 81T30, 83E30, 53C29, 32Q25, 14J32, 83E50, 14E15, 53C80, 32S45, 14D20.

ISBN 0-8218-3715-X

Published by the American Mathematical Society, Providence, RI,
for the Clay Mathematics Institute, Cambridge, MA.
Printed in the United States of America.

∞ The paper used in this book is acid-free and falls within the guidelines
established to ensure permanence and durability.
Visit the AMS home page at http://www.ams.org/
Visit the Clay Mathematics Institute home page at http://www.claymath.org/

10 9 8 7 6 5 4 3 2 1 09 08 07 06 05 04

Contents

Preface

The 2002 Clay School on Geometry and String Theory was held at the Isaac Newton Institute for Mathematical Sciences, Cambridge, UK from 24 March - 20 April 2002. It was organized jointly by the organizers of two concurrent workshops at the Newton Institute, one on Higher Dimensional Complex Geometry organized by Alessio Corti, Mark Gross and Miles Reid, and the other on M-theory organized by Robbert Dijkgraaf, Michael Douglas, Jerome Gauntlett and Chris Hull, in collaboration with Arthur Jaffe, then president of the Clay Mathematics Institute.

This volume is one of two books which will provide the scientific record of the school, and focuses on the topics of manifolds of special holonomy and supergravity. Articles in algebraic geometry, Dirichlet branes and related topics are also included. It begins with an article by Michael Douglas that provides an overview of the geometry arising in string theory and sets the subsequent articles in context. A second book, in the form of a monograph to appear later, will more systematically cover mirror symmetry from the homological and SYZ points of view, derived categories, Dirichlet branes, topological string theory, and the McKay correspondence.

On behalf of the Organizing Committee, we thank the directors of the Isaac Newton Institute, H. Keith Moffatt and John Kingman, for their firm support. We thank the Isaac Newton Institute staff, Wendy Abbott, Mustapha Amrani, Tracey Andrew, Caroline Fallon, Jackie Gleeson, Louise Grainger, Robert Hunt, Rebecca Speechley and Christine West, for their superlative job in bringing such a large project to fruition, and providing the best possible environment for the school. We thank the dining hall staff at Kings College, Magdelene College, Corpus Christi College and Emmanuel College, and especially the Kings College singers, for some memorable evenings. We thank the staff of the Clay Mathematics Institute, and especially Barbara Drauschke, for their behind-the-scenes work, which made the school possible. Finally we thank Arthur Greenspoon, Vida Salahi and Steve Worcester for their efforts in helping to produce this volume.

Michael Douglas, Jerome Gauntlett and Mark Gross
September 2003

Clay Mathematics Proceedings
Volume 3, 2004

The Geometry of String Theory

Michael R. Douglas

Abstract. An overview of the geometry of string theory, which sets the various contributions to this proceedings in this context.

1. Introduction

The story of interactions between mathematics and physics is very long and very rich, too much so to summarize in a few pages. But from the beginning, a central aspect of this interaction has been the evolution of the concept of geometry, from the static conceptions of the Greeks, through the 17th century development of descriptions of paths and motions through a fixed space, to Einstein's vision of space-time itself as dynamical, described using Riemannian geometry.

String/M theory, the unified framework subsuming superstring theory and supergravity, is at present by far the best candidate for a unified quantum theory of all matter and interactions, including gravity. One might expect that a worthy successor to Einstein's theory would be based on a fundamentally new concept of geometry. At present, it would be fair to say that this remains a dream, but a very live dream indeed, which is inspiring a remarkably fruitful period of interaction between physicists and mathematicians.

Our school focused on the most recent trends in this area, such as compactification on special holonomy manifolds, and approaches to mirror symmetry related to Dirichlet branes. But before we discuss these, let us say a few words about how these interactions began. To a large extent, this can be traced to before the renaissance of string theory in 1984, back to informal exchanges and schools during the mid-1970's, at which physicists and mathematicians began to realize that they had unexpected common interests.

Although not universally known, one of the most important of these encounters came at a series of seminars at Stony Brook, in which C. N. Yang would invite mathematicians to speak on topics of possible mutual interest. In 1975, Jim Simons gave a lecture series on connections and curvature, and the group quickly realized that this mathematics was the geometric foundation of Yang-Mills theory, and could be used to understand the recently discovered non-Abelian instanton and monopole solutions [**23, 65**]. These foundations are by now so familiar that it is

2000 *Mathematics Subject Classification.* 81T30, 83E30.

a bit surprising to realize that, for Yang-Mills theory, they date back only to this time.

Among the participants in the Simons-Yang seminars was Is Singer, who carried news of these developments to Atiyah at Oxford. Before long the mathematicians were taking the lead in exploiting these solutions, culminating in the early 1980's with Donaldson's use of instanton moduli spaces to formulate his celebrated invariants [**17**], which revolutionized the study of four-dimensional topology.

While this case study in math-physics interaction might have ended there, with the lesson being that mathematicians can find useful inspiration in physical developments but then must apply them to their own problems, of course it did not. The deeper aspects of this interaction began with Witten's 1988 reformulation of the Donaldson invariants as observables in a topological field theory [**58**]. This set the stage for the eventual application of deeper physical arguments, which led to the 1994 Seiberg-Witten solution of $\mathcal{N} = 2$ supersymmetric gauge theory [**50**]. Applying this to the topological field theory formulation led to the dramatic discovery of the Seiberg-Witten invariants [**60**]. While the special role of four dimensions in physics had led many to suspect physics would lead to new insights into special features of four-dimensional space, this success went far beyond what anyone had expected.

An extreme reading of these striking developments would take them as evidence that fundamental theoretical physics and mathematics (or at least some subfields of mathematics) have merged into a single unified field. However, few workers in either field would agree with this claim, and consideration of the developments following the Seiberg-Witten work bears this out: with some noteworthy exceptions such as [**42**], the subsequent mathematical developments [**43**] and physical developments (well discussed in Dorey's lectures in this volume) have had little overlap so far. Of course it would be overly pessimistic to rule out equally dramatic interactions in the future, but the point we want to make is simply that the two fields are not the same: they have different goals, different questions are raised, and even the sense of when a question has been answered or an area understood is rather different in the two fields. Thus the basic questions of why such a sustained interaction should be possible at all, and whether we should expect this trend to continue indefinitely or not, deserve serious consideration.

While we will not address these questions in depth here, the most basic point to keep in mind is that a sustained interaction is only possible in an area which is of fundamental interest on both sides, and in practice this interest is going to be for very different reasons on the two sides. This was certainly the case for the study of gauge theory, and it is equally true for string theory. Just because so many of the recent physical developments start with interesting mathematics, say the existence of manifolds of G_2 holonomy, or the classification of Calabi-Yau threefolds, and then suggest intriguing mathematical conjectures, does not in itself make these interactions deep or significant. Rather, they have true significance only to the extent that the ideas which cross over from one side to the other turn out to be important in addressing the fundamental questions on the "other" side.

We will try to say a few words about this deeper significance in our conclusions, but for now the main point we take from these comments is that one needs to keep the fundamental questions from math and from physics clearly in mind in evaluating the likely progress and significance of any given point of interaction.

Thus I start in section 2 with a very brief overview for mathematicians of where string theorists stand at present, and what for them are the fundamental physical questions. While this will necessarily be a sketch with a lot of undefined terms, besides serving as cultural background, it will outline the long road we have taken from our experimental foundations to our present situation, in which we believe interaction with mathematicians will provide us with essential insights. I will not presume to do the same for the mathematical questions, instead referring to [**47**].

In the remainder of the article, we introduce the various frameworks used to study string theory. Of course a basic problem in math-physics interaction is that there is at present no precise definition of string theory or even most quantum field theories, even to the satisfaction of physicists. On the other hand, the problem is not as bad as one might think, in that these theories are highly overconstrained – it is not hard to list axioms which almost all physicists would agree should characterize them uniquely. Rather, the problem with treating the theories mathematically is that along with these axioms, a very large number of supplementary assumptions are used in physical arguments; one wants to reduce these down to a manageable list which deserve the name "axioms." Having even a redundant set of axioms for a particular QFT or string theory, which sufficed to make interesting physical arguments, would be a valuable advance, even before we reach the (probably still distant) goal of proving the existence of the theories they characterize.

Most math-physics interaction is based on theories for which this problem is not too serious, and we concentrate on these. In particular, classical field theory, in which the basic mathematical framework is simply that of partial differential equations, is far more generally applicable to these problems than one might think. The only essential generalizations beyond what was needed to formulate general relativity and Yang-Mills theory are fairly trivial mathematically, for example to allow anticommuting variables.

While at first one might think classical field theory would be too simple to address any of the questions of current interest, this is not true at all. Indeed, the formulation of the Standard Model, which summarizes most existing observational and experimental data, and most of the proposals which try to go beyond it, is made using "effective field theory," which summarizes quantum and string theoretic results in classical terms.

The concept of effective field theory is also central in the connections to mathematics – in particular, the Seiberg-Witten solution is just the effective field theory related to the original Donaldson theory. Thus we give an introduction to this idea in section 3, emphasizing the sense in which effective field theory can be thought of as a generalization of "moduli space." As such, concepts developed in algebraic geometry, such as formal deformation spaces and stacks, have simple effective field theory analogs, and this is an important dictionary to learn.

In section 4, we discuss the physics of higher-dimensional field theories, Kaluza-Klein reduction, and supersymmetry. For mathematician-friendly introductions to supersymmetry, we recommend [**24, 13**]. Again, these are classical field theories, but there is a precise sense in which a limit (the low energy limit) of string theory is described by the ten- and eleven-dimensional supergravity theories, and many of the recent developments involving duality were first discovered by working on this level.

In section 5, we discuss results from the study of perturbative string theory, the limit of string theory which can be defined in terms of a string world-sheet governed by two-dimensional quantum field theory. This was the traditional definition of string theory, and is still the basis of most of our understanding of string theory beyond the low energy limit. As by far the most concrete formulation which goes beyond conventional ideas of geometry, it has been a major focus of interaction with algebraic geometers, leading to discoveries such as mirror symmetry, and its more recent open string/Dirichlet brane analogs.

In section 6 we briefly discuss the more recent developments of superstring duality and M theory. Most of the physics lectures in the volume address these topics. While physicists have had a fairly clear picture of what M theory should be for several years, it is still rather challenging to summarize this in a way which mathematicians will find accessible. See [**64**] for another attempt at this.

In section 7 we conclude.

2. String theory and its physical goals

The stated goal of "fundamental physics" is to provide a single set of precise laws which governs all observed physical phenomena. While ambitious, from many points of view this goal has already been realized: it is difficult to find any hard evidence for phenomena which are not very well described by the combination of the "Standard Model," a specific four-dimensional quantum field theory based on Yang-Mills theory, and general relativity.

On the other hand, it is clear on theoretical grounds that this combination of theories is incomplete. Most importantly, quantum effects in general relativity predict the breakdown of the theory at a very high energy scale called the "Planck scale," $M_{pl} \sim 10^{19}$ GeV $\equiv 10^{28}$ eV (electron volts). For comparison purposes, the mass of the electron is about 5×10^5 eV.

To illustrate this breakdown, consider an experiment in which we collide an electron and positron each moving with energy E. While the Standard Model by itself makes definite predictions for any E, since the gravitational interaction has strength roughly E^2/M_{pl}^2, as $E \sim M_{pl}$ this interaction is no longer a small correction, and must be taken into account on an equal footing with the other forces. Providing any candidate set of laws which does this remained an open problem for many years. Over the course of the 1980's, fairly convincing arguments were developed to show that superstring theory could solve this problem.

Besides this basic problem, there are many features of the Standard Model which strongly suggest that there are deeper layers of structure to be found. The two best examples are known as "coupling unification" and the "hierarchy problem." To explain these, we start with an exceedingly brief overview of gauge symmetry as it relates to the Standard Model.

2.1. The Standard Model and gauge symmetry. There are three fundamental forces of nature described by gauge theory, namely electromagnetism, the weak interaction, and the strong interaction, These look very different in low energy experiments: the electric and magnetic interactions are long range and relatively weak; the weak interaction operates over nuclear distances and operates between "currents" much like the electromagnetic current, while the strong interaction, while also operating only over nuclear distances, is exceedingly complicated and on the surface bears little resemblance to Maxwell's theory.

On the other hand, at the highest energies we attain in accelerator experiments today, around 100 GeV, the three gauge forces look very similar, and one can directly observe particles which form representations of the gauge groups $U(1)$, $SU(2)$ and $SU(3)$. Their very different behaviors at low energy are explained through the existence of three phases of gauge theory, somewhat analogous to the phases solid, liquid and gas of ordinary matter. One possibility, the "Coulomb phase", is the familiar behavior of electromagnetism as described by Maxwell's equations. The second is the "spontaneously broken" or "Higgs phase," in which the force is short range and weak; the weak interaction of the Standard Model arises through breaking of $SU(2) \times U(1)$ gauge symmetry. Finally, the strong force is in the "confined" phase, in which the gauge interactions between charged objects (quarks) are not directly observable; indeed, the most basic physical effect is the formation of a non-trivial vacuum "condensate" of quarks, so that the lightest strongly interacting particles (the pions) are "spin waves" of this condensate.

While we will not get into the physical details of this, it is valuable to have some mathematical analogies or counterparts for these effects, because they are central in physicists' thinking, and the actual study of supersymmetry and string theory relies heavily on these ideas. Let us do this in the following framework: we grant that the additional structure of string theory, M theory or whatever candidate fundamental theory lies beneath the Standard Model can be described by a family of "configurations". Traditionally, in string theory a configuration is a choice of structure for the compact manifold of "extra dimensions of space-time," and might include auxiliary structures such as vector bundles on this manifold, submanifolds (branes), and so on. Eventually, we may want to describe string theory using something other than differential geometry, and we might then use some more abstract family of configurations. In any case, a possible physical history will be a map from four-dimensional space-time into the family of configurations.

In this language, the gauge symmetry we observe arises because our vacuum is a configuration with an isometry or endomorphism group. Since this group acts at each point in observable space-time, it will lead to gauge symmetry, almost by definition. Thus, physical experiment tells us that whatever may be the additional structure which leads to the Standard Model, it must be preserved by an $SU(3) \times SU(2) \times U(1)$ symmetry in some sense.

What about the three phases? The broken or Higgs phase is relatively easy to explain in these terms: it corresponds to an "approximate symmetry." More specifically, it arises in the situation in which a family of configurations (say connections) is naturally defined by a quotient construction. Reducible configurations then have additional endomorphisms (by definition) and thus lead to enhanced gauge symmetry. A configuration which is close to a reducible configuration (in terms of some metric), will have broken gauge symmetry, at a scale proportional to the distance to the reducibility locus. Thus, while our expectation of an $SU(3) \times SU(2) \times U(1)$ symmetry for the additional structure is correct, it must be that the effective potential (to be explained below) has a minimum which is $SU(3)$ symmetric, and near but not precisely at the $SU(2) \times U(1)$ symmetric point.

The confined phase would seem the most difficult to explain in these terms. In one way, it is not: from the point of view of families of configurations, if we grant that the group G is confined, the consequence is that at low energies we must identify all configurations related by G symmetry; in other words we can answer all physical questions after performing the G quotient. Of course this begs the question of whether or not a given symmetry G which appears at a particular reducibility locus is confined; this is a central question we will return to below.

2.2. Coupling unification. At 100 GeV, the strengths of each of the gauge interactions can be parameterized by a dimensionless "coupling constant" $\alpha_1 \sim 1/60$, $\alpha_2 \sim 1/30$ and $\alpha_3 \sim 1/10$. Now, due to renormalization effects, these coupling constants depend on the energy at which they are measured. Going down to everyday energies, a combination of α_1 and α_2 becomes the familiar "fine structure constant" $\sim \frac{1}{137}$. On the other hand, if one extrapolates to high energies one finds that all three become equal to $\alpha_{GUT} \sim 1/25$ at a single energy scale $M_{GUT} \sim 10^{16}$ GeV called the "GUT scale," where GUT stands for "grand unified theory."

Needless to say, while getting two functions (coupling as a function of energy) to agree at some value would not be too impressive, the fact that three agree at the same energy looks like clear evidence that something important happens at that energy. The simplest hypothesis is that all three interactions arise from the spontaneous breakdown of a larger gauge symmetry, say the group $SU(5)$ which contains $SU(3)\times SU(2)\times U(1)$ as a (maximal) subgroup. This hypothesis of "grand unification" implies a great deal of additional structure, and leads to very concrete predictions, for example that the proton has a finite (although extremely long) lifetime.

This picture was suggested in 1974 [**27**] and subsequently many attempts were made to find a simple GUT from which all the observed structure of the Standard Model would naturally arise. Although some progress was made, it is now generally believed that there is no GUT which achieves this in a particularly simple and natural way; the GUTs are almost as complicated as the Standard Model they purport to explain. Rather, it appears that some other level of structure is required to explain the origin of most of the features of the Standard Model, such as the three generations of quarks and leptons.

2.3. The hierarchy problem. The Standard Model contains an explicit parameter setting the scale of electroweak $SU(2)\times U(1)$ symmetry breaking, the mass squared of the Higgs boson. While not yet observed, if the Standard Model is right, this must be between 110 GeV (the current experimental lower bound) and about 500 GeV. Either of these figures is far below the other two scales we mentioned, the Planck scale and the GUT scale, and this already might lead one to deep scepticism about any attempt to provide a complete fundamental theory – how can we hope to extrapolate more than 13 orders of magnitude in energy, without experimental input? On a more positive note, one can remark that no previously known physical phenomenon relates such disparate scales, so if it is true here, there must be some striking new physics which makes this possible, which we can hope to discover.

Of the various proposals in this direction, by far the most widely believed is that this new physics is low energy supersymmetry, meaning not only that each particle observed so far has a counterpart of the opposite statistics, but that these "superpartners" must have mass around the same $100 - 500$ GeV energy range. Besides theoretical arguments, there is a powerful supporting piece of evidence in the fact that the coupling unification we just mentioned has now been tested to very high precision, and actually fails to work if one has only the Standard Model particle content. On the other hand, it works extremely well if one assumes low energy supersymmetry.

This is very good news for physicists as it would mean that the superpartners should be observed in upcoming experiments at the Large Hadron Collider now being constructed at CERN. These experiments should come into full swing around the year 2008, giving physicists some time to try to sharpen our predictions in preparation.

If we grant this conventional wisdom, the problem facing string theorists is fairly clear. We must show that there is some solution of string theory which leads to effectively four-dimensional physics at the energies of interest, contains low energy supersymmetry, and agrees in detail with the tested predictions of the Standard Model. Of course, we need not grant the conventional wisdom, and other scenarios have been suggested, for example that some of the extra dimensions of string theory will be observed in accelerator experiments. Other experimental surprises might completely change this picture as well. Of course, any attempt to make physical predictions is by definition speculative, and this is only a question of degree.

Leaving aside physical arguments as to where to place our bets, what is more important for the present discussion is that the "conventional low energy supersymmetry" scenario is the only one which has been developed physically to the point where any of the important questions can be made mathematically precise, and thus we continue our discussion in this framework.

3. Effective field theory

Once we establish the existence of solutions of string theory with low energy supersymmetry, we will need to go on and derive their physical predictions. In particular, we need a framework which is general enough to incorporate "stringy" and quantum corrections, which will be essential for understanding effects such as the breaking of supersymmetry at low energy.

In practice, the framework which is used is "effective field theory." The physical definition of an effective theory is as follows: one chooses an energy scale E, and asks for the simplest theory which can reproduce all physical phenomena at energies up to E, without knowing anything about the "degrees of freedom" present at energies above E. The simplest example is quantum electrodynamics, which describes the physics of electrons and photons. This theory is not complete, since of course there are many other particles which can be created by the interaction of electrons and photons. On the other hand, if one only does experiments at energies less than 105 MeV (the mass of the second lightest charged particle, the muon), the other particles cannot be created directly. They still have "virtual" effects, but these can be summarized by modifications (renormalizations) of the parameters entering the Lagrangian, including those controlling "higher-dimensional operators."

Thus, one can describe all physics at low energies solely in terms of the particles we see, at the cost of needing to determine some additional parameters. The process of finding the contributions of the more massive particles (equivalently, fields) to these parameters is called "integrating out." One can show that these effects are controlled by powers of E/M, where M is the lightest mass of a particle not explicitly present in the theory, so one only needs a finite (and small) number of parameters to make real predictions.

This paradigm would appear quite appropriate for discussing the relation between string theory and the real world, at least until that happy day when we can directly produce strings or higher-dimensional excitations, not described by conventional field theory, in our experiments. For various reasons, at present most string theorists do not consider this outcome very likely at LHC or any accelerator we can imagine constructing, though in the end of course this is a question for experiment to decide. We should also say that for some questions, especially those involving gravitational effects, the applicability of effective field theory is controversial; see for example [**4**] for arguments making this point.

In any case, the Standard Model is universally thought of as an effective field theory, and thus it is a central problem to derive this particular effective field theory from string theory.

For our purposes, an effective field theory is defined by the data $(\mathcal{C}, V, \mathcal{F}, G)$:

- $\mathcal{C}$ is a manifold with metric called the "configuration space." This is not space-time; rather it is the space in which scalar fields take values. Coordinates on this space are traditionally called "scalar fields" and are often denoted ϕ^i.
- V is a real function on $\mathcal{C}$, called the potential.
- $\mathcal{F}$ is a complex vector bundle over $\mathcal{C}$, in which the fermions take values. This also comes with a metric.
- G is a Lie group with an action on $\mathcal{C}$, which preserves the other data.

Traditionally, all of this data is finite dimensional, and G is compact and semisimple. It can be useful to consider infinite dimensional manifolds and groups as well.

The data would be used to define the effective field theory as follows (see [**24**] for details). One takes as degrees of freedom a map $M \to \mathcal{C}$ (the scalar fields), a compatible section of $\Gamma(M, \mathcal{F} \otimes Spin(M))$ (the fermionic fields), etc. One then uses the data to write an action for these fields. This requires postulating additional data as well, most importantly the "Yukawa couplings," a bilinear form $Y : \mathcal{F} \otimes \mathcal{F}|_\phi \to C$ at each $\phi \in \mathcal{C}$, but this will not appear in our discussion.

Finally, for present purposes, one treats this action **classically**; in other words a possible dynamics for the system is given by a gauge equivalence class of solutions of the classical equations of motion. Thus, a "configuration" as in section 2 is a G-orbit of the form $g(p, \xi) \in (\mathcal{C}, \mathcal{F}|_p)$ with $g \in G$.

The only solution we will discuss is the vacuum solution, in which the scalar fields are constant,

$$\frac{\partial \phi^i(x)}{\partial x} = 0; \qquad \frac{\partial V}{\partial \phi^i} = 0,$$

and the fermions (the section of $\mathcal{F}$) are zero. We furthermore identify solutions under gauge equivalence; in other words a solution is a G-orbit of critical points. Such a solution preserves the maximal possible symmetry of four-dimensional space-time and contains no particles anywhere in space; thus it is referred to as a "vacuum." For a stable vacuum, the solution should be stable to small perturbations. This will be true if the critical point is a local minimum.

We say "a" and not "the" vacuum for the evident reason that a general function V can have many minima. *A priori*, any one of the minima is a candidate for describing physics. All further physical predictions depend on the choice of minimum ϕ: for example, the spectrum of scalar particle masses is the matrix of second derivatives $V''(\phi)$ evaluated in a local orthonormal basis at ϕ.

Our previous comments about gauge symmetry and its breaking can be summarized in this language as follows: the unbroken gauge symmetry of a vacuum ϕ is its stabilizer $H \subset G$, while an orbit at distance d from such a stabilizer has H symmetry broken at energy $E = d$. Finally, whereas in general one can identify G and distinguish points on a G-orbit by performing experiments at sufficiently high energy, if an unbroken gauge group H confines at a point ϕ_H, then there is a "confinement scale" Λ such that two points on the same H-orbit at distance $d < \Lambda$ from ϕ_H cannot be distinguished physically.

So far, we have not really distinguished classical field theory and effective field theory. As we suggested earlier, the main physical point of the latter is that we can incorporate all quantum mechanical effects, as well as all effects due to strings or other objects which are not explicitly descibed by the effective field theory, in the choice of the data of the effective Lagrangian. In particular, if we start with a particular classical Lagrangian and quantize it, the resulting effective field theory will obtain modifications or "renormalizations" of V and the metric on $\mathcal{C}$.

In general, even data such as the choice of G or the topology of $\mathcal{C}$ can be different in the quantum theory from that one might have guessed at from a naive interpretation of the classical Lagrangian, for example because confinement can make the identification of G ambiguous, as we just discussed. While such possibilities can be controlled using perturbative techniques at small $\hbar$, one needs more subtle arguments to treat large $\hbar$; much of the lectures of Acharya, Dorey and Gauntlett are explicitly or implicitly devoted to this problem.

We can now summarize our problem as string theorists: it is to find the effective field theories which can arise from string theory, with quantum corrections taken into account, find their vacua, and find out which of these can agree with the Standard Model.

3.1. The Standard Model and unification. In the Standard Model, $G \cong SU(3) \times SU(2) \times U(1)$. G acts linearly on $\mathcal{C}$ and $\mathcal{F}$, so the action is described by a choice of linear representation. The irreps are usually denoted $(N_3, N_2)_Y$, where N_3 and N_2 are dimensions of $SU(3)$ and $SU(2)$ representations, and the notation $\bar{N}_3$ is used to denote the complex conjugate representation. Finally, Y is the weight for the $U(1)$ action.

First, we specify $\mathcal{C} \cong (1,2)_{1/2}$, while V is an invariant quartic polynomial, for which $\phi = 0$ is a local maximum. The metric can be taken to be flat.

We then have

$$\mathcal{F} \cong 3[Q\ (3,2)_{1/6} \oplus \bar{D}\ (\bar{3},1)_{1/3} \oplus \bar{U}\ (\bar{3},1)_{-2/3} \oplus L\ (1,2)_{-1/2} \oplus \bar{E}\ (1,1)_1], \tag{3.1}$$

where we label the quark" fields Q, $\bar{U}$ and $\bar{D}$, and the lepton fields L and $\bar{E}$. In other words, we have 3 copies of a specific 15-dimensional reducible representation (a "generation" of quarks and leptons).

One must also specify (up to change of basis) three matrices $m_1, m_2, m_3 \in \mathrm{Mat}_3(\mathbb{C})$, the Yukawa couplings. Together with the three gauge couplings, these form 19 real parameters which must be determined by experiment.[1]

At first sight, a mathematician might say that while $\mathcal{C}$ and V seem reasonably simple, $\mathcal{F}$ is a rather ugly and random looking collection of data. Needless to say, physicists have had a lot of time to consider the matter, and to some extent have

[1] The discovery of neutrino oscillations is the first solid evidence for physics beyond the Standard Model, but can be accommodated with a few more parameters and fields.

come to the opposite point of view: the G-representation which appears in (3.1) can easily come out of a simpler starting point, but the presence of $\mathcal{C}$ and V is quite difficult to understand.

Perhaps the simplest hypothesis which can lead to (3.1) is to postulate that $G = SO(10)$, $\mathcal{F}$ is 3 copies of an irreducible spinor representation, and $\mathcal{C}$ contains a copy of the adjoint representation of G. It is then easy to find a ϕ whose stabilizer is the Standard Model gauge group, and it is a simple exercise to check that the 16 dimensional spinor of $SO(10)$ decomposes into the spectrum (3.1) plus an additional trivial representation. Thus this is a "grand unified theory." Other nice unification groups, for which simple choices for $\mathcal{F}$ can decompose into (3.1), include $SU(5)$ and E_6.

Although this striking observation may look like direct evidence for grand unification, one can find other explanations for this matter content. For example, a consistency condition called anomaly cancellation strongly constrains the allowed matter representations in gauge theory, and determines the $U(1)$ weight assignments in (3.1). Still, this observation combined with coupling unification is very suggestive.

Points which remain unexplained are the number 3 of generations and the structure of the Yukawa couplings. Most significantly, as mentioned already, one also has the hierarchy problem: the potential V must have a minimum of order 100 GeV, while other couplings are of order 10^{16} GeV. This is not easy to arrange, and the only simple way to do it involves supersymmetry.

3.2. Supersymmetric effective field theory. An $\mathcal{N} = 1$ supersymmetric effective field theory can be defined by the data $(\mathcal{C}, G, \mu, W)$.

- $\mathcal{C}$ is a complex manifold with Kähler metric.
- G is a Lie group with a holomorphic action by isometries on $\mathcal{C}$.
- μ are moment maps for the G action.
- W is a holomorphic G-invariant function on $\mathcal{C}$, called the superpotential.

Note that there is less data than in the non-supersymmetric case, and that it obeys the far more constraining conditions of complex geometry.

This data determines a specific effective theory which can also be described in the general terms we gave earlier, but with very specific relations among the particles and the couplings. In particular, each particle has a "superpartner" – quarks come with scalar "squarks", and gauge bosons come with partner "gauginos." While none of these have been observed yet, there is a precise hypothesis which accounts for this: supersymmetry is spontaneously broken.

What this means for present purposes is the following. Given the data, the potential of the effective theory is

$$V = (\partial W, \partial W) + |\mu|^2, \tag{3.2}$$

where ∂ is the holomorphic gradient, $(,)$ is the Hermitian inner product derived from the Kähler metric, and μ are the moment maps (the metric which appears here is derived from the gauge couplings).

Such a potential has two types of critical points. One type, in which

$$0 = \partial W = \mu, \tag{3.3}$$

is a "supersymmetric vacuum." One can also have critical points $V' = 0$ at which (3.3) does not hold, called "non-supersymmetric" vacua.

In a supersymmetric vacuum, the naive consequences of supersymmetry hold – for example, superpartners must have the same mass. This is clearly not true in our universe, so if supersymmetry is relevant for real physics, it must be spontaneously broken. In other words, we must be in a non-supersymmetric vacuum. Expanding about such a vacuum, one can compute masses of all particles; now superpartners have different masses, and it turns out to be natural for the presently observed particles to be lighter than the others.

Thus, spontaneously broken supersymmetry is not easy to detect experimentally, until one reaches the energy scales set by $|\partial W|$ and $|\mu|$, which sets the mass of the superpartners. On the other hand, whether or not the supersymmetry is spontaneously broken, supersymmetric theories enjoy much better renormalization properties, in particular the superpotential is not renormalized to all orders in $\hbar$. This is an essential ingredient in most solutions to the hierarchy problem, as it allows terms of very different order to coexist. It is not the entire story, as one must somehow generate these terms in the first place. The field of "dynamical supersymmetry breaking" began with the realization that W could obtain corrections from instanton effects, which are exponentially small in $\hbar$. One can find theories in which supersymmetric vacua which would have been present without these effects become nonsupersymmetric vacua in the exact theory, leading to a solution to the hierarchy problem.

A lot of work has been done on supersymmetric extensions of the Standard Model. Indeed, the simplest such extensions have the same $G = SU(3) \times SU(2) \times U(1)$, while $\mathcal{C}$ is now the direct sum of $\mathcal{F}$ from (3.1), with the $\mathcal{C}$ as in the Standard Model, with a "second Higgs" $(1,2)_{-1/2}$ (necessary because W is holomorphic, so the two representations are different). The Yukawa couplings of the Standard Model, are then derived from cubic terms in the superpotential,

$$W = \bar{U}\ m_1\ QH_1 + \bar{D}\ m_2\ QH_2 + \bar{E}\ m_3\ LH_2, \tag{3.4}$$

where H_i are the two Higgs fields, and m_1, m_2 and m_3 are 3×3 matrices encoding the specific Yukawa couplings.

Thus, the Standard Model fits relatively easily into a supersymmetric extension, and its grand unified extensions do as well. This is somewhat non-trivial as one can easily write effective field theories which do not. Checking these constraints from supersymmetry largely depends on discovering the Higgs boson, but it is also generally believed that supersymmetry forces the Higgs bosons to be relatively light, so this should also happen before long. Either way, the status of low energy supersymmetry should become much clearer in a few years.

Having gotten this far, we can briefly mention the next important problem in constructing a supersymmetric GUT, the problem of "fast proton decay," and the related "doublet-triplet splitting" problem. First, note the non-generic form of (3.4), in which no terms appear which are odd in quarks or leptons separately. Such terms would describe processes in which quarks decay (or transmute) into leptons, but the observation that the proton lifetime is greater than 10^{33} years forces the coefficient of any such term to be vanishingly small. The problem is then that this is not easy to get from a GUT, which in the first instance treats quarks and leptons on the same footing. It is not impossible, but we refer to [**63**] and references there for further discussion of how this can be done.

3.3. Summary. We gave a brief sketch of the Standard Model, and a few of the most promising ideas for what physicists believe lies beyond it, grand unification and supersymmetry. All of these ideas can be described in the general framework of four-dimensional effective field theory.

Attempts to derive these effective field theories from superstring theory and understand their physics have been a major focus of the field for almost twenty years. It is this line of work which has led to most of the mathematical "spin-offs," and explaining this will take up most of the rest of the lecture.

4. Supersymmetric Kaluza-Klein compactification

The field of superstring compactification started in 1985 with the work of Candelas, Horowitz, Strominger and Witten on compactification of the heterotic string on Calabi-Yau manifolds, which led to the first quasi-realistic models with low energy supersymmetry [**6**].

While this construction was discovered in the context of string theory, all of the work was actually done using field theory, and in concept directly follows the original picture of Kaluza and Klein, in which Maxwell's equations were derived as a consequence of compactification of five-dimensional general relativity. Thus we review this picture very quickly before proceeding.

The modern version of the Kaluza-Klein idea can be stated succinctly as the idea that all properties of the four-dimensional effective field theory of our universe can be traced to geometric properties of extra dimensions of space. The central problem in "the geometry of string theory," is a definition of "geometry" which is general enough to encompass all work on string and M theory compactification in a reasonably unified way.

4.1. Kaluza-Klein compactification. Consider a solution of general relativity which is close to the product manifold $M \times K$, where M is four-dimensional Minkowski space-time with its flat metric, and K is a D-dimensional compact manifold. The data of $(D+4)$-dimensional general relativity is a $D+4$ metric and its associated Levi-Civita connection; we now try to interpret this as 4-dimensional data.

Besides a 4-dimensional metric, we find two additional "fields." First, there is the D-dimensional metric on K. This is a choice which can be made independently at each point in M, and thus this choice is a configuration in the terms of section 2. The vacua are direct product solutions of the Einstein equation, with a fixed Ricci-flat metric on K. This can be described in effective field theory terms by taking $\mathcal{C} = \mathrm{Met}(K)$, the infinite dimensional space of metrics on K, and V to be the Einstein-Hilbert action functional on these metrics.

Close to the direct product solutions are solutions in which the metric on K is slowly varying as we move in M. These can be described to a good approximation by a scalar field, a map from M into $\mathcal{C}$.

This approximation is typically controlled by the ratio between the scale L of this variation in M to the radius R of K. More precisely, it is controlled by the ratio E/h, where E is the energy and $h \sim 1/R$ is the gap (the lowest non-zero eigenvalue) for the scalar (and other) Laplacians) on K. If we want an effective field theory for energies far below h, the description using an infinite dimensional $\mathcal{C}$ is far more complicated than we need.

Rather, one can take for $\mathcal{C}$ the moduli space of Ricci-flat metrics on K. Since all of these are solutions, one takes $V = 0$. This $\mathcal{C}$ carries a natural metric, the Weil-Petersson metric, inherited from the natural metric on $\mathrm{Met}(K)$.

The remaining data contained in the $(D{+}4)$-dimensional metric can be regarded as a connection on the fibration of K over M. While in general this connection need not be linear, if we restrict attention to solutions which are slowly varying in M we find that the connection must be trivial up to a possible action by an isometry of K. In particular, the original result of Kaluza and Klein was that $K = S^1 \cong U(1)$, so in this case one recovers a $U(1)$ connection. More generally, one obtains a G-connection, where G is the isometry group of K. Furthermore, there is an induced action of G on $\mathcal{C}$; in physics terms one says that the scalar field is "charged" under G.

While in pure $(D+4)$-dimensional general relativity K must be Ricci-flat, by considering other theories this can be generalized. The simplest possibility [**25**] is to consider a $(D+4)$-dimensional theory with an additional 3-form gauge potential (generalizing Einstein-Maxwell theory). As discussed by Acharya and Gauntlett in this volume, one can show that the resulting field equations can be solved by K with constant positive scalar curvature. This greatly enlarges the possible isometry group, at the cost of obtaining M with constant negative curvature (anti-de Sitter space).

To end, we summarize the proposal of Kaluza and Klein as the idea that a configuration is a manifold with metric, up to the natural equivalence by diffeomorphisms. Gauge symmetry arises at fixed points of the quotient by this equivalence.

4.2. String theory. For a long time, the Kaluza-Klein idea was completely ignored, for the simple reason that it requires starting with a higher-dimensional analog of general relativity. Even in four dimensions, it is very hard (most would say impossible) to quantize general relativity by a direct approach, for the reason mentioned earlier that the interaction becomes strong at short distances. This problem gets worse in higher dimensions and, while mitigated in supergravity, is not cured.

In superstring theory, point particles are replaced by strings. These come with a natural length scale, their average size or "string length" l_s, and it turns out that all interactions become weak at distances shorter than l_s, overcoming this problem. On the other hand, at distances larger than l_s, it is a good approximation to regard the string as a convenient summary for a particular spectrum of particles, one for each possible joint state of its modes of vibration. Only a small number of these particles have mass small compared to $1/l_s$, and thus string theory will reduce to a field theory with a finite number of fields at distances $L >> l_s$ and energies $E << 1/l_s$. This is the limit in which we want to think of string theory as reducing to general relativity.

The fact that superstring theory includes general relativity is still one of the most striking and mysterious facts about the subject, and we will not be able to explain it satisfactorily here. Most of the explanations are indirect, developing other properties of the theory and eventually seeing that these imply the existence of a massless spin two particle, which by general arguments imply that the theory must include general relativity.

Perhaps the simplest indirect explanation of this sort makes gravity a consequence of supersymmetry. We will only describe this impressionistically: the motion of a string can be described by a two-dimensional field theory of maps from Σ, a two-dimensional manifold of topology $S^1 \times \mathbb{R}$ parameterizing the space-time history of a loop of string, into space-time $M \times K$. There are also two-dimensional field theories whose fields are fermionic, taking values in a space Λ with anticommuting coordinates, as well as more complicated possibilities. A theory of maps into $M \times K \times \Lambda$ will naturally have symmetries which mix all of the dimensions, and these are the supersymmetries.

We now grant that string theory can realize supersymmetry (in fact string theory came first, and was one of the clues which led to the discovery of supersymmetry), and consider the consequences of supersymmetry.

4.3. Supersymmetry and supergravity. After the dimension of space-time, the most basic characterization of a supersymmetric theory is the number of supercharges (in Minkowski space-time), which we denote Ns. The $\mathcal{N} = 1$ theories in four dimensions have $Ns = 4$, but the upper bound (for the known unitary theories) is $Ns = 32$.

There are three basic supergravity theories with $Ns = 32$. The simplest one, discussed in Acharya and Gauntlett's lectures, is formulated in 11 space-time dimensions; the supercharges form a real spinor representation of $SO(1, 10)$. Besides the metric, it contains a fermionic "gravitino," and a three-form gauge potential. The other two basic supergravities are formulated in 10 space-time dimensions, and the supercharges form spinor representations of $SO(1, 9)$ – depending on the choice of representation, one obtains the IIa or the IIb theory.

The basic example of a supersymmetric theory with $Ns = 16$ is super Yang-Mills theory in ten dimensions; this theory contains a fermionic superpartner called a "gaugino." As with pure Yang-Mills theory, these theories are uniquely determined by the choice of "gauge group," a finite dimensional compact semisimple Lie group G. These theories can also be coupled to gravity, to obtain "type I supergravity." In this case, the gauge coupling is determined by the expectation value of a complex scalar field usually called the "dilaton-axion."

The explicit Lagrangians and equations of motion for these theories are of course known and discussed in other contributions in this volume. More important for us here are the supersymmetry transformation laws, and the corresponding conditions for unbroken supersymmetry,

$$0 = \delta\psi_i = (D_i + \Gamma \cdot G)\epsilon \tag{4.1}$$

$$0 = \delta\chi = \Gamma^{ij} F_{ij}\epsilon. \tag{4.2}$$

Here G denotes a direct sum of the various gauge field strengths of the particular supergravity theory.

A background in which the supersymmetry variations of all fields are zero is one which preserves supersymmetry. In classical backgrounds, fermionic fields have zero expectation value, so this condition reduces to setting the right hand side of (4.1) to zero. The analysis of these conditions, and variants of them which we will come to, is the first step in the study of supersymmetric compactification.

4.4. Low energy supersymmetry from string theory. Let us start with the simplest case of $F = G = 0$. In this case, the conditions (4.1) reduce to $D_i\epsilon = 0$; in other words ϵ must be a covariantly constant spinor.

As is familiar, covariantly constant spinors are associated to a reduction of the global holonomy group of K from $SO(D)$ to a subgroup H, such that the spinor representation contains an H-invariant subspace. The possibilities, listed in Gauntlett's lectures, include all of those on Berger's list which support Ricci-flat metrics. The simplest, used in the early works, is $H = SU(n)$ with $n = D/2$. Manifolds with $SU(n)$ holonomy admit Ricci-flat metrics, and are thus solutions of supergravity. In particular, compactification of the $Ns = 16$ supergravity on a Calabi-Yau threefold, leads to an $Ns = 4$ supergravity in four dimensions.

On a Calabi-Yau, the second equation in (4.1) reduces to the Hermitian Yang-Mills equations (in the special case $c_1(V) = 0$), $F^{(2,0)} = F^{(1,1)} \wedge J \wedge J = 0$ (where J is the Kähler form). Now one appeals to the theorems of Donaldson and Uhlenbeck-Yau to argue that for a μ-stable holomorphic vector bundle V on K, these equations have a unique solution.

This class of theories provides the simplest candidates for realistic physics, as we discuss shortly. There are two more essential ingredients in the string theoretic discussion. First, it turns out that to quantize a field theory in ten dimensions, the spectrum of fermions must satisfy the very strong constraint of "anomaly cancellation." Without going into details, this arises because the determinant of the Dirac operator acting on an irreducible spinor representation in ten (or any even number of) dimensions is gauge invariant only up to a phase, and one must arrange for the non gauge invariant phase to cancel between representations. In ten-dimensional supergravity, this is true for the type II theories, and for type I theories only for the gauge groups $E_8 \times E_8$ and $Spin(32)/\mathbb{Z}_2$. Both can be realized in the heterotic string.

A full discussion of this point brings us to the second essential string theoretic modification, the "Green-Schwarz term." This imposes a topological constraint on the gauge bundle,

$$c_2(K) = c_2(V).$$

The resulting four-dimensional theory will have as a factor in its gauge group the endomorphism group of V as an $E_8 \times E_8$ or $Spin(32)/\mathbb{Z}_2$ bundle, while $\mathcal{C}$ will include the formal deformation space of V, as we shortly discuss.

4.5. Deriving the supersymmetric effective field theory. Suppose we grant the above, and find an interesting M and V; how would we go on to derive the effective field theory, to compare it with the supersymmetric Standard Model?

The first point to recognize is that the discussion above, based on Kaluza-Klein reduction of a supergravity theory, does not lead to exact results in superstring theory. Physically, one usually thinks of supergravity as the "weak coupling" and "large volume" limit of superstring theory. These limits are characterized by two parameters, controlling "stringy" and "quantum" corrections respectively.

The "stringy" corrections are functions of the length scales L of the internal manifold K, measured in units of the string length l_s; in other words of l_s/L, which go to zero as $L \to \infty$. These include "world-sheet instantons," maps from two-dimensional surfaces Σ into M, controlled by $\exp(-l_s^2/L^2)$ where L^2 is the volume of a two-cycle. In general they also include "perturbative" corrections, given by a general Taylor series in l_s/L. The "quantum" corrections are functions of the string coupling g_s (the stringy analog of Planck's constant $\hbar$) and can also be power-like or exponentially suppressed.

In any case, physicists argue that as $l_s/L \to 0$ and $g_s \to 0$ the supergravity description becomes good. Thus the style of analysis described above can be extended to derive the entire low energy effective Lagrangian in this limit.

Actually, g_s is controlled by a scalar field, the dilaton-axion, so a better way to think about the situation is that the effective Lagrangian data one derives in the supergravity approximation is an approximation to the true data, which is controlled in a subset of $\mathcal{C}$, and progressively less reliable away from this subset. The parameters l_s/L are also controlled by fields (the Kähler class of the Ricci-flat metric), so this statement of the situation is general: we only have direct information about a limit of the effective field theory.

For quite a while, this realization, and general arguments that our universe could not be described by the supergravity limit, turned most string theorists away from detailed exploration of the problem we now discuss. On the other hand, it does exhibit a lot of the structure one wants, and later considerations involving duality make even limiting results much more interesting. So let us continue.

On some level, the result is not hard to describe, in the case $G = 0$. The space $\mathcal{C}$ is a fibration, whose base is the product of the moduli space of Ricci-flat metrics on M with its natural metric with a copy of the hyperbolic disc (in which the dilaton-axion takes values).

The fiber F is harder to describe. One's first attempt would be to take it to be the moduli space of holomorphic $E_8 \times E_8$ (or $Spin(32)/\mathbb{Z}_2$) bundles on K. Within this moduli space, interesting bundles V lie on the reducibility locus with endomorphism group $SU(3)\times SU(2)\times U(1)$, while infinitesimal deformations away from this locus correspond to the spectrum of quarks and leptons (3.1).

Of course, this is problematic: in general there is no good moduli space, and even if we place restrictions to make it so (for example, consider μ-stable bundles), it is not a manifold. In favorable cases it might be a variety with quotient singularities, with some complicated obstruction theory. This is all relevant physical data; for example the Massey products translate into Yukawa couplings in the effective field theory.

One can postpone the problems of defining a moduli space by taking the fiber to be the space of holomorphic connections on K, described as deformations of a reference connection by an End G-valued $(0,1)$-form. Since K is Kähler, this space has a natural Kähler metric. The superpotential W is the holomorphic Chern-Simons action,

$$W[\bar{A}] = \int_M \Omega \wedge \mathrm{Tr}(\bar{A}\bar{\partial}\bar{A} + \frac{2}{3}\bar{A}^3). \tag{4.3}$$

This is a functional of $\bar{A}$ with the property that $\delta W/\delta\bar{A} = F^{(0,2)} = 0$; in other words its stationary points are connections on holomorphic bundles. Finally, the G-action is the action of the group of holomorphic maps from K into $G_{\mathbb{C}}$, and the corresponding moment maps μ are the Hermitian Yang-Mills equations, as explained in [**17**]. Reducible connections lead to enhanced gauge symmetry, as in our previous discussion, while the cubic coupling in (4.3) directly encodes the cubic terms in W which become the Yukawa couplings.

While this is morally an effective field theory, it is infinite dimensional and not concrete enough to address physicists' questions. We need ways to reduce this description to a finite number of variables. Now, many generalizations of "moduli space" have been suggested by mathematicians to capture this information and get a usable description of F. Personally, while one of my goals for the school was to start a serious math-physics discussion of these ideas, I am not sure this effort was a success; still, I hope some student will soon prove me wrong!

Perhaps more importantly, as we mentioned, the effective field theories one obtains are only approximations valid in the large volume limit; as we move away from this limit, the general result will obtain "quantum" and "stringy" corrections. Rather than make this approximate effective field theory explicit, and then add these corrections, one would rather find an equally general "geometric" characterization of the true effective field theory, and then find its explicit realizations.

5. Beyond differential geometry: the string world-sheet

So far we have been discussing applications of existing geometry – differential and algebraic – to string theory. Why should we expect string theory to require, or suggest, new concepts of geometry?

There are several intuitive reasons to expect this. Of course the most fundamental is that the primary definitions are not in terms of point-like particles but instead extended objects, strings and the "branes" we discuss below. One thus feels that concept of geometry based on points is somehow not "to the point," and one should start from new foundations.

These new foundations do not yet exist, though we will certainly come back to this topic below. Now, even if we try to formulate string theory using the standard language of field theory, we find not general relativity and Yang-Mills theory, but a two-parameter deformation of these theories. The case of finite string length l_s is still far better understood than that of finite g_s, and in this section we consider it. In string theory terms, it corresponds to "perturbative" string theory (so, processes involving one or a few strings), but moving through a background geometry in which curvature and other geometric length scales can be arbitrarily small.

The propagation of a string moving through any background geometry, which solves the stringy analog of Einstein's equations, is described by a two-dimensional conformal field theory, and thus we start here. This subject is on a somewhat better mathematical footing than general field theory or string theory, as there are many exactly solved cases (Wess-Zumino-Witten models, Gepner models, and so on) which either have been or are in the process of being rigorously formulated. We refer to [**56**] for a status report on this work, and continue our overview.

The most "geometric" definition of quantum field theory is the nonlinear sigma model. This is defined as a functional integral over maps from Σ, here a two-dimensional Riemannian manifold, to a target space K, a manifold with metric; denote this $g \in \mathrm{Met}(K)$. One can also supersymmetrize this model, adding fermions valued in the tangent bundle to K.

The foundational result here [**26**] is that the renormalization theory is geometric, *i.e.* can be done in a generally covariant way. While the arguments are perturbative, so is renormalization, so it is very plausible that an exact definition would share this property; of course more rigorous arguments would be very valuable.

The primary result of the renormalization theory is the renormalization group (RG) flow, a vector field on $\mathrm{Met}(K)$. This is a deformation of the Ricci flow, by corrections in a Taylor series in l_s and local tensors constructed from the metric. The additional couplings of string theory, such as the "B-field" and dilaton, can also be treated; adding the latter leads to the modified flow equations studied (for example) in [**46**].

Fixed points of the RG flow are conformal field theories. Very few examples are known for the original nonlinear sigma model, and these have either flat metric or additional background fields (for example, the Wess-Zumino-Witten model has a B-field) and at least hints at some underlying integrable structure. On the other hand, it has been argued for the supersymmetric sigma model that any Ricci-flat metric on K will correspond to a fixed point. Note that the fixed point metric is typically not Ricci-flat, due to the l_s corrections [**31**]; rather the obstruction to solving the corrected fixed point equations vanishes in perturbation theory [**45**].

5.1. Strings as loops, and T-duality. The approach we just described, based on constructing the stringy deformation order by order, is not too satisfying; one would rather characterize the theory at finite l_s, presumably in terms of some radical reformulation of geometry.

While this goal is widely accepted, other satisfactory starting points are not easy to find or work with. Perhaps the one which has been most studied is to formulate stringy geometry as a geometry of loop spaces. So far, attempts to do this, both physical and mathematical, rapidly get lost in the technicalities of infinite dimensional manifolds, renormalization theory, and so on. Some sort of conceptual breakthrough seems to be required to make useful progress in this direction.

Even without this, the intuitions coming from the picture of loops and extended objects have motivated many important developments. Perhaps the most simple intuitive ideas which came out of perturbative string theory are T-duality, of which mirror symmetry is an example, and the noncommutativity of open string theory.

T-duality is a symmetry of perturbative string theory which appears in compactification on tori, or torus fibrations. For example, a compactification of the type IIa string on an S^1 of circumference $2\pi R$ is physically identical to a compactification of the IIb string on an S^1 of circumference $2\pi l_s^2/R$.

The picture which leads to this equivalence is the following. Compactification of field theory on an S^1 leads to a spectrum of "Kaluza-Klein" or "momentum" modes, whose masses are given by the square roots of the eigenvalues of the Laplacian, say $M_n = n/R$ for $n \in \mathbb{Z}$. This is part of the string theory spectrum, obtained by quantizing string loops with trivial topology (zero winding number).

By itself, this spectrum is obviously not invariant under $R \to l_s^2/R$. On the other hand, string theory contains additional "winding modes," defined as string loops around the S^1 with fixed winding number, call this m. These have mass equal to the string tension, $1/2\pi l_s^2$, multiplied by the length of the geodesic with this winding number, $2\pi Rm$, so $M_m = Rm/l_s^2$. Now $R \to l_s^2/R$ is a symmetry of the union of these two mass spectra, exchanging momentum and winding modes. This intuitive argument can be made precise using string world-sheet techniques.

To generalize this to higher-dimensional tori, one starts by noting that the moduli space of string compactifications on T^d with $d > 1$ is larger than the space of metrics: one can also choose a harmonic two-form called the "Neveu-Schwarz B-field." The combined moduli space of metric and B-field can be shown to be the homogeneous space $SO(d,d,\mathbb{R})/SO(d) \times SO(d)$ before T-duality, and T-duality leads to identifications by any element of the discrete group $SO(d,d,\mathbb{Z})$. The resulting quotient space is a submanifold of $\mathcal{C}$ for type II string compactification on tori, but not all of $\mathcal{C}$ – we will discuss this further under "duality" below.

What is most important for later developments is that there is much evidence that T-duality applies not just to tori, but generalizes to any fibration with a torus action which (at least approximately) preserves the metric.

5.2. Mirror symmetry and topological string theory. One can also take the point of view that the framework of two-dimensional conformal field theory is itself the starting point for "stringy geometry." On the surface, this certainly looks very different from conventional geometry. For example, some conformal field theories can be defined in purely algebraic terms, without assuming that the fields take values in any target space. One then might try to find geometric notions which capture the structure of this algebraic problem, in the spirit of modern algebraic geometry.

While attractive, this point of view has not been easy to implement in generality, because by themselves the axioms of conformal field theory are far too weak to draw interesting conclusions. There is a special case of the problem, rational conformal field theory [**41**], which is more accessible, and this led to one of the most interesting early results, the construction of Gepner [**28**]. This was an explicit algebraic construction of certain $(2, 2)$ theories, which Gepner proposed (and others confirmed) were equivalent to sigma models with Calabi-Yau target space at particular points in moduli space. On the other hand, the class of rational theories is not preserved by deformation, and thus seems too special to form the foundation of a new geometry.

The most interesting CFT results so far are for the special case of the $(2, 2)$ models. To make a long story short (see [**39, 14, 29**]), the supersymmetric nonlinear sigma model with target space a complex Kähler manifold has additional world-sheet supersymmetry, the $(2, 2)$ superconformal algebra. This structure is very central in all relations between string theory and algebraic geometry. It also leads to the definition of "topological twisting" and the A and B twisted sigma models.

In early study of the $(2, 2)$ models with Calabi-Yau target space [**16, 35**], it was noted that the complex and Kähler moduli entered on a "symmetric" footing, and that their roles could be interchanged by a simple automorphism of the $(2, 2)$ algebra. This observation was developed by Greene and Plesser [**30**], who used Gepner's construction and identifications to propose a specific equivalence between a pair of sigma models with Calabi-Yau target. This proposal was the basis for the famous work of Candelas *et al* [**7**], counting curves using mirror symmetry.

While the direct CFT point of view remains difficult, much progress has been made in recent years by the use of the "linear sigma model" construction [**59**]. This essentially constructs Calabi-Yau target spaces as subvarieties of toric varieties obtained by an explicit quotient of $\mathbb{C}^n$ by holomorphic isometries. We refer to [**39**] for a detailed discussion of this topic, and the associated mathematical and physical proofs of mirror symmetry.

5.3. Dirichlet branes. Although historically Dirichlet branes were not much studied before the "duality revolution" we discuss below, they are also defined in terms of the string world-sheet, and thus belong logically to the current discussion.

Strings can be closed, with topology S^1, or open, with the topology of the interval. Intuitively, a Dirichlet brane is an allowed endpoint for an open string. Thus, in a geometrical framework, a Dirichlet brane could be specified by a choice of a submanifold of the target space, on which a particular type of open string is allowed to end.

There is an additional possible choice: since the end of an open string is a point, it behaves physically like a particle, and can be coupled to a Yang-Mills field. Thus, one can also specify a non-trivial Yang-Mills connection on the Dirichlet brane. In fact, when one quantizes the string, one finds that this degree of freedom emerges as a physical field, much in the same way that gravity emerged from the closed string. Thus, in the physical discussion one cannot leave this out.

On a more general level, the Dirichlet brane turns out to be the simplest way to embed Yang-Mills theory into the general framework of string theory, and has led to many striking connections between the mathematical and physical study of Yang-Mills theory. The simplest, which led to the original argument that Dirichlet branes must be included in string theory [**12**], is that T-duality acts on a Dirichlet brane as a sort of "Fourier-Mukai" or "spectral cover" transform. For example, a brane embedded in T^d carries a $U(1)$ connection, which by the Yang-Mills equations must be flat. Flat $U(1)$ connections of course parameterize points in a dual T^d, which turns out physically to be the moduli space of a T-dual Dirichlet brane which sits at a point in the dual torus.

The next such relation to be discovered was the equivalence between "point-like instantons" and Dirichlet branes [**61, 18**]. If one considers a Dirichlet brane with spatial dimension $p \geq 4$, the associated Yang-Mills field theory will have self-dual instanton solutions. Among their moduli is a scale size, and one can ask how string theory resolves the singularity which appears when the scale size goes to zero. In fact, one finds that, in the limit, the instanton becomes equivalent to a second Dirichlet brane of spatial dimension $p - 4$. Following this idea, one can find a purely physical rederivation of the ADHM construction of instanton moduli space [**3**], which admits many generalizations and has found many physical applications, as discussed in Dorey's lectures.

5.4. Noncommutative geometry. The physics of Dirichlet branes is rich and complicated, and one soon feels the need for organizing principles. One of the simplest of these starts with the elementary observation that an open string has two ends. Furthermore, if we consider oriented strings, an interaction of two open strings can take place when the end of the first string adjoins the beginning of the second. This can be distinguished from another possibility – the end of the second string could adjoin the beginning of the first – which leads to a different interaction.

These interactions can be formulated more precisely as the operator product algebra on the two-dimensional world-sheet. Whereas for closed strings this product is defined by taking a coincidence limit of two points on the sphere, a limit without any natural ordering, for open strings one takes a coincidence limit of points on the boundary of the disk, which are ordered. This distinction can be encoded algebraically as noncommutativity – whereas closed string operator products form a commutative algebra, open string operator products naturally define a noncommutative algebra.

One furthermore has a choice of boundary conditions – a choice of submanifold and connection in the geometric description, or some more abstract choice in general CFT. Since an open string has two ends, its spectrum depends on a choice of a pair of boundary conditions, say A and B, and a string ending on (say) B can only interact with another string beginning on B.

One way to encode these constraints is to say that Dirichlet branes (boundary conditions) are objects in a category, and open strings are morphisms. This is not the only way to think of this; one could instead say that open strings form a groupoid algebra, as in [**9**], but given a general and diverse set of boundary conditions the categorical language is perhaps more natural. In any case, the essential feature is the noncommutativity – in this sense, the Dirichlet branes define a "noncommutative geometry" associated to any CFT.

This idea has been made precise in several different ways. In principle, the most general would be to work with open string field theory, along the lines of [**57**], but so far this framework remains cumbersome. To avoid its difficulties, one must consider a special case or limit in which the operator product algebra of CFT reduces to an ordinary algebra. This essentially requires operator products to be independent of the positions of operators on the world sheet; in other words one must study topological string theory.

One special case in which the string reduces to a topological string is the study of a background Neveu-Schwarz B-field, as discussed by Cattaneo and Felder [**8**] following earlier work by Kontsevich [**34**]. In particular, in toroidal compactification, this leads to the noncommutative torus algebra, as was argued in works of Connes *et al* [**10**], Douglas and Hull [**21**], Schomerus [**52**] and Seiberg and Witten [**51**].

5.5. Open string mirror symmetry. Another special case in which the string can be made topological is the sigma model with Calabi-Yau target. Now the topologically twisted open string theories provide categories of Dirichlet branes, which can be defined using the data of the topological closed string theory.

For the B model, this data is a Calabi-Yau K considered as a complex variety. The obvious category which arises is the category of coherent sheaves on K.

For the A model, the closed string data is a Calabi-Yau K considered as a symplectic manifold. It is also clear from elementary considerations that world-sheet instantons with disk topology contribute to the operator product algebra. The natural category which these considerations suggest is the Fukaya category of isotopy classes of Lagrangian submanifolds, with morphisms given by Floer cohomology.

A naive conjecture of open string mirror symmetry would equate Coh M with Fuk W on its mirror. However, Kontsevich, who was the first to get this far, realized that this could not be – the two categories look too different in general. This observation was subsequently made concrete in many different ways. For example, mirror symmetry actually relates the B model on W, to a collection of A models on various CY's M, M', etc. which are birationally equivalent, but non-homeomorphic. While Fuk W does not depend on the B model data (the complex structure), Coh $M \neq$ Coh M' for non-homeomorphic Calabi-Yau's, so one gets a contradiction.

This contradiction can be fixed by instead postulating that the category arising from the topological B model open string is the derived category $D(\text{Coh } M)$, which (as first suggested by Bondal and Orlov [**5**]) is equivalent for birationally equivalent CY's. This is one way to arrive at Kontsevich's homological mirror symmetry conjecture, discussed in detail in the monograph [**40**].

Physical understanding of this idea came much later. A simpler physical picture of open string mirror symmetry is to relate it to T-duality. As we discussed, this naturally acts on torus fibrations, and this can fit with the known action of mirror symmetry on homology only if these are T^3 fibrations. This picture leads to the Strominger-Yau-Zaslow formulation of mirror symmetry [**53**], also discussed in [**40**], in which Dirichlet branes wrapping the T^3 on W are mirror to point-like branes on M.

As explained by Gauntlett, supersymmetry requires the relevant branes on W to wrap special Lagrangian cycles. This suggests a mirror relation between the complete set of special Lagrangian cycles on W and the coherent sheaves on M. However, early pictures of this relation relied too much on properties of the large volume approximation, and it seems fair to say that the full picture has not yet been spelled out. Making this full picture will clearly require a good deal of math-physics interaction, as indicated by a few results we now mention.

One of the simpler mathematical consequence of this symmetry, discussed in Szendrői's contribution, is a mirror relation between Fourier-Mukai transforms, which realize the autoequivalences on $D(\mathrm{Coh}\ M)$, and symplectomorphisms on W. Both transformations are "generalized T-dualities" in physics language, but the first is non-geometric, further demonstrating that $D(\mathrm{Coh}\ M)$ must play a physical role. This has been better understood, along lines discussed by Douglas in [**22**] and to be explained in [**40**].

Another small volume subtlety is discussed in Kapustin's contribution: to get mirror symmetry in general, the Fukaya category must be enlarged to include coisotropic submanifolds, objects which can also be shown to be supersymmetric Dirichlet branes.

5.6. Resolution of orbifold singularities. The simplest non-trivial non-commutative algebras of this sort arise as quotients of $\mathbb{C}^n$ by a discrete group action. Taking infinite groups provides the rich set of examples discussed in [**9**], but quotients by finite groups are very interesting as well, especially in the context of algebraic geometry, where one has well understood methods to resolve and study the resulting orbifold singularities.

Since quotients are natural, one would expect that all reasonable definitions of the quotient will agree. This is certainly borne out by the Dirichlet brane example, in which orbifolding leads directly to the concrete quiver theories which arose in the study of the McKay and generalized McKay correspondence [**48**]. This connection allowed immediately generalizing Witten's small instanton work to gauge fields on orbifold resolutions [**19**], and subsequent physical work has made fairly detailed contact with the mathematical work on the McKay correspondence, explained in contributions by Craw and Ishii to this volume. All of these topics will be discussed at length in [**40**].

6. The duality revolution

So far, we discussed the new geometry associated with l_s. The problem of understanding quantum g_s corrections is much more far-reaching, and much more difficult. Clearly one should start by understanding the g_s deformation in the absence of l_s corrections.

6.1. Quantum supersymmetric field theory. The typical problem here is to start with a classical Lagrangian, not including g_s corrections, and compute the exact answer with g_s corrections. Many such problems are discussed in the lectures of Dorey and Acharya. Traditionally, one starts by showing that the quantity of interest receives either no corrections in perturbation theory (say the superpotential in $d = 4$, $N = 1$ gauge theory), or very specific corrections (say the metric on $\mathcal{C}$ in $N = 2$ gauge theory, for which the only correction in the g_s expansion is the "one-loop" correction independent of g_s). This allows determining its behavior in certain limits and, using an intricate combination of symmetry arguments and global constraints on holomorphic functions, this can in simple cases be pushed to get exact results.

Perhaps the most famous example is the Seiberg-Witten solution of four-dimensional $\mathcal{N} = 2$ (or $Ns = 8$ in our conventions) gauge theory. In the language of $\mathcal{N} = 1$ effective field theory, this theory has $\mathcal{C} \cong \mathcal{G}_{\mathbb{C}}$ (the complexified Lie algebra) with the adjoint action of G, and the space of vacua is $\mathcal{C} //G \cong \mathcal{T}_{\mathbb{C}} / \mathrm{Weyl}(G)$ (the complexified Cartan subalgebra modulo the Weyl group). In the "exact solution" (meaning the exact low energy effective field theory; almost no other observables are computable at present), one learns (essentially) two things.

First, one learns the exact metric on $\mathcal{C}$, which (much as in mirror symmetry) is given by an infinite series summing instanton contributions (in a *tour de force* work [**44**], this was recently checked against the direct computation).

Second, one learns the locus in $\mathcal{C}$ on which new charged particles become massless. These are the magnetic monopoles and dyons visible as solitonic solutions at weak coupling, and this is the result which led to the formulation of the Seiberg-Witten equations: they are the equations of motion for the effective field theory of the monopole, and as such "must" (when suitably interpreted) reproduce the Donaldson invariants. The argument is that since these invariants are "topological" and do not depend on the metric, they cannot depend on the overall scale of energy, therefore they must be the same for the effective field theory as they were in the full quantum field theory.

Physically, this result determines the phase structure of the theory: in particular, the related $\mathcal{N} = 1$ theory obtained by adding a quadratic superpotential is in the confining phase. These ideas can be pushed further to compute the effective superpotential as a functional of the original superpotential (taken to be a general G-invariant function on $\mathcal{C}$). Indeed, many more exact $N = 1$ superpotentials are known for simple cases.

If we add a cubic or higher order superpotential, something even more unusual happens: the result is a "non-trivial fixed point theory" [**1**], a theory with quantum fluctuations at all length scales, analogous to the models of critical phenomena. Another example can be obtained by coupling matter in sufficiently many copies of the fundamental representation of G [**49**], and there is a whole zoo of further examples. Such theories do not behave classically in the low energy limit and thus are not well described by effective field theory; one needs more elaborate tools, more analogous to those of two-dimensional conformal field theory (indeed, it is generally believed that fixed point theories in any dimension are conformally invariant).

Shortly after the school, a general prescription for computing exact $N = 1$ superpotentials at finite g_s was found by Dijkgraaf and Vafa [**15**]. It states that the quantum superpotential can be obtained by a Legendre-type transform of the result of an integral over $\mathcal{C}$ treated as a zero-dimensional (hence, matrix model) integral.

6.2. The web of dualities. We turn to discuss the entire string theory at finite g_s. The remarkable and still amazing discovery of the "second superstring revolution," is that these theories are simple not just at small g_s, but also in the limit of very large g_s.

The prototypical example is the relation between IIa superstring theory in ten dimensions and eleven-dimensional supergravity. Naively, the IIa theory has a dimensionless parameter, the string coupling g_s. In fact, this parameter must be thought of as the expectation value of a field, the dilaton. As it becomes large, the theory actually becomes equivalent to eleven-dimensional supergravity compactified on $K \cong S^1$, of radius $R \sim g_s^{3/2}$.

Conversely, the spectrum of IIa supergravity can be obtained from eleven-dimensional supergravity by Kaluza-Klein reduction on S^1. Even the superstring can be obtained by wrapping the membrane on S^1, *i.e.* embedding one of the two spatial dimensions with the S^1, one obtains an object which looks like a string. In the small radius limit, the claim is that this becomes the string.

Arguing in this vein, one concludes that the well-definedness of superstring theory implies the existence of a "completion" of eleven-dimensional supergravity, which makes sense at all energy scales, and contains the membrane and a $(5+1)$-dimensional solution called the "fivebrane." This theory is generally called "M theory." It turns out that similar arguments can be used to connect the IIb, type I and heterotic string theories (suitably compactified), and this connected web of dual theories is sometimes also called "M theory." Evidently any starting point would have this property, and thus the more democratic and inclusive term "string/M theory" is also used.

While we do not have space to treat superstring duality in the depth it deserves, the interested reader will find many overviews which do in [**55**] and references there.

6.3. Branes and singularities. A point in the preceding discussion worthy of emphasis is the role of the spectrum of extended objects, or branes. This entered into our discussion of IIa-M theory duality, and can be generalized into the claim that, in a limit in which a wrapped brane becomes light, it will become a fundamental object in the dual theory. So, for example, the wrapped fivebrane on M theory compactified on the $K3$ manifold turns out to have the spectrum of the heterotic string compactified on T^3 (both seven-dimensional theories); thus these theories are dual.

Another variation on this argument is to consider a compactification with nontrivial cycles of small volume. In this case, the wrapped branes will lead to new localized degrees of freedom. This situation can be arranged by compactifying on a singular manifold and then resolving or deforming it slightly. Depending on the spectrum of branes, this can lead to instanton effects and a quantum deformation of the effective field theory, or it can lead to new light fields in the effective field theory.

To review the case of the fundamental string (which counts as a brane for this argument), the first effect is seen in mirror symmetry (world-sheet instantons associated with small two-cycles), while the second is seen in T-duality (new "winding states") associated with π_1.

A central ingredient in generalizing these results to higher-dimensional branes is calibrated geometry. As explained in Gauntlett's lectures, this provides the conditions for a wrapped brane to preserve supersymmetry. It can be generalized to compactifications with the non-metric fields (or "fluxes") turned on as well.

Of the many other cases, a particularly interesting one is the resolution of a canonical (or ADE) two-dimensional complex singularity, locally described by the orbifold $\mathbb{C}^2/\Gamma$ for $\Gamma \in SU(2)$. In a theory with a membrane, such as M theory, this leads to an effective field theory with enhanced gauge symmetry, with the corresponding ADE gauge group. In a bit more detail, the three-form gauge potential leads to the Cartan subalgebra of the gauge symmetry, while the wrapped membranes lead to the non-Abelian gauge bosons. This beautiful physical connection between two interpretations of ADE has numerous generalizations, for example to compactification on threefolds [**32, 20**].

Furthermore, it is perhaps the simplest example we know of for which the basic questions of "the geometry of string theory" have not been satisfactorily answered. A "correct stringy geometric" description of the singularity would make this G-symmetry manifest, and predict the corresponding behaviors in families, for higher-dimensional singularities, and so on. No fitting mathematicization or even summary of the somewhat *ad hoc* physical arguments has yet emerged.

6.4. G_2 and other supersymmetric compactifications. The basic analysis of supersymmetry we gave in section 4 of course has a role for most of the special holonomies. For example, compactification on a manifold of G_2 holonomy preserves 1/8 of supersymmetry. This is particularly interesting starting from M theory, as one then gets $\mathcal{N} = 1$ supersymmetry in four dimensions, and candidate physical models.

While we could have begun the discussion of G_2 compactification in an earlier section, it only becomes physically interesting when one considers compactification on singular manifolds. This is because an easy index theorem argument shows that on a smooth G_2 manifold the matter must form real G-representations. On the other hand, the Standard Model spectrum (3.1) involves a complex representation (physically, one says the Standard Model has chiral fermions).

The arguments of the last section tell us how to get non-Abelian gauge symmetry; we need to compactify on a G_2 manifold with an ADE singularity. This will lead to non-Abelian gauge symmetry localized in real codimension four. As discussed by Acharya, the simplest way to get chiral fermions is to have two such singularities intersect at a point (the metric behavior near the singularity turns out to allow this).

The novelty of these results has led to a lot of interest in G_2 compactification, but it must be admitted that the field is difficult as the appropriate mathematical foundations barely exist. The work of Joyce, reviewed in this volume, is of course the necessary starting point, and further discussion can be found in the contributions of Hitchin and Kovalev as well as Acharya.

6.5. AdS/CFT. One of the most beautiful, yet startling, results of superstring duality was the AdS/CFT correspondence, discovered by Maldacena [**36**]. To put the simplest example in a nutshell, $\mathcal{N} = 4$ supersymmetric Yang-Mills theory in four dimensions is dual to ten-dimensional IIb supergravity compactified on S^5. This duality between theories of different dimension will obviously require some explanation; we refer to the review [**2**] and the lectures of Acharya and Gauntlett in this volume.

As the name suggests, the central point is that $(D + 1)$-dimensional anti-de Sitter space-time admits as symmetry group $SO(D, 2)$, which is the same as the group of conformal isometries of D-dimensional Minkowski space-time. Thus, the simplest geometric questions which appear in this context involve string compactifications which lead to anti-de Sitter space, and variations of these solutions of supergravity.

In a real sense this duality goes beyond geometry, instead describing a sense in which geometry can emerge from "pre-geometrical" ingredients. Indeed, for a variety of reasons, many physicists have long suspected that a full theory of quantum gravity would **not** have the same relation to geometry as classical general relativity. After all, if the metric is a fluctuating quantum variable, it becomes difficult to formulate the standard geometric observables such as geodesic or holonomy. On the other hand, one might imagine that other, simpler observables might appear. In AdS/CFT, these are boundary conditions on the fields in anti-de Sitter space, dual to couplings in the gauge theory.

Some more mathematical papers inspired by these ideas include [**62, 37**].

7. Some reflections

As we have seen, string theory is a very rich subject, and the project of understanding it mathematically has barely begun. Most of us who attended the school, and many readers, would happily agree with the claim that this project is destined to be interesting for both physics and mathematics, and does not require further justification. Let me nevertheless offer a few thoughts on where this interaction is going, and what we can hope will come out of it.

We devoted most of our attention to the "classic" problem of string compactification, to explain or at least reproduce the physics we have seen in experiments and observations. While the original scenarios for this involved a fairly direct (though ingenious) generalization of the Kaluza-Klein idea, during the 1980's string theorists came to realize that string theory was not adequately described by existing notions of geometry, due to the l_s and g_s corrections. This led to the idea that string theory would lead to completely new ideas of geometry.

On the other hand, the progress on this problem over the years has largely come through a more modest agenda: one works on a subproblem which can be formulated in geometric terms, but for which string theory provides a "twist." For example, both sides of closed string mirror symmetry can be formulated in these terms, counting curves and varying Hodge structures; the twist is to connect these seemingly very different problems. A related but more recent example is the description of Dirichlet branes as objects in the derived category of coherent sheaves, a notion formulated by mathematicians, but satisfying a modified stability condition motivated by string theory, which incorporates the l_s corrections [**22**].

Such developments seem very likely to continue, and this should be to the benefit of both sides, but at present seem of more clear interest to mathematicians than to physicists. Indeed, for those who feel string theory has something to say about the real world, the following quote may be telling:

> There is probably less difference between the positions of a mathematician and of a physicist than is generally supposed, and the most important seems to me to be this, that the mathematician is in much more direct contact with reality.

This was from G. H. Hardy's *A Mathematician's Apology*, but it seems to me strangely appropriate as a description of the string theorists' present situation.

Of course, Dirac is particularly well known for the point of view that the beauty of equations is itself evidence for their correctness, but Hardy's point is also well taken: the mathematician knows what he (or she) is talking about, and in the act of precisely formulating it makes it real, at least real as mathematics.[2]

String theory has far passed the threshold of convincing us that it is real in this sense, despite our inability to precisely formulate it. On the other hand, it is still hard to say how we will be convinced that it describes the physical world. It could be that strings will be discovered in accelerators, but if not, it seems very possible that we will have to live for some time with the argument that it is the only candidate which is not "obviously wrong."

One of the striking points which the school brought home to me is how different the cultures of mathematics and physics are. It is interesting to ask whether or not this interaction will ultimately lead to a merging of the two, or perhaps the formation of some sort of intermediate culture.

One of the well-known differences in culture is the attitude towards generalization. Physicists are always happier to consider examples, and feel that these bring home the essential "point" of an idea or development much better than any abstract treatment. On the other hand, while mathematicians regard examples as essential as well, one of the defining characteristics of mathematics is the development of general theory.

This is a very deep difference. It comes to the fore when one tries to prove that some property or attribute of the objects under consideration is **not** possible, as most such proofs require a general definition which does not presuppose the existence of an object with this property. In some cases, one can make *reductio ad absurdum* proofs, but on reflection one realizes that this presupposes a far better understanding of the objects under discussion than physicists usually can get. Indeed, there is a saying among physicists that "no-go" theorems, which state that some desired property or construction is impossible, would better be renamed "go-go" theorems, as there are so many examples of loopholes or other ways around the supposed impossibility. While this is more an illustration of selective memory than anything else, it shows the deep distrust of physicists for general arguments.

If we ever attain a reasonably complete formulation of string/M theory, this attitude will have to change. After all, by definition, we will have reached the point where the axioms are not just obstacles to work one's way around, but inevitable constraints on the possibilities. Indeed, we must hope that these constraints are strong enough in the end to enable us to make predictions. The most basic test of whether a theory has content is whether it can be wrong or, more precisely, falsified by the observations. Assuming that physicists do discover supersymmetry and additional phenomena, but only those which can be described by four-dimensional effective field theory, the primary question which string theory will need to answer is: can the effective field theory of our world arise from string theory? One can answer this question by finding the "correct compactification," but this is meaningless if any effective theory can be obtained from some compactification. From this point of view, the question of showing that certain possibilities cannot come out of string theory is just as important as showing that the one which describes

[2]On the other hand, Jerome Gauntlett reminded me of another quote: "Mathematics may be defined as the subject where we never know what we are talking about, nor whether what we are saying is true." (Bertrand Russell, *Mysticism and Logic*).

our universe can arise. This is at heart a mathematical question, in the sense I just described, and it seems to me that physicists will have to adopt something of the spirit of the mathematicians to answer it.

Comparing my colleagues' attitudes with those I recall from my graduate studies in the early 80's, I can only say that this process has come a long way already, and probably has farther to go. As to whether mathematicians are being influenced in equally deep ways, I leave for them, and especially the students in this school, to tell me.

References

1. P. C. Argyres and M. R. Douglas, *New phenomena in SU(3) supersymmetric gauge theory*, Nucl. Phys. B **448**, 93 (1995) [arXiv:hep-th/9505062].
2. O. Aharony, S. S. Gubser, J. M. Maldacena, H. Ooguri, and Y. Oz, *Large N field theories, string theory and gravity*, Phys. Rept. **323** (2000), 183–386.
3. M. Atiyah, V. Drinfeld, N. Hitchin and Y. Manin, *Construction of instantons*, Phys. Lett. **A65**:185, 1978.
4. T. Banks, M. Dine and E. Gorbatov, *Is there a string theory landscape?*, arXiv:hep-th/0309170.
5. A. Bondal and D. Orlov, *Derived Categories of Coherent Sheaves*, in Proceedings of ICM 2002, vol. 2, pp. 47–56, Higher Education Press, Beijing.
6. P. Candelas, G. T. Horowitz, A. Strominger and E. Witten, *Vacuum configurations for superstrings*, Nucl. Phys. B. 258:46-74, 1985.
7. P. Candelas, X. C. De La Ossa, P. S. Green and L. Parkes, *A Pair Of Calabi-Yau Manifolds As An Exactly Soluble Superconformal Theory*, Nucl. Phys. B **359**, 21 (1991).
8. A. S. Cattaneo and G. Felder, *A path integral approach to the Kontsevich quantization formula*, Commun. Math. Phys. **212**, 591 (2000) [arXiv:math.qa/9902090].
9. A. Connes, Noncommutative Geometry, 1994, Academic Press, San Diego.
10. A. Connes, M. R. Douglas and A. Schwarz, *Noncommutative Geometry and Matrix Theory: Compactification on Tori*, JHEP 9802 (1998) 003, [arXiv:hep-th/9711162].
11. E. Cremmer, B. Julia, and J. Scherk, *Supergravity theory in 11 dimensions*, Phys. Lett. **B76** (1978), 409–412.
12. J. Dai, R. G. Leigh and J. Polchinski, *New Connections Between String Theories*, Mod. Phys. Lett. A **4**, 2073 (1989).
13. P. Deligne and J. W. Morgan, *Notes on Supersymmetry (following J. Bernstein)*, in *Quantum Fields and Strings: A course for Mathematicians*, eds. Deligne *et al*, pp. 41–97, 2 vols, AMS, Providence, 1999.
14. R. Dijkgraaf, *Fields, strings and duality*, in *Les Houches 1995, Quantum symmetries,* eds. A. Connes, K. Gawędzki and J. Zinn-Justin, pp. 3–148, North Holland, 1998 [arXiv:hep-th/9703136].
15. R. Dijkgraaf and C. Vafa, *A perturbative window into non-perturbative physics*, [arXiv:hep-th/0208048].
16. L. Dixon, in *Superstrings, Unified Theories and Cosmology 1987*, eds. G. Furlan *et al*, World Scientific 1988.
17. S. K. Donaldson and P. B. Kronheimer, *The Geometry of Four-Manifolds,* Oxford Univ. Press, 1990.
18. M. R. Douglas, *Branes within branes*, in *Strings, Branes and Dualities,* eds. L. Baulieu et al., pp. 267–276, Kluwer 1999. [arXiv:hep-th/9512077]
19. M. R. Douglas and G. Moore, *D-branes, Quivers, and ALE Instantons,* [arXiv:hep-th/9603167].
20. M. R. Douglas, S. Katz and C. Vafa, *Small instantons, del Pezzo surfaces and type I' theory*, Nucl. Phys. B **497**, 155 (1997) [arXiv:hep-th/9609071].
21. M. R. Douglas and C. Hull, *D-branes and the Noncommutative Torus*, JHEP 9802 (1998) 008, [arXiv:hep-th/9711165].
22. M. R. Douglas, *Dirichlet branes, homological mirror symmetry, and stability*, in Proceedings of ICM 2002, *op. cit.*, vol. 3, pp. 395–408, [arXiv:math.AG/0207021].
23. R. G. Douglas, private communication.

24. D. S. Freed, *Five Lectures on Supersymmetry*, AMS, Providence, 1999.
25. P. G. Freund and M. A. Rubin, *Dynamics Of Dimensional Reduction*, Phys. Lett. B **97**, 233 (1980).
26. D. H. Friedan, *Nonlinear Models In Two + ϵ Dimensions*, Annals Phys. **163**, 318 (1985).
27. H. Georgi, H. R. Quinn and S. Weinberg, *Hierarchy Of Interactions In Unified Gauge Theories*, Phys. Rev. Lett. **33**, 451 (1974).
28. D. Gepner, *Space-time supersymmetry in compactified string theory and superconformal models*, Nucl. Phys. **B296** (1988) 757.
29. B. R. Greene, *Lectures on the Quantum Geometry of String Theory*, in *Les Houches 1995, Quantum symmetries,* op. cit.
30. B. R. Greene and M. R. Plesser, *Duality In Calabi-Yau Moduli Space*, Nucl. Phys. B **338**, 15 (1990).
31. M. T. Grisaru, A. E. van de Ven and D. Zanon, *Four-Loop Divergences For The N=1 Supersymmetric Nonlinear Sigma Model In Two-Dimensions*, Nucl. Phys. B **277**, 409 (1986).
32. K. A. Intriligator, D. R. Morrison and N. Seiberg, *Five-dimensional supersymmetric gauge theories and degenerations of Calabi-Yau spaces*, Nucl. Phys. B **497**, 56 (1997) [arXiv:hep-th/9702198].
33. M. Kontsevich, *Homological Algebra of Mirror Symmetry,* Proceedings of the 1994 ICM, [arXiv:alg-geom/9411018].
34. M. Kontsevich, *Deformation quantization of Poisson manifolds, I*, [arXiv:q-alg/9709040].
35. W. Lerche, C. Vafa and N. P. Warner, *Chiral Rings In N=2 Superconformal Theories*, Nucl. Phys. B **324**, 427 (1989).
36. J. M. Maldacena, *The large N limit of superconformal field theories and supergravity*, Adv. Theor. Math. Phys. **2** (1998), 231–252.
37. Y. I. Manin and M. Marcolli, *Holography principle and arithmetic of algebraic curves*, Adv. Theor. Math. Phys. **5**, 617 (2002) [arXiv:hep-th/0201036].
38. M. Mariño, R. Minasian, G. W. Moore, and A. Strominger, *Nonlinear instantons from supersymmetric p-branes*, JHEP **01** (2000), 005.
39. *Mirror Symmetry*, proceedings of the 2000 Clay School, eds. K. Hori *et al*, AMS 2003.
40. *Mirror Symmetry II*, proceedings of the 2002 Clay School, eds. P. Aspinwall *et al*, in preparation.
41. G. W. Moore and N. Seiberg, *Classical And Quantum Conformal Field Theory*, Commun. Math. Phys. **123**, 177 (1989).
42. G. W. Moore and E. Witten, *Integration over the u-plane in Donaldson theory*, Adv. Theor. Math. Phys. **1**, 298 (1998) [arXiv:hep-th/9709193].
43. J. W. Morgan, *The Seiberg-Witten equations and applications to the topology of smooth four-manifolds*, Princeton Univ. Press, 1996.
44. N. A. Nekrasov, *Seiberg-Witten prepotential from instanton counting*, in Proceedings of ICM 2002, *op. cit.*, vol. 3, pp. 477-496, [arXiv:hep-th/0306211].
45. D. Nemeschansky and A. Sen, *Conformal Invariance Of Supersymmetric Sigma Models On Calabi-Yau Manifolds*, Phys. Lett. B **178**, 365 (1986).
46. G. Perelman, *The entropy formula for the Ricci flow and its geometric applications*, [arXiv:math.DG/0211159].
47. Proceedings of ICM 2002, ed. Li Tatsien, 3 vols., Higher Education Press, Beijing, 2002.
48. M. Reid, *La correspondance de McKay,* Séminaire Bourbaki (novembre 1999), no. 867, [arXiv:math.AG/9911165].
49. N. Seiberg, *Electric - magnetic duality in supersymmetric non-Abelian gauge theories*, Nucl. Phys. B **435**, 129 (1995) [arXiv:hep-th/9411149].
50. N. Seiberg and E. Witten, *E*lectric-magnetic duality, monopole condensation, and confinement in N=2 supersymmetric Yang-Mills theory, Nucl. Phys. B **426**, 19 (1994) [Erratum-ibid. B **430**, 485 (1994)] [arXiv:hep-th/9407087].
51. N. Seiberg and E. Witten, *String Theory and Noncommutative Geometry*, JHEP 9909 (1999) 032, [arXiv:hep-th/9908142].
52. V. Schomerus, *D-branes and Deformation Quantization*, JHEP 9906 (1999) 030, [arXiv:hep-th/9903205].
53. A. Strominger, S. T. Yau and E. Zaslow, *Mirror symmetry is T-duality*, Nucl. Phys. B **479**, 243 (1996) [arXiv:hep-th/9606040].

54. R.P. Thomas, *Examples of bundles on Calabi-Yau 3-folds for string theory compactifications,* Adv. Theor. Math. Phys. 4 (2000) 231-247.
55. Unity from Duality: Gravity, gauge theory and strings, eds. C. Bachas *et al*, proceedings of the 2001 Les Houches Advanced Study Institute, Springer, 2002.
56. Vertex Operator Algebras in Mathematics and Physics, eds. Y.-Z. Huang *et al*, Fields Institute Communications, AMS, Providence, 2003.
57. E. Witten, *Noncommutative geometry and string field theory*, Nucl. Phys. **B268** (1986) 253.
58. E. Witten, *Topological Quantum Field Theory*, Commun. Math. Phys. **117** (1988) 353.
59. E. Witten, *Phases of $N = 2$ theories in two dimensions*, Nucl. Phys. B **403**, 159 (1993) [arXiv:hep-th/9301042].
60. E. Witten, *Monopoles and four-manifolds*, Math. Res. Lett. **1**, 769 (1994) [arXiv:hep-th/9411102].
61. E. Witten, *Small Instantons in String Theory*, Nucl. Phys. B. **460**, 541 (1996) [arXiv:hep-th/9511030].
62. E. Witten and S. T. Yau, *Connectedness of the boundary in the AdS/CFT correspondence*, Adv. Theor. Math. Phys. **3**, 1635 (1999) [arXiv:hep-th/9910245].
63. E. Witten, *Deconstruction, G_2 Holonomy, and Doublet-Triplet Splitting*, [arXiv:hep-ph/0201018].
64. E. Witten, *Singularities in String Theory,* in Proceedings of ICM 2002, *op. cit.*, vol. 1, pp. 495-506.
65. C. N. Yang, as discussed in *Selected Papers 1945–1980*, pp. 73–76, W. H. Freeman and Co., New York, 1983.

Department of Physics and Astronomy, Rutgers University, Piscataway, NJ 08855-0849, USA, Isaac Newton Institute for Mathematical Sciences, Cambridge, CB3 0EH, U.K., I.H.E.S., Le Bois-Marie, Bures-sur-Yvette, 91440 France

E-mail address: `mrd@physics.rutgers.edu`

Clay Mathematics Proceedings
Volume 3, 2004

M theory, G_2-manifolds and Four Dimensional Physics

Bobby S. Acharya

ABSTRACT. M theory on a manifold X of G_2-holonomy is a natural framework for obtaining vacua with four large spacetime dimensions and $\mathcal{N} = 1$ supersymmetry. The standard features of particle physics, namely non-Abelian gauge symmetries and chiral fermions, emerge from singularities of X. The aim of these lectures is to describe in detail how the above picture emerges. Along the way we will see how interesting aspects of strongly coupled gauge theories, such as confinement and mass gap, receive relatively simple explanations within the context of M theory. All of the singularities of G_2-manifolds we discuss here are are constructed naturally from the familiar ADE-singularities. For instance, codimension seven singular points which support chiral fermions are obtained from a natural modification of the Kronheimer construction of ALE-spaces.

1. Introdution

Supersymmetry is one of our best candidates for physics beyond the Standard Model. M theory goes further in the sense that it is supersymmetric, contains gravity and is quantum mechanically consistent. Since the theory is formulated on spacetimes with eleven dimensions, a natural question to ask is whether there are vacua of M theory with four macroscopic spacetime dimensions *and* a realistic particle physics spectrum? Since supersymmetry is intrinsic to M theory, it is perhaps more natural to look for vacua with supersymmetric particle physics in four dimensions. Since non-minimal or extended supersymmetry in four dimensions cannot accommodate chiral fermions, to answer this question we should really be studying M theory vacua with $\mathcal{N} = 1$ supersymmetry.

There are two (to date) natural looking ways to obtain four large spacetime dimensions with $\mathcal{N} = 1$ supersymmetry from M theory. Both of these require the seven extra dimensions to form a manifold X whose metric obeys certain properties. The first consists of taking X to be a manifold whose boundary ∂X is a Calabi-Yau threefold [**15**]. The second possibility, which will be the subject of these lectures is to take X to be a manifold of G_2-holonomy[1]. We will explain what this means in section two.

2000 *Mathematics Subject Classification.* Primary 81T30, 83E30, 53C29, 81T60.
Key words and phrases. M theory, G_2-holonomy.

[1]We will often refer to X as a G_2-manifold.

If X is large compared to the Planck scale (the only scale in M theory) and smooth, then at low energies a good approximation is provided by eleven-dimensional supergravity. Compactifications of the latter have been studied for several decades. See [**13**] for a review. Unfortunately, Witten proved that none of these could give rise to chiral fermions [**22**]. However, this does not mean that G_2-manifolds are useless for obtaining models of particle physics from a fundamental theory. This is because we have learned in recent years that additional light degrees of freedom can be "hidden" at *singularities* of X. These are typically branes wrapped on submanifolds of X which have shrunk at the singularity[2]. In M theory these are either the $M2$-brane or the $M5$-brane. This can provide a novel picture of conventional field theory dynamics and can even lead to new theories. The supergravity approximation breaks down at such singularities and the analysis of [**22**] no longer applies.

Within the past couple of years there has been a tremendous amount of progress in understanding M theory physics near singularities in manifolds of G_2-holonomy $[4-10]$. In particular we now understand at which kinds of singularities in G_2-manifolds the basic requisites of the standard model - non-Abelian gauge groups and chiral fermions - are to be found. The purpose of these lectures will be to explain how this picture was developed in detail. Along the way we will see how important properties of strongly coupled gauge theories such as confinement and the mass gap can receive a semi-classical description in M theory on G_2-manifolds.

The subject of these lectures may also be of interest to mathematicians. One compelling mathematical theme is ADE-singularities. As we will see, all of the singularities of interest here are built out of the basic ADE orbifold singularities. Yang-Mills fields in four dimensions emerge by considering singularities of X parameterised by an associative 3-manifold $W \subset X$, near which X looks like $\mathbb{C}^2/\Gamma_{\mathbb{ADE}} \times W$. Chiral fermions emerge from additional singular points in X through which W passes and near which the description of X is given by a simple modification of the Kronheimer construction of ADE-singularities: one simply picks a $U(1)$ in the Kronheimer gauge group and does not set its corresponding moment map to zero. One then obtains a 7-manifold which fibers over $\mathbb{R}^3$ (given by the set of values of this privileged moment map) and the fibers of this map are ADE-singularities. At the origin in $\mathbb{R}^3$ the 7-manifold becomes more singular since some collection of S^2's in the fibers collapse at that point.

Another point which may also be of interest to mathematicians - especially to those interested in mirror symmetry - will be a duality which we exploit here between M theory on $K3$-fibered G_2-manifolds and heterotic string theory on T^3-fibered Calabi-Yau threefolds. On the string theory side the threefold is endowed with a Hermitian-Yang-Mills connection A and chiral fermions emerge from zero modes of the Dirac operator twisted by A. On the M theory side this gets mapped to a statement about the singularities of X, which is how the aforementioned description of singularities is obtained.

By the end of these lectures, we should have a good and clear picture of the basic properties that X should have in order that M theory on X produces something like a realistic model of particle physics in four dimensions. An important problem

[2]A more conventional example of light states at a singularity is provided by string theory on an orbifold. Typically one finds extra light states confined to the singularity. These arise in the so-called twisted sectors.

- which we would like to pose to the mathematicians - is to construct compact G_2-manifolds with these properties.

At the beginning of each main section we will offer a section summary. In section two we derive the basic properties of M theory when X is large and smooth. We also derive some basic properties of G_2-manifolds. Section three explains how classical supersymmetric Yang-Mills theory can be obtained from M theory on a singular G_2-manifold X. We describe these singularities in detail. Section four describes how quantum properties of the Yang-Mills theory, confinement and a mass gap, can be understood from M theory. The reason that this can be done successfully is that M theory contains semi-classical limits which are not present in the quantum gauge theory. Having understood how non-Abelian gauge groups emerge, section five goes on to describe how additional singularities of X give rise to chiral fermions.

2. Supersymmetry, G_2-holonomy and Kaluza Klein spectrum.

In this section we will describe in detail why G_2-holonomy manifolds naturally emerge in the context of supersymmetric M theory compactification. We will describe some of the basic properties of G_2-manifolds. We will then discuss the Kaluza-Klein spectrum of M theory on a large and smooth G_2-manifold.

2.1. Supersymmetry and G_2-holonomy. At low energies M theory admits a description in terms of eleven-dimensional supergravity. This description is valid on smooth spacetimes whose smallest length scale is much larger than the eleven-dimensional Planck length. The supergravity contains three fields, a metric g, a three-form potential C and a gravitino Ψ. In addition to being generally covariant and supersymmetric, the theory has a gauge invariance under which

$$\delta C = d\lambda \tag{2.1}$$

with λ a 2-form, so the gauge invariant field is the derivative of C, denoted by G. The action for the bosonic fields is of the form

$$S = \int \sqrt{g}R - \frac{1}{2}G \wedge *G - \frac{1}{6}C \wedge G \wedge G. \tag{2.2}$$

The equations of motion for C and g are of the form,

$$d * G = G \wedge G \tag{2.3}$$

and

$$R_{MN} - \frac{1}{2}g_{MN}R = T_{MN}(C) \tag{2.4}$$

where T is the energy-momentum tensor for the C-field.

Since the theory is supersymmetric, it is natural to look for supersymmetric vacua. In the classical theory these are just the conditions that the supersymmetry variations of the three fields vanish. In a Lorentz invariant background the expectation value of Ψ is zero, in which case the variations of g and C vanish automatically. In order to find classically supersymmetric field configurations we must find values of C and g for which the variation of Ψ is zero:

$$\delta_\eta \Psi_M \equiv \nabla_M \eta + \frac{1}{288}(\Gamma_M^{PQRS} G_{PQRS} - 6\Gamma^{PQR} G_{MPQR})\eta = 0. \tag{2.5}$$

The simplest way to solve these equations is to take $G = 0$ in which case we are looking for 11-manifolds with metric g which admit a covariantly constant or parallel spinor:

$$\nabla_M \eta = 0. \tag{2.6}$$

We will re-write this equation in the more symbolic form,

$$\nabla_g \eta = 0 \tag{2.7}$$

where by ∇_g we mean the Levi-Civita connection constructed from g. Solutions to these conditions can be classified via the holonomy group of the connection ∇_g.

The holonomy group of a connection acting on a field like η or a vector field can be understood in terms of parallel transport. One takes a closed loop on the manifold and literally transports the field around it. In the case of the Levi-Civita connection, the field comes back to itself up to a rotation in $SO(n)$ in the case of Riemannian n-manifolds or $SO(10,1)$ in the case of M theory. The set of all such rotations based at some point on the manifold generates a group, the holonomy group of ∇_g, $Hol(g)$. If g is sufficiently *generic* then $Hol(g) = SO(10,1)$. However, if we take special choices of g, then $Hol(g)$ can be a proper subgroup of $SO(10,1)$. For instance, if g is such that parallel spinors (supersymmetries) can be found, then η is a field which undergoes no parallel transport at all and therefore $Hol(g)$ must be a subgroup of $SO(10,1)$ for which there are spinors in the trivial representation.

We will concern ourselves with compactifications of M theory to four dimensions on a 7-manifold X. More precisely we will take the eleven-manifold to be a product $X \times \mathbb{R}^{3,1}$ with g a product metric of a metric $g(X)$ and the Minkowski metric on $\mathbb{R}^{3,1}$. With this choice of eleven-metric, we have explicitly broken $SO(10,1)$ to $SO(7) \times SO(3,1)$. The second factor now plays the role of the Lorentz group of the compactified theory. The conditions for supersymmetry can be satisfied by taking $g(X)$ to be such that it admits a spinor θ obeying

$$\nabla_{g(X)} \theta = 0 \tag{2.8}$$

and choosing

$$\eta = \theta \otimes \epsilon \tag{2.9}$$

with ϵ a basis of constant spinors in Minkowski space.

The condition

$$\nabla_{g(X)} \theta = 0 \tag{2.10}$$

implies that $Hol(g(X))$ is G_2 or a subgroup. This is because G_2 is the maximal proper subgroup of $SO(7)$ under which the spinor representation contains a singlet. Specifically, a spinor of $SO(7)$ can be regarded as a fundamental of G_2 and a singlet:

$$\mathbf{8} \to \mathbf{7} + \mathbf{1}. \tag{2.11}$$

Therefore, if X is a compact manifold of precisely G_2-holonomy, the effective theory in four dimensions will be minimally $\mathcal{N} = 1$ supersymmetric. We get precisely $\mathcal{N} = 1$ and no more because there is only one singlet spinor according to the above group theory.

2.2. Properties of G_2-manifolds. From the parallel spinor θ we can construct other covariantly constant fields on X. More precisely, any p-form with components

$$\theta^T \Gamma_{i_1 i_2 \ldots i_p} \theta \tag{2.12}$$

is obviously parallel with respect to $\nabla_{g(X)}$. In fact, since the antisymmetric representations of $SO(7)$, when decomposed as representations of G_2, contain singlets only when p is $0, 3, 4, 7$ the above p-forms are non-zero precisely for these values. The 0-form is just a constant on X. The 7-form is the volume form. Locally the three-form, which we will conventionally denote by φ, can be regarded as a set of structure constants for the octonion algebra. This stems from the fact that G_2 is the automorphism group of the octonion algebra, where we regard the tangent space at a point on X as a copy of $\mathbf{ImO}$, the imaginary octonions. A specific representation of φ locally is

$$\varphi_0 = dx_{123} + dx_{145} + dx_{167} + dx_{246} - dx_{257} - dx_{347} - dx_{356} \tag{2.13}$$

where the subscript refers to the fact that we are considering a local model. The covariantly constant 4-form is the Hodge dual of φ, which in the local model is given by

$$*\varphi_0 = dx_{4567} + dx_{2367} + dx_{2345} + dx_{1357} - dx_{1346} - dx_{1256} - dx_{1247}. \tag{2.14}$$

In addition to implying that there are other parallel fields on X, the existence of a parallel spinor (or a G_2-holonomy metric) also has other implications. One of these is that the metric of G_2-holonomy, $g(X)$, is Ricci-flat. To see this, observe that the commutator of the covariant derivative is the Riemann curvature. Acting on θ this implies

$$[\nabla_{g(X)}, \nabla_{g(X)}]_{mn} \theta = \frac{1}{4} R_{mnpq} \Gamma^{pq} \theta = 0. \tag{2.15}$$

Now, contract again with a Γ-matrix to obtain,

$$\Gamma^n \Gamma^{pq} R_{mnpq} \theta = 0 \tag{2.16}$$

The Bianchi identity for R_{mnpq} which asserts that the components totally antisymmetric in $[npq]$ are zero then implies that

$$\Gamma^j R_{ij} \theta = 0 \tag{2.17}$$

which implies that the Ricci tensor vanishes. This shows that G_2-manifolds obey the equations of motion of $d = 11$ supergravity when the 4-form field strength G and the gravitino are zero: these equations are simply the vacuum Einstein equations.

A final implication - whose proof goes beyond the scope of these lectures but which can be found in [**16**] - is that compact manifolds with $Hol(g) = G_2$ have a finite fundamental group. This implies that the first Betti number vanishes.

2.3. Kaluza-Klein Reduction. At low energies, the eleven-dimensional supergravity approximation is valid when spacetime is smooth and large compared to the eleven-dimensional Planck length. So, when X is smooth and large enough, we can obtain an effective four-dimensional description by considering a Kaluza- Klein analysis of the fields on X. This analysis was first carried out in [**20**].

In compactification of eleven-dimensional supergravity, massless scalars in four dimensions can originate from either the metric or the C-field. If $g(X)$ contains k

parameters, i.e. there is a k-dimensional family of G_2-holonomy metrics on X, then there will be correspondingly k massless scalars in four dimensions.

The scalars in four dimensions which originate from C arise via the Kaluza-Klein ansatz,

$$C = \Sigma_I \omega^I(x)\phi_I(y) + ... \tag{2.18}$$

where ω^I form a basis for the harmonic 3-forms on X. These are zero modes of the Laplacian on X and are also closed. There are $b_3(X)$ linearly independent such forms. The dots refer to further terms in the Kaluza-Klein ansatz which will be prescribed later. The $\phi_I(y)$ are scalar fields in four-dimensional Minkowski space with coordinates y. With this ansatz, these scalars are classically massless in four dimensions. To see this, note that,

$$G = \Sigma_I \omega^I \wedge d\phi_I \tag{2.19}$$

and $d * G$ is just

$$d * G = *\Sigma_I \omega^I d * d\phi_I. \tag{2.20}$$

Since $G\wedge G$ vanishes identically, the equations of motion actually assert that the scalar fields ϕ_I are all massless in four dimensions. Thus, the C-field gives rise to $b_3(X)$ real massless scalars in four dimensions.

In fact it now follows from $\mathcal{N} = 1$ supersymmetry in four dimensions that the Kaluza-Klein analysis of g will yield an additional $b_3(X)$ scalars in four dimensions. This is because the superpartners of C should come from g as these fields are superpartners in eleven dimensions. We should also add that (up to duality transformations) all representations of the $\mathcal{N} = 1$ supersymmetry algebra which contain one massless real scalar actually contain two scalars in total which combine into complex scalars. We will now describe how these scalars arise explicitly.

We began with a G_2-holonomy metric $g(X)$ on X. $g(X)$ obeys the vacuum Einstein equations,

$$R_{ij}(g(X)) = 0. \tag{2.21}$$

To obtain the spectrum of modes originating from g we look for fluctuations in $g(X)$ which also satisfy the vacuum Einstein equations. We take the fluctuations in $g(X)$ to depend on the four-dimensional coordinates y in Minkowski space. Writing the fluctuating metric as

$$g_{ij}(x) + \delta g_{ij}(x, y) \tag{2.22}$$

and expanding to first order in the fluctuation yields the Lichnerowicz equation

$$\Delta_L \delta g_{ij} \equiv -\nabla_M^2 \delta g_{ij} - 2R_{ijmn}\delta g^{mn} + 2R^k_{(i}\delta g_{j)k} = 0. \tag{2.23}$$

Next we make a Kaluza-Klein ansatz for the fluctuations as

$$\delta g_{ij} = h_{ij}(x)\rho(y). \tag{2.24}$$

Note that the term ∇^2 is the square of the full $d = 11$ covariant derivative. If we separate this term into two terms:

$$\nabla_M^2 = \nabla_\mu^2 + \nabla_i^2, \tag{2.25}$$

then we see that the fluctuations are scalar fields in four dimensions with squared masses given by the eigenvalues of the Lichnerowicz operator acting on the h_{ij}:

$$h_{ij}\nabla_\mu^2 \rho(y) = -(\Delta_L h_{ij})\rho(y) = -\lambda h_{ij}\rho(y). \tag{2.26}$$

Thus, zero modes of the Lichnerowicz operator give rise to massless scalar fields in four dimensions. We will now show that we have precisely $b_3(X)$ such zero modes.

On a 7-manifold of $SO(7)$ holonomy, the h_{ij} - being symmetric 2-index tensors - transform in the **27**-dimensional representation. Under G_2 this representation remains irreducible. On the other hand, the 3-forms on a G_2-manifold, which are usually in the **35** of $SO(7)$, decompose under G_2 as

$$\mathbf{35} \longrightarrow \mathbf{1} + \mathbf{7} + \mathbf{27}. \tag{2.27}$$

Thus, the h_{ij} can also be regarded as 3-forms on X. Explicitly,

$$\varphi_{n[pq}h^n_{r]} = \omega_{pqr}. \tag{2.28}$$

The ω's are 3-forms in the same representation as h_{ij} since φ is in the trivial representation. The condition that h is a zero mode of Δ_L is equivalent to ω being a zero mode of the Laplacian:

$$\Delta_L h = 0 \leftrightarrow \Delta\omega = 0. \tag{2.29}$$

This shows that there are precisely $b_3(X)$ additional massless scalar fields coming from the fluctuations of the G_2-holonomy metric on X.

As we mentioned above, these scalars combine with the ϕ's to give $b_3(X)$ massless scalars, $\Phi^I(y)$, which are the lowest components of massless chiral superfields in four dimensions. There is a very natural formula for the complex scalars $\Phi^I(y)$. Introduce a basis α_I for the third homology group of X, $H_3(X, \mathbf{R})$. This is a basis for the noncontractible holes in X of dimension three. We can choose the α_I so that

$$\int_{\alpha_I} \omega^J = \delta^J_I. \tag{2.30}$$

Since the fluctuating G_2-structure is

$$\varphi' = \varphi + \delta\varphi = \varphi + \Sigma_I \rho^I(y)\omega^I(x), \tag{2.31}$$

we learn that

$$\Phi^I(y) = \int_{\alpha_I} \varphi' + iC. \tag{2.32}$$

The fluctuations of the four-dimensional Minkowski metric give us the usual fluctuations of four-dimensional gravity, which due to supersymmetry implies that the four-dimensional theory is locally supersymmetric.

In addition to the massless chiral multiplets, we also get massless vector multiplets. The bosonic component of such a mulitplet is a massless Abelian gauge field which arises from the C-field through the Kaluza-Klein ansatz,

$$C = \Sigma_\alpha \;\; \beta^\alpha(x) \wedge A_\alpha(y) \tag{2.33}$$

where the β's are a basis for the harmonic 2-forms and the A's are one-forms in Minkowski space, i.e. Abelian gauge fields. Again, the equations of motion for C imply that the A's are massless in four dimensions. This gives $b_2(X)$ such gauge fields. As with the chiral multiplets above, the fermionic superpartners of the gauge fields arise from the gravitino field. Note that we could have also included an ansatz giving 2-forms in four dimensions by summing over harmonic 1-forms on X. However, since $b_1(X) = 0$, this does not produce any new massless fields in four dimensions.

We are now in a position to summarise the basic effective theory for the massless fields. The low energy effective theory is an $\mathcal{N}=1$ supergravity theory coupled to $b_2(X)$ Abelian vector multiplets and $b_3(X)$ massless, *neutral* chiral multiplets. This theory is relatively uninteresting physically. In particular, the gauge group is Abelian and there are no light charged particles. We will thus have to work harder to obtain the basic requisites of the standard model - non-Abelian gauge fields and chiral fermions - from G_2-compactifications. The basic point of these lectures is to emphasise that these features emerge naturally from singularities in G_2-manifolds.

3. Super Yang-Mills from G_2-manifolds: Classical

In this section we will describe how to obtain non-Abelian gauge groups from singular G_2-manifolds. We have known for some time now that non-Abelian gauge groups emerge from M theory when space has a so-called ADE-singularity. We learned this in the context of the duality between M theory on $K3$ and the heterotic string on a flat three-torus, $\mathbf{T}^3$ [**23**]. So, our basic strategy will be to embed ADE-singularities into G_2-manifolds. After reviewing the basic features of the duality between M theory on $K3$ and heterotic string theory on $\mathbf{T}^3$, we describe ADE-singularities explicitly. We will then describe the M theory physics near such a singularity. This is the physics of the McKay correspondence.

We then develop a picture of a G_2-manifold near an embedded ADE-singularity. Based on this picture we analyse what kinds of *four*-dimensional gauge theories these singularities give rise to. We then go on to describe local models for such singular G_2-manifolds as finite quotients of smooth ones.

3.1. *M* **theory - Heterotic Duality in Seven Dimensions.** M theory compactified on a $K3$ manifold is widely believed to be equivalent to the heterotic string theory compactified on a 3-torus $\mathbf{T}^3$. As with G_2 compactification, both of these are compactifications to flat Minkowski space. Up to diffeomorphisms, $K3$ is the only simply connected, compact 4-manifold admitting metrics of $SU(2)$-holonomy. $SU(2)$ is the analog in four dimensions of G_2 in seven dimensions. Interestingly enough, in this case $K3$ is the only simply connected example, whereas there are many G_2-manifolds.

There is a 58-dimensional moduli space of $SU(2)$-holonomy metrics on $K3$ manifolds of fixed volume. This space $\mathcal{M}(K3)$ is locally a coset space:

$$\mathcal{M}(K3)=\mathbb{R}^{+}\times\frac{SO(3,19)}{SO(3)\times SO(19)}. \tag{3.1}$$

An $SU(2)$ holonomy metric also admits two parallel spinors, which when tensored with the $\mathbf{8}$ constant spinors of 7-dimensional Minkowski space give 16 global supercharges. This corresponds to minimal supersymmetry in seven dimensions (in the same way that G_2-holonomy corresponds to minimal supersymmetry in four dimensions). If we work at a smooth point in $\mathcal{M}$ we can use Kaluza-Klein analysis and we learn immediately that the effective $d=7$ supergravity has 58 massless scalar fields which parameterise $\mathcal{M}$. These are the fluctuations of the metric on $K3$. Additionally, since $H^2(K3,\mathbb{R})\equiv\mathbb{R}^{22}$ there are twenty-two linearly independent classes of harmonic 2-forms. These may be used a la equation (32) to give a $U(1)^{22}$ gauge group in seven dimensions. We now go on to describe how this spectrum is the same as that of the heterotic string theory on $\mathbf{T}^3$, at generic points in $\mathcal{M}$.

The heterotic string in ten dimensions has a low energy description in terms of a supergravity theory whose massless bosonic fields are a metric, a 2-form B, a dilaton ϕ and non-Abelian gauge fields of structure group $SO(32)$ or $E_8 \times E_8$. There are sixteen global supersymetries. Compactification on a flat $\mathbf{T}^3$ preserves all supersymmetries which are all products of constant spinors on both $\mathbf{T}^3$ and Minkowski space. A flat metric on $\mathbf{T}^3$ involves six parameters so the metric gives rise to six massless scalars, and since there are three independent harmonic two forms we obtain from B three more. The condition for the gauge fields to be supersymmetric on $\mathbf{T}^3$ is that their field strengths vanish: these are so called flat connections. They are parameterised by Wilson lines around the three independent circles in $\mathbf{T}^3$. These are representations of the fundamental group of $\mathbf{T}^3$ in the gauge group. Most of the flat connections actually arise from Wilson loops which are actually in the maximal torus of the gauge group, which in this case is $U(1)^{16}$. Clearly, this gives a 48-dimensional moduli space giving 58 scalars altogether. Narain showed that this moduli space is actually also locally of the same form as $\mathcal{M}$ [**19**].

From the point of view of the heterotic string on $\mathbf{T}^3$, the effective gauge group in 7 dimensions (for generic metric and B-field) is the subgroup of $SO(32)$ or $E_8 \times E_8$ which commutes with the flat connection on $\mathbf{T}^3$. At generic points in the moduli space of flat connections, this gauge group will be $U(1)^{16}$. This is because the generic flat connection defines three generic elements in $U(1)^{16} \subset G$. We can think of these as diagonal 16 by 16 matrices with all elements on the diagonal non-zero. Clearly, only the diagonal elements of G will commute with these. So, at a generic point in moduli space the gauge group is Abelian.

Six more $U(1)$ gauge fields arise as follows from the metric and B-field. $\mathbf{T}^3$ has three harmonic one forms, so Kaluza-Klein reduction of B gives three gauge fields. Additionally, since $\mathbf{T}^3$ has a $U(1)^3$ group of isometries, the metric gives three more. In fact, the local actions for supergravity theories in seven dimensions are actually determined by the number of massless vectors. So, in summary, we have shown that at generic points in $\mathcal{M}$ the low energy supergravity theories arising from M theory on $K3$ or the heterotic string on $\mathbf{T}^3$ are the same.

At special points, some of the eigenvalues of the flat connections will vanish. At these points the unbroken gauge group can get enhanced to a non-Abelian group. This is none other than the Higgs mechanism: the Higgs fields are just the Wilson lines. Additionally, because seven-dimensional gauge theories are infrared trivial (the gauge coupling has dimension a positive power of length), the low energy quantum theory actually has a non-Abelian gauge symmetry.

If M theory on $K3$ is actually equivalent to the heterotic string in seven dimensions, it too should therefore exihibit non-Abelian symmetry enhancement at special points in the moduli space. These points are precisely the points in moduli space where the $K3$ develops orbifold singularities. We will not provide a detailed proof of this statement, but will instead look at the $K3$ moduli space in a neighbourhood of this singularity, where all the interesting behaviour of the theory is occurring. So, the first question is: what do these orbifold singularities look like?

3.1.1. **ADE-singularities.** An orbifold singularity in a Riemannian 4-manifold can locally be described as $\mathbb{R}^4/\Gamma$, where Γ is a finite subgroup of $SO(4)$. For generic enough Γ, the only singular point of this orbifold is the origin. These are the points in $\mathbb{R}^4$ left invariant under Γ. A very crucial point is that on the heterotic side, supersymmetry is completely unbroken all over the moduli space, so our orbifold singularities in $K3$ should also preserve supersymmetry. This means that Γ is a finite subgroup of $SU(2) \subset SO(4)$. The particular $SU(2)$ can easily be identified as follows. Choose some set of complex coordinates so that $\mathbb{C}^2 \equiv \mathbb{R}^4$. Then, a point in $\mathbb{C}^2$ is labelled by a 2-component vector. The $SU(2)$ in question acts on this vector in the standard way:

$$\begin{pmatrix} u \\ v \end{pmatrix} \longrightarrow \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix}. \tag{3.2}$$

The finite subgroups of $SU(2)$ have a classification which may be described in terms of the simply laced semi-simple Lie algebras: A_n, D_k, E_6, E_7 and E_8. There are two infinite series corresponding to $SU(n+1) = \mathsf{A}_n$ and $SO(2k) = \mathsf{D}_k$ and three exceptional subgroups corresponding to the three exceptional Lie groups of E-type. The subgroups, which we will denote by Γ_{A_n}, Γ_{D_k}, Γ_{E_i} can be described explicitly.

$\Gamma_{A_{n-1}}$ is isomorphic to $\mathbb{Z}_n$ - the cyclic group of order n - and is generated by

$$\begin{pmatrix} e^{\frac{2\pi i}{n}} & 0 \\ 0 & e^{\frac{-2\pi i}{n}} \end{pmatrix}. \tag{3.3}$$

Γ_{D_k} is isomorphic to $\mathbb{D}_{k-2}$ - the binary dihedral group of order $4k-8$ - and has two generators α and β given by

$$\alpha = \begin{pmatrix} e^{\frac{\pi i}{k-2}} & 0 \\ 0 & e^{\frac{-\pi i}{k-2}} \end{pmatrix} \quad \beta = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}. \tag{3.4}$$

Γ_{E_6} is isomorphic to $\mathbb{T}$ - the binary tetrahedral group of order 24 - and has two generators given by

$$\begin{pmatrix} e^{\frac{\pi i}{2}} & 0 \\ 0 & e^{\frac{-\pi i}{2}} \end{pmatrix} \quad and \quad \frac{1}{\sqrt{2}} \begin{pmatrix} e^{\frac{2\pi i7}{8}} & e^{\frac{2\pi i7}{8}} \\ e^{\frac{2\pi i5}{8}} & e^{\frac{2\pi i}{8}} \end{pmatrix} \tag{3.5}$$

Γ_{E_7} is isomorphic to $\mathbb{O}$ - the binary octahedral group of order 48 - and has three generators. Two of these are the generators of $\mathbf{T}$ and the third is

$$\begin{pmatrix} e^{\frac{2\pi i}{8}} & 0 \\ 0 & e^{\frac{2\pi i7}{8}} \end{pmatrix}. \tag{3.6}$$

Finally, Γ_{E_8} is isomorphic to $\mathbb{I}$ - the binary icosahedral group of order 120 - and has two generators given by

$$-\begin{pmatrix} e^{\frac{2\pi i3}{5}} & 0 \\ 0 & e^{\frac{2\pi i2}{5}} \end{pmatrix} \quad and \quad \frac{1}{e^{\frac{2\pi i2}{5}} - e^{\frac{2\pi i3}{5}}} \begin{pmatrix} e^{\frac{2\pi i}{5}} + e^{\frac{-2\pi i}{5}} & 1 \\ 1 & -e^{\frac{2\pi i}{5}} - e^{\frac{-2\pi i}{5}} \end{pmatrix} \tag{3.7}$$

Since all the physics of interest is happening near the orbifold singularities of $K3$, we can replace the $K3$ by $\mathbb{C}^2/\Gamma_{\mathbb{ADE}}$ and study the physics of M theory on $\mathbb{C}^2/\Gamma_{\mathbb{ADE}} \times \mathbb{R}^{6,1}$ near its singular set which is just $\mathbf{0} \times \mathbb{R}^{6,1}$. Since the $K3$ went from smooth to singular as we varied its moduli we expect that the singular orbifolds $\mathbb{C}^2/\Gamma_{\mathbb{ADE}}$ are singular limits of *non-compact* smooth 4-manifolds X^{ADE}. Because of supersymmetry, these should have $SU(2)$-holonomy. This is indeed the case. The metrics of $SU(2)$-holonomy on the X^{ADE} are known as ALE-spaces, since they asymptote to the locally Euclidean metric on $\mathbb{C}^2/\Gamma_{\mathbb{ADE}}$. Their existence was proven by Kronheimer [**18**] - who constructed a gauge theory whose Higgs branch is precisely the $\mathbb{C}^2/\Gamma_{\mathbb{ADE}}$ with its $SU(2)$-holonomy (or hyper-Kähler) metric.

A physical description of this gauge theory arises in string theory. Consider Type IIA or IIB string theory on $\mathbb{C}^2/\Gamma_{\mathbb{ADE}}\times\mathbb{R}^{5,1}$. Take a flat D$p$-brane (with $p \leq 5$) whose world-volume directions span $\mathbb{R}^{p,1} \subset \mathbb{R}^{5,1}$, i.e. the D-brane is sitting at a point on the orbifold. Then the world-volume gauge theory, which was first derived in [**12**], is given by the Kronheimer gauge theory. This theory has eight supersymmetries, which implies that its Higgs branch is a hyper-Kähler manifold. For one D-brane this theory has a gauge group which is a product of unitary groups of ranks given by the Dynkin indices (or dual Kac labels) of affine Dynkin diagram of the corresponding ADE-group. So, for the A_n-case the gauge group is $U(1)^{n+1}$. The matter content is also given by the affine Dynkin diagram - each link between a pair of nodes represents a hypermultiplet transforming in the bifundamental representation of the two unitary groups. This is an example of a quiver gauge theory - a gauge theory determined by a quiver diagram.

We will make this explicit in the simplest case of Γ_{A_1}. Γ_{A_1} is isomorphic to $\mathbb{Z}_2$ and is in fact the center of $SU(2)$. Its generator acts on $\mathbb{C}^2$ as

$$\begin{pmatrix} u \\ v \end{pmatrix} \longrightarrow \begin{pmatrix} -u \\ -v \end{pmatrix}. \tag{3.8}$$

In this case, the Kronheimer gauge theory has a gauge group which is $U(1)$ and has two fields Φ_1 and Φ_2. These are hypermultiplets in the string theory realisation on a D-brane. The hypermultiplets each contain two complex scalars (a_i, b_i) and the a's transform with charge $+1$ under $U(1)$, whilst the b's transform with charge -1.

The potential energy of these scalar fields on the D-brane is

$$V = |\vec{D}|^2 \equiv |\vec{\mu}| \tag{3.9}$$

where the three D-fields $\vec{D}$ (which are also known as the hyper-Kähler moment maps $\vec{\mu}$ associated with the $U(1)$ action on the $\mathbb{C}^4$ parameterised by the fields) are given by

$$D_1 = |a_1|^2 + |b_1|^2 - |a_2|^2 - |b_2|^2 \tag{3.10}$$

and

$$D_2 + iD_3 = a_1 b_1 - a_2 b_2. \tag{3.11}$$

The space of zero energy minima of V is the space of supersymmetric ground states S of the theory on the brane;

$$S = \{\vec{D} = 0\}/U(1). \tag{3.12}$$

In supersymmetric field theories, instead of solving these equations directly, it is equivalent to simply construct the space of gauge invariant holomorphic polynomials of the fields and impose only the holomorphic equation above (this is the classical F-term).

In the case at hand the gauge invariant polynomials are simply

$$X = a_1 b_1, \;\; Y = a_2 b_2, \;\; Z = a_1 b_2, \;\; W = a_2 b_1. \tag{3.13}$$

These obviously parameterise $\mathbb{C}^4$ but are subject to the relation

$$XY = WZ \tag{3.14}$$

However, the complex D-term equation asserts that

$$X = Y \tag{3.15}$$

hence

$$X^2 = WZ. \tag{3.16}$$

The space of solutions is precisely a copy of $\mathbb{C}^2/\Gamma_{A_1}$. To see this, we can parameterise $\mathbb{C}^2/\Gamma_{A_1}$ algebraically in terms of the Γ_{A_1} invariant coordinates on $\mathbb{C}^2$. These are u^2, v^2 and uv. If we denote these three coordinates as w, z, x, then obviously

$$x^2 = wz. \tag{3.17}$$

We prefer to re-write this equation by changing coordinates again. Defining $x = u^2 - v^2$, $y = iu^2 + iv^2$ and $z = 2uv$ gives a map from $\mathbb{C}^2/\Gamma_{A_1}$ to $\mathbb{C}^3$. Clearly however,

$$x^2 + y^2 + z^2 = 0 \tag{3.18}$$

which means that $\mathbb{C}^2/\Gamma_{A_1}$ is the hypersurface in $\mathbb{C}^3$ defined by this equation.

The orbifold can be deformed by adding a small constant to the right hand side,

$$x^2 + y^2 + z^2 = r^2. \tag{3.19}$$

If we take x, y and z to all be real and r to be real then it is clear that the deformed 4-manifold contains a 2-sphere of radius r. This 2-sphere contracts to zero size as r goes to zero. The total space of the deformed 4-manifold is in fact the cotangent bundle of the 2-sphere, $\mathbf{T}^*\mathbf{S}^2$. To see this write the real parts of the x, y and z as x_i and their imaginary parts as p_i. Then, since r is real, the x_i are coordinates on the sphere which obey the relation

$$\Sigma_i x_i p_i = 0. \tag{3.20}$$

This means that the p_i's parameterise tangential directions. The radius r sphere in the center is then the zero section of the tangent bundle. Since the manifold is actually complex it is natural to think of this as the cotangent bundle of the Riemann sphere, $\mathbf{T}^*\mathbf{CP}^1$. In the context of Euclidean quantum gravity, Eguchi and Hanson constructed a metric of $SU(2)$-holonomy on this space, asymptotic to the locally flat metric on $\mathbb{C}^2/\Gamma_{A_1}$.

In the Kronheimer gauge theory on the D-brane, deforming the singularity corresponds to setting the D-terms or moment maps not to zero but to constants. On the D-brane these constants represent the coupling of the background closed string fields to the brane. These fields parameterise precisely the metric moduli of the Eguchi-Hanson metric.

3.1.2. *M* **theory Physics at the Singularity.** This metric, whose precise form we will not require, actually has three parameters which control the size and shape of the two-sphere which desingularises the orbifold - the three possibilities for setting the moment maps to constants . From a distance it looks as though there is an orbifold singularity, but as one looks more closely one sees that the singularity has been smoothed out by a two-sphere. The 2-sphere is dual to a compactly supported harmonic 2-form, α. Thus, Kaluza-Klein reducing the C-field using α gives a $U(1)$ gauge field in seven dimensions. A vector multiplet in seven dimensions contains precisely one gauge field and three scalars and the latter are the parameters of the $\mathbf{S}^2$. So, when $\mathbf{T}^*\mathbf{CP}^1$ is smooth the massless spectrum is an Abelian vector multiplet.

From the duality with the heterotic string we expect to see an enhancement in the gauge symmetry when we vary the scalars to zero i.e. when the sphere shrinks to zero size. In order for this to occur, $W^{\pm}$-bosons must become massless at the singularity. These are electrically charged under the $U(1)$ gauge field which originated from C. From the eleven-dimensional point of view the object which is charged under C is the $M2$-brane. If the $M2$-brane wraps around the two-sphere, it appears as a particle from the seven-dimensional point of view. This particle is electrically charged under the $U(1)$ and has a mass which is classically given by the volume of the sphere. Since the $M2$-brane has tension its dynamics will push it to wrap the smallest volume two-sphere in the space. This least mass configuration is in fact invariant under half of the supersymmetries [3] - a fact which means that it lives in a short representation of the supersymmetry algebra. This in turn means that its classical mass is in fact uncorrected quantum mechanically. The $M2$-brane wrapped around this cycle with the opposing orientation has the opposing $U(1)$ charge to the previous one.

Thus, when the two-sphere shrinks to zero size we find that two oppositely charged BPS multiplets become massless. These have precisely the right quantum numbers to enhance the gauge symmetry from $U(1)$ to $\mathsf{A}_1 = SU(2)$. Super Yang-Mills theory in seven dimensions depends only on its gauge group. In this case we are asserting that in the absence of gravity, the low energy physics of M theory on $\mathbb{C}^2/\Gamma_{A_1}\times\mathbb{R}^{6,1}$ is described by super Yang-Mills theory on $\mathbf{0}\times\mathbb{R}^{6,1}$ with gauge group A_1.

The obvious generalisation also applies: in the absence of gravity, the low energy physics of M theory on $\mathbb{C}^2/\Gamma_{\mathbb{ADE}}\times\mathbb{R}^{6,1}$ is described by super Yang-Mills theory on $\mathbf{0}\times\mathbb{R}^{6,1}$ with ADE gauge group. To see this, note that the smoothing out of the orbifold singularity in $\mathbb{C}^2/\Gamma_{\mathbb{ADE}}$ contains $rank(ADE)$ two-spheres which intersect according to the Cartan matrix of the ADE group. At smooth points in the moduli space the gauge group is thus $U(1)^{rank(ADE)}$. The corresponding wrapped membranes give rise to massive BPS multiplets with precisely the masses and quantum numbers required to enhance the gauge symmetry to the full ADE-group at the origin of the moduli space.

Mathematically, the quantum numbers of wrapped branes are exactly the simple roots of the ADE-Lie algebra. This is our first encounter with the physics of the McKay correspondence.

3.2. ADE-singularities in G_2-manifolds. We have thus far restricted our attention to the ADE singularities in $K3\times\mathbb{R}^{6,1}$. However, the ADE singularity is a much more local concept. We can consider more complicated spacetimes $X^{10,1}$ with ADE singularities along more general seven-dimensional spacetimes, $Y^{6,1}$. Then, if X has a modulus which allows us to scale up the volume of Y, the large volume limit is a semi-classical limit in which X approaches the previous maximally symmetric situation discussed above. Thus, for large enough volumes we can assert that

[3] This is because the least volume two-sphere is an example of a calibrated or supersymmetric cycle.

the description of the classical physics of M theory near Y is in terms of seven-dimensional super Yang-Mills theory on Y - again with gauge group determined by which ADE singularity lives along Y.

In the context of G_2-compactification on $X\times\mathbb{R}^{3,1}$, we want Y to be of the form $W\times\mathbb{R}^{3,1}$, with W the locus of ADE singularities inside X. Near $W\times\mathbb{R}^{3,1}$, $X\times\mathbb{R}^{3,1}$ looks like $\mathbb{C}^2/\Gamma_{\mathbb{ADE}}\times W\times\mathbb{R}^{3,1}$. In order to study the gauge theory dynamics without gravity, we can again focus on the physics near the singularity itself. So, we want to focus on seven-dimensional super Yang-Mills theory on $W\times\mathbb{R}^{3,1}$.

3.2.1. *M* **theory Spectrum Near The Singularity.** In flat space the super Yang-Mills theory has a global symmetry group which is $SO(3)\times SO(6,1)$. The second factor is the Lorentz group, the first is the R-symmetry. The theory has gauge fields transforming as $(\mathbf{1},\mathbf{7})$, scalars in the $(\mathbf{3},\mathbf{1})$ and fermions in the $(\mathbf{2},\mathbf{8})$ of the universal cover. All fields transform in the adjoint representation of the gauge group. Moreover the sixteen supersymmetries also transform as $(\mathbf{2},\mathbf{8})$.

On $W\times\mathbb{R}^{3,1}$ - with an arbitrary W - the symmetry group gets broken to $SO(3)\times SO(3)'\times SO(3,1)$. Since $SO(3)'$ is the structure group of the tangent bundle on W, covariance requires that the theory is coupled to a background $SO(3)'$ gauge field - the spin connection on W. Similarly, though perhaps less intuitively, $SO(3)$ acts on the normal bundle to W inside X, hence there is a background $SO(3)$ gauge field also.

The supersymmetries transform as $(\mathbf{2},\mathbf{2},\mathbf{2})+(\mathbf{2},\mathbf{2},\bar{\mathbf{2}})$. For large enough W and at energy scales below the inverse size of W, we can describe the physics in terms of a *four*-dimensional gauge theory. But this theory as we have described it is not supersymmetric as this requires that we have covariantly constant spinors on W. Because W is curved, there are none. However, we actually want to consider the case in which W is embedded inside a G_2-manifold X. In other words we require that our local model - $\mathbb{C}^2/\Gamma_{\mathbb{ADE}}\times W$ - admits a G_2-holonomy metric. When W is curved this metric cannot be the product of the locally flat metric on $\mathbb{C}^2/\Gamma_{\mathbb{ADE}}$ and a metric on W. Instead the metric is warped and is more like the metric on a fiber bundle in which the metric on $\mathbb{C}^2$ varies as we move around in W. Since the space has G_2-holonomy we should expect the four-dimensional gauge theory to be supersymmetric. We will now demonstrate that this is indeed the case by examining the G_2-structure more closely. In order to do this however, we need to examine the $SU(2)$ structure on $\mathbb{C}^2/\Gamma$ as well.

4-dimensional spaces of $SU(2)$-holonomy are actually examples of hyperKähler manifolds. They admit three parallel 2-forms ω_i. These are analogous to the parallel forms on G_2-manifolds. These three forms transform locally under $SO(3)$ which locally rotates the complex structures. On $\mathbb{C}^2$ these forms can be given explicitly as

$$\omega_1+i\omega_2 = du\wedge dv, \tag{3.21}$$

$$\omega_3 = \frac{i}{2}du\wedge d\bar{u}+dv\wedge d\bar{v}. \tag{3.22}$$

$\Gamma_{\mathbb{ADE}}$ is defined so that it preserves all three of these forms. The $SO(3)$ which rotates these three forms is identified with the $SO(3)$ factor in our seven-dimensional gauge theory picture. This is because the moduli space of $SU(2)$-holonomy metrics is the moduli space of the gauge theory and this has an action of $SO(3)$.

In a locally flat frame we can write down a formula for the G_2-structure on $\mathbb{C}^2/\Gamma_{\mathbb{ADE}}\times W$,

$$\varphi = \frac{1}{6}\omega_i \wedge e_j\delta^{ij} + e_1 \wedge e_2 \wedge e_3 \tag{3.23}$$

where e_i are a flat frame on W. Note that this formula is manifestly invariant under the $SO(3)$ which rotates the w_i *provided* that it also acts on the e_i in the same way.

The key point is that when the $SO(3)$ of the gauge theory acts, in order for the G_2-structure to be well defined the e_i's must transform in precisely the same way as the ω_i. But $SO(3)'$ acts on the e_i, because it is the structure group of the tangent bundle to W. Therefore, if $\mathbb{C}^2/\Gamma_{\mathbb{ADE}}\times W$, admits a G_2-holonomy metric, we must identify $SO(3)$ with $SO(3)'$. In other words, the connection on the tangent bundle is identified with the connection on the normal $SO(3)'$ bundle. This breaks the symmetries to the diagonal subgroup of the two $SO(3)$'s and implies that the effective four-dimensional field theory is classically supersymmetric. Identifying the two groups breaks the symmetry group down to $SO(3)''\times SO(3,1)$ under which the supercharges transform as $(\mathbf{1},\mathbf{2}) + (\mathbf{3},\mathbf{2}) + cc$. We now have supersymmetries since the $(\mathbf{1},\mathbf{2})$ and its conjugate can be taken to be constants on W.

An important point, which we will not actually prove here, but will require in the sequel is that the locus of ADE-singularities - namely the copy of W at the center of $\mathbb{C}^2/\Gamma$ is actually a supersymmetric cycle (also known as a calibrated cycle). This follows essentially from the fact that $\Gamma_{\mathbb{ADE}}$ fixes W and therefore the φ restricts to be the volume form on W. This is the condition for W to be supersymmetric.

Supposing we could find a G_2-manifold of this type, what exactly is the four-dimensional supersymmetric gauge theory it corresponds to? This we can answer also by Kaluza-Klein analysis [**1, 2**], since W will be assumed to be smooth and 'large'. Under $SO(3)''\times SO(3,1)$, the seven-dimensional gauge fields transform as $(\mathbf{3},\mathbf{1}) + (\mathbf{1},\mathbf{4})$, the three scalars give $(\mathbf{3},\mathbf{1})$ and the fermions give $(\mathbf{1},\mathbf{2}) + (\mathbf{3},\mathbf{2}) + cc$. Thus the fields which are scalars under the four-dimensional Lorentz group are two copies of the $\mathbf{3}$ of $SO(3)''$. These may be interpreted as two one forms on W. These will be massless if they are zero modes of the Laplacian on W (w.r.t. its induced metric from the G_2-manifold). There will be precisely $b_1(W)$ of these i.e. one for every harmonic one-form. Their superpartners are clearly the $(\mathbf{3},\mathbf{2}) + cc$ fermions, which will be massless by supersymmetry. This is precisely the field content of $b_1(W)$ chiral supermultiplets of the supersymmetry algebra in four dimensions.

The $(\mathbf{1},\mathbf{4})$ field is massless if it is constant on W and this gives one gauge field in four dimensions. The requisite superpartners are the remaining fermions which transform as $(\mathbf{1},\mathbf{2}) + cc$.

All of these fields transform in the adjoint representation of the seven dimensional gauge group. Thus the final answer for the massless fields is that they are described by $\mathcal{N} = 1$ super Yang-Mills theory with $b_1(W)$ massless adjoint chiral supermultiplets. The case with pure "superglue" i.e. $b_1(W) = 0$ is a particularly interesting gauge theory at the quantum level: in the infrared the theory is believed to confine colour, undergo chiral symmetry breaking and have a mass gap. We will actually exhibit some of these very interesting properties semi-classically in M theory! Much of the sequel will be devoted to explaining this. But before we can do that we must first describe concrete examples of G_2-manifolds with the properties we desire.

One idea is to simply look for *smooth* G_2-manifolds which are topologically $\mathbb{C}^2 \times W$ but admit an action of $SU(2)$ which leaves W invariant but acts on $\mathbb{C}^2$ in the natural way. Then we simply pick a $\Gamma_{\mathbb{ADE}} \subset SU(2)$ and form the quotient space $\mathbb{C}^2/\Gamma_{\mathbb{ADE}} \times W$.

Luckily, such non-compact G_2-manifolds were constructed some time ago [**9**]! Moreover, in these examples, $W = S^3$, the simplest possible compact 3-manifold with $b_1(W) = 0$. Perfect.

3.2.2. **Examples of G_2-manifolds with ADE-singularities.** The G_2 holonomy metrics on $\mathbb{C}^2 \times S^3$ were constructed in [**9**]. These metrics are smooth and complete and depend on one parameter a. There is a radial coordinate r and the metric shows that there is a finite size S^3 (of size a) in the center which grows as we move out along the r direction. They asymptote at infinity in a radial coordinate r to a conical form

$$ds^2 = dr^2 + r^2 d\Sigma^2 \tag{3.24}$$

where $d\Sigma^2$ is an Einstein metric on $S^3 \times S^3$. Taking r large is equivalent to taking a to zero, so the finite volume S^3 shrinks to zero size. This Einstein metric $d\Sigma^2$ is not the standard product metric on the product of two spheres, although it is the homogeneous metric on G/H with $G = SU(2)^3$ and $H = SU(2)$. H acts on G as the 'diagonal' subgroup of the three $SU(2)$'s. G/H is obviously isomorphic as a manifold to $S^3 \times S^3$ since S^3 is isomorphic to $SU(2)$. This description of the conical G_2-metric obviously has an asymptotic $SU(2)^3 \times \Sigma_3$ group of isometries with Σ_3 the group of permutations of the three $SU(2)$ factors.

The conical metric is obviously incomplete, since the base of the cone goes to zero size at $r = 0$. The complete G_2 metrics can thus be regarded as completions of the cones obtained by smoothing out the singularity at its apex. Topologically the conical manifold is $\mathbb{R}^+ \times S^3 \times S^3$, which gets smoothed out to $\mathbb{R}^4 \times S^3$. Concretely this amounts to choosing an $SU(2)$ factor in G and 'filling it in' to form $\mathbb{R}^4$. We remind the reader that $\mathbb{R}^4 - 0$ is the same as $\mathbb{R}^+ \times S^3$ and filling in the origin gives back $\mathbb{R}^4$. In the case at hand we actually complete the cone by gluing in an S^3. This is the S^3 of size a.

Clearly there are three natural ways to carry out this procedure since G consists of three copies of $SU(2)$. Obviously each of these gives the same topological manifold but it is very important for what follows that we realise that there are actually three G_2 manifolds that we can make in this way. The point is that the classical moduli space of G_2-holonomy metrics consists of three real lines which intersect at one point - the conical singularity. Moving off of the conical manifold in the three different directions amounts to choosing an S^3 in G and filling it in. Along these three directions three different S^3's develop a finite volume. Another way to say more or less the same thing in a perhaps more physical way is that there are three smooth G_2-manifolds with the prescribed behaviour at infinity: the metric on G/H. The transition between the three-manifolds via the conical singularity is called a 'flop' - by analogy with what happens to spheres in complex manifolds.

We can give a simple algebraic description of the phenomenon of collapsing one sphere and growing another with the following model taken from [**6**]. Consider the hypersurface in $\mathbb{R}^4 \times \mathbb{R}^4$ cut out by the following equation

$$\Sigma_i x_i^2 - y_i^2 = a \tag{3.25}$$

where the x's and y' s are linear coordinates on the two $\mathbb{R}^4$'s. For a positive, we have a radius a S^3 at the origin in the second $\mathbb{R}^4$. This manifold is clearly topologically $S^3 \times \mathbb{R}^4$. For a negative, its again the same manifold topologically, but the roles of x and y have been interchanged. Therefore as a passes from positive to negative an S^3 shrinks to zero volume, the manifold becomes singular at 0 and *another* S^3 grows and the space remains smooth. This is obviously a much cruder description of the space as a function of a, since as we have seen above there are actually three directions in the moduli space and not two, but it has the advantage that it makes the basic picture transparent.

3.2.3. **M theory Physics on $X = \mathbb{R}^4 \times S^3$.** We saw earlier that the moduli of the G_2-metric get complexified in M theory by the addition of the C-field. This is necessary for supersymmetry. We observed that on a compact G_2-manifold the low energy four-dimensional theory contains one massless scalar for every parameter in the G_2-metric. The situation on a non-compact manifold X can in general be quite different, since the metric fluctuations need not be localised on X. The more delocalised these fields are, the more difficult it is to excite them. Indeed to obtain a four-dimensional action we have to integrate over X, and this integral will diverge if the fluctuations are not L^2-normalisable. If this is indeed the case then we should not regard the corresponding four-dimensional fields as fluctuating: rather they are background parameters, coupling constants, and we should study the four-dimensional physics as a function of them.

In the case at hand, by examining the first order fluctation in the G_2-metric one can readily see that a should indeed be treated as a coupling constant. It follows from supersymmetry that its complex partner should also. We refer the reader to [**7**] for the simple calculation which shows this explicitly. Our formula (32) can now be applied to write this complex coupling constant as

$$\tau = \int_{S^3} \varphi + iC = Vol(S^3) + i \int C \tag{3.26}$$

where we integrate over the minimal volume three-sphere in the center of X. This sphere generates the third homology. Note that there is no prime here, since the field is not fluctuating.

So, we arrive at the conclusion that M theory on our G_2-manifold X is actually a one complex dimensional family of theories parameterised by τ. There are three semi-classical regimes corresponding to the three regions in which the spheres are large, X is smooth and thus supergravity is valid. The four-dimensional spectrum is massive in each of these semi-classical regimes since $b_2(X)$ is zero and the zero mode of the Lichnerowicz operator does not fluctuate. The physics in each of these three regions is clearly the same since the metrics are the 'same'. What about the physics on the bulk of the τ-plane?

Our intuition asserts that there is only one further interesting point, $\tau = 0$ where the manifold develops a conical singularity. In minimally supersymmetric field theories in four dimensions which do have massless complex scalars which parameterise a moduli space M, singularities in the physics typically only occur at subloci in M which are also complex manifolds. In our case, we don't have a moduli space, but rather a parameter space, but we can think of τ as a background superfield.

In any case, at $\tau = 0$ we have zero size S^3's and instanton effects can become important here, since the action of an $M2$-brane instanton is τ. These effects could generate a non-zero quantum value of the C-field period and remove this potential singularity, in which case we would be in the situation that there are no physical phase transitions as a function of τ. Of course, physical quantities will depend on τ, but the qualitative nature of the physics will remain the same for any value of τ. This was first suggested in [**6**] and proven rigorously in [**7**].

The parameter space spanned by τ is thus a Riemann sphere with three special points which correspond to the three large G_2-manifolds. This is the picture Robbert Dijkgraaf alluded to in his lectures. To make contact with what he was discussing we need to understand the S^1 quotients of our three G_2-manifolds. This is because if the M theory 11-manifold Y is "fibered" by circles then the base 10-manifold Q can be regarded as a spacetime in Type IIA string theory. The IIA theory has a 1-form field A which is a connection on this circle bundle.

The quotient by S^1 of our three G_2-manifolds are basically the three geometries Robbert was describing - the resolved conifold, its flop and the deformed conifold with open strings!

To see this we define the S^1 action on the first big G_2-manifold to be the standard action on the S^3 in $\mathbb{R}^4 \times S^3$ regarded as a Hopf fibration over $\mathbb{CP}^1$. The quotient is thus an $\mathbb{R}^4$-bundle over $\mathbb{CP}^1$. Because A is now topologically non-trivial the metric one gets on this 6-manifold is not the Calabi-Yau metric on the resolved conifold. However, it is a metric with the same symplectic structure! Thus, the closed string Gromov-Witten invariants are the same as on the Calabi-Yau conifold. The S^1 quotient of G_2-manifold number two gives the 'same' 6-manifold in Type IIA - but now this is better regarded as the flop of the previous one.

Finally, S^1-quotient number three acts on the $\mathbb{R}^4$ and not the S^3. The quotient by $U(1)$ of $\mathbb{R}^4$ is $\mathbb{R}^3$. The quotient of the G_2-manifold is thus an $\mathbb{R}^3$-bundle over S^3. This action has a fixed point which is the S^3 of minimal volume in the middle of the G_2 space. The fact that there is a fixed point of the circle action means in Type IIA language that A is singular along a codimension three submanifold in the 10-dimensional spacetime. This is thus a 7-manifold which supports a Dirac monopole. This is a D6-brane in string theory and is a place where open strings can end. We thus learn that the third quotient gives Type IIA string theory on the deformed conifold plus one D6-brane wrapping the S^3.

Therefore M theory explains why the open string Gromov-Witten invariants of the deformed conifold are exactly the same as the usual closed string Gromov-Witten invariants of the resolved conifold. This is simply the statement that all three Type IIA geometries are the same G_2-manifold in M theory metrically.

In M theory these invariants should be regarded as counting associative 3-cycles in a G_2-manifold.

4. Super Yang-Mills from G_2-manifolds: Quantum.

We now move on to study the interplay between quantum super Yang-Mills theory and M theory on G_2-manifolds. The results of this section are based upon [**2, 6, 7, 3**]. We will be studying the physics of M theory on the G_2-manifolds with ADE-singularities whose construction we described at the end of section 3.2.1. We begin by reviewing the basic properties of super Yang-Mills theory. We then go on to describe how these features are reflected in M theory. We first show

how membrane instantons can be seen to generate the superpotential of the theory. Second, we go on to exhibit confinement and a mass gap semi-classically in M theory. The result about confinement highlights the power of the physical McKay correspondence.

The G_2-manifolds that actually interest us are obtained as quotients of $\mathbb{R}^4 \times S^3$ by $\Gamma_{\mathbb{ADE}}$. We saw previously that, for large volume and low energies, four-dimensional super Yang-Mills theory is a good description of the M theory physics. We will thus begin this section with a review of the basic properties of the gauge theory.

4.1. Super Yang-Mills Theory. For completeness and in order to compare easily with M theory results obtained later we briefly give a review of $\mathcal{N}=1$ pure super Yang-Mills theory. We begin with gauge group $SU(n)$. $\mathcal{N}=1$ $SU(n)$ super Yang-Mills theory in four dimensions is an extensively studied quantum field theory. The classical Lagrangian for the theory is

$$\mathcal{L} = -\frac{1}{4g^2}(F^a_{\mu\nu})^2 + \frac{1}{g^2}\bar{\lambda}^a i\not{D}\lambda^a + i\frac{\theta}{32\pi^2}F^a_{\mu\nu}\tilde{F}^{a\mu\nu}. \tag{4.1}$$

F is the gauge field strength and λ is the gaugino field.

It is widely believed that this theory exhibits dynamics very similar to that of ordinary QCD: confinement, chiral symmetry breaking, a mass gap. There are n supersymmetric vacua. Supersymmetry constrains the dynamics of the theory so strongly that the values of the low energy effective superpotential in the n vacua are known. These are of the form

$$W_{eff} = c\mu^3 e^{2\pi i\tau/n}. \tag{4.2}$$

Here τ is the complex coupling constant,

$$\tau = \frac{\theta}{2\pi} + i\frac{4\pi}{g^2} \tag{4.3}$$

and μ the mass scale. Shifting θ by 2π gives n different values for W.

In particular, the form of this potential suggests that it is generated by dynamics associated with "fractional instantons", i.e. instantonic objects in the theory whose quantum numbers are formally of instanton number $\frac{1}{n}$. Such states are also closely related to the spontaneously broken chiral symmetry of the theory. Let us briefly also review some of these issues here.

Under the $U(1)R$-symmetry of the theory, the gauginos transform as

$$\lambda \rightarrow e^{i\alpha}\lambda \tag{4.4}$$

This is a symmetry of the classical action but not of the quantum theory (as can easily be seen by considering the transformation of the fermion determinant in the path integral). However, if the above transformation is *combined* with a shift in the theta angle of the form

$$\tau \rightarrow \tau + \frac{2n}{2\pi}\alpha. \tag{4.5}$$

then this cancels the change in the path integral measure. This shift symmetry is a bona fide symmetry of the physics if $\alpha = \frac{2\pi}{2n}$, so that even in the quantum theory a $\mathbb{Z}_{2n}$ symmetry remains. Associated with this symmetry is the presence of a non-zero value for the following correlation function,

$$\langle\lambda\lambda(x_1)\lambda\lambda(x_2)...\lambda\lambda(x_n)\rangle \tag{4.6}$$

which is clearly invariant under the $\mathbb{Z}_{2n}$ symmetry. This correlation function is generated in the 1-instanton sector and the fact that $2n$ gauginos enter is due to the fact that an instanton of charge 1 generates $2n$ chiral fermion zero modes.

Cluster decomposition implies that the above correlation function decomposes into 'n constituents' and therefore there exists a non-zero value for the gaugino condensate:

$$\langle\lambda\lambda\rangle\neq 0. \tag{4.7}$$

Such a non-zero expectation value is only invariant under a $\mathbb{Z}_2$ subgroup of $\mathbb{Z}_{2n}$ implying that the discrete chiral symmetry has been spontaneously broken. Consequently this implies the existence of n vacua in the theory.

In fact, it can be shown that

$$\langle\lambda\lambda\rangle = 16\pi i\frac{\partial}{\partial\tau}W_{eff} = -\frac{32\pi^2}{n}c\mu^3 e^{2\pi i\tau/n}. \tag{4.8}$$

In view of the above facts it is certainly tempting to propose that 'fractional instantons' generate the non-zero gaugino condensate directly. But this is difficult to see directly in super Yang-Mills on $\mathbb{R}^{3,1}$. We will return to this point later.

More generally, if we replace the $SU(n)$ gauge group by some other gauge group H, then the above statements are also correct but with n replaced *everywhere* with $c_2(H)$, the dual Coxeter number of H. For ADE gauge groups $c_2(H)$ $= \Sigma_{i=1}^{r+1} a_i$, where r is the rank of the gauge group and the a_i are the Dynkin indices of the affine Dynkin diagram associated to H. For A_n, all the $a_i = 1$; for D_n groups the four 'outer' nodes have index 1 whilst the rest have $a_i = 2$. E_6 has indices $(1,1,1,2,2,2,3)$, E_7 has $(1,1,2,2,2,3,3,4)$ whilst E_8 has indices $(1,2,2,3,3,4,4,5,6)$.

4.2. Theta angle and Coupling Constant in M theory. The physics of M theory supported near the singularities of $\mathbb{C}^2/\Gamma\times R^{6,1}$ is described by super Yang-Mills theory on $\mathbb{R}^{6,1}$. The gauge coupling constant of the theory is given by

$$\frac{1}{g_{7d}^2} \sim \frac{1}{l_p^3} \tag{4.9}$$

where l_p is the eleven-dimensional Planck length. In seven dimensions, one analog of the theta angle in four dimensions is actually a three-form Θ. The reason for this is the seven-dimensional interaction

$$L_I \sim \Theta\wedge F\wedge F \tag{4.10}$$

(with F the Yang-Mills field strength). In M theory Θ is given by C, the three-form potential for the theory.

If we now take M theory on our G_2-manifold $\mathbb{R}^4/\Gamma\times W$ we have essentially compactified the seven-dimensional theory on W and the four-dimensional gauge coupling constant is roughly given by

$$\frac{1}{g_{4d}^2} \sim \frac{V_W}{l_p^3} \tag{4.11}$$

where V_W is the volume of W. The four-dimensional theta angle can be identified as

$$\theta = \int_W C. \tag{4.12}$$

The above equation is correct because under a global gauge transformation of C which shifts the above period by 2π times an integer - a transformation which is a symmetry of M theory - θ changes by 2π times an integer. Such shifts in the theta angle are also global symmetries of the field theory.

Thus the complex gauge coupling constant of the effective four-dimensional theory may be identified as the τ parameter of M theory

$$\tau = \int_W i\frac{C}{2\pi} + \frac{\varphi}{l_p^3}. \tag{4.13}$$

This is of course entirely natural, since τ is the only parameter in M theory on this space!

4.3. Superpotential in M theory. There is a very elegant way to calculate the superpotential of super Yang-Mills theory on $\mathbb{R}^{3,1}$ by first compactifying it on a circle to three dimensions [**21**]. The three-dimensional theory has a perturbative expansion since the Wilson lines on the circle behave as Higgs fields whose vev's break the gauge symmetry to the maximal torus. The theory has a perturbative expansion in the Higgs vevs, which can be used to compute the superpotential of the compactified theory. One then takes the four-dimensional limit. In order to compute the field theory superpotential we will mimic this idea in M theory [**1**]. Compactifying the theory on a small circle is equivalent to studying perturbative Type IIA string theory on our G_2-manifold.

4.3.1. **Type IIA theory on $X = \mathbb{R}^4/\Gamma_{\mathbb{ADE}}\times S^3$.** Consider Type IIA string theory compactified to three dimensions on a seven-manifold X with holonomy G_2. If X is smooth we can determine the massless spectrum of the effective supergravity theory in three dimensions as follows. Compactification on X preserves four of the 32 supersymmetries in ten dimensions, so the supergravity theory has three-dimensional $\mathcal{N} = 2$ local supersymmetry. The relevant bosonic fields of the ten-dimensional supergravity theory are the metric, B -field, dilaton plus the Ramond-Ramond one- and three-forms. These we will denote by g, B, ϕ, A_1, A_3 respectively. Upon Kaluza-Klein reduction the metric gives rise to a three-metric and $b_3(X)$ massless scalars. The latter parameterise the moduli space of G_2 -holonomy metrics on X. B gives rise to $b_2(X)$ periodic scalars φ_i. ϕ gives a three-dimensional dilaton. A_1 reduces to a massless vector, while A_3 gives $b_2(X)$ vectors and $b_3(X)$ massless scalars. In three dimensions a vector is dual to a periodic scalar, so at a point in moduli space where the vectors are free we can dualise them. The dual of the vector field originating from A_1 is the period of the RR 7-form on X, whereas the duals of the vector fields coming from A_3 are given by the periods of the RR 5-form A_5 over a basis of 5-cycles which span the fifth homology group of X. Denote these scalars by σ_i. All in all, in the dualised theory we have, in addition to the supergravity multiplet, $b_2(X) + b_3(X)$ scalar multiplets. Notice that $b_2(X)$ of the scalar multiplets contain two real scalar fields, both of which are periodic.

Now we come to studying the Type IIA theory on $X = \mathbb{R}^4/\Gamma_{\mathbb{ADE}}\times S^3$. Recall that $X = \mathbb{R}^4/\Gamma_{\mathbb{ADE}}\times S^3$ is defined as an orbifold of the standard spin bundle of S^3, denoted by $S(S^3)$. To determine the massless spectrum of IIA string theory on X we can use standard orbifold techniques. However, the answer can be phrased in a simple way. X is topologically $\mathbb{R}^4/\Gamma_{\mathbb{ADE}}\times S^3$. This manifold can be desingularised to give a smooth seven-manifold $M^{\Gamma_{\mathbb{ADE}}}$ which is topologically $X^{\Gamma_{\mathbb{ADE}}}\times S^3$, where $X^{\Gamma_{\mathbb{ADE}}}$ is homeomorphic to an ALE space. The string theoretic cohomology groups of X are isomorphic to the usual cohomology groups of $M^{\Gamma_{\mathbb{ADE}}}$. The reason for this is simple: X is a global orbifold of $S(S^3)$. The string theoretic cohomology groups count massless string states in the orbifold CFT. The massless string states in the twisted sectors are localised near the fixed points of the action of $\Gamma_{\mathbb{ADE}}$ on the spin bundle. Near the fixed points we can work on the tangent space of $S(S^3)$ near a fixed point and the action of $\Gamma_{\mathbb{ADE}}$ there is just its natural action on $\mathbb{R}^4 \times \mathbb{R}^3$.

Note that blowing up X to give $M^{\Gamma_{\mathbb{ADE}}}$ cannot give a metric with G_2 -holonomy which is continuosly connected to the singular G_2 -holonomy metric on X, since this would require that the addition to homology in passing from X to $M^{\Gamma_{\mathbb{ADE}}}$ receives contributions from four-cycles. This is necessary since these are dual to elements of $H^3(M)$ which generate metric deformations preserving the G_2 -structure. This argument does not rule out the possibility that $M^{\Gamma_{\mathbb{ADE}}}$ admits 'disconnected' G_2 -holonomy metrics, but is consistent with the fact that pure super Yang-Mills theory in four dimensions does not have a Coulomb branch.

The important points to note are that the twisted sectors contain massless states consisting of r scalars and r vectors where r is the rank of the corresponding ADE group associated to Γ. The r scalars can intuitively be thought of as the periods of the B field through r two-cycles. In fact, for a generic point in the moduli space of the orbifold conformal field theory the spectrum contains massive particles charged under the r twisted vectors. These can be interpreted as wrapped D2-branes whose quantum numbers are precisely those of W -bosons associated with the breaking of an ADE gauge group to $U(1)^r$. This confirms our interpretation of the origin of this model from M theory: the values of the r B -field scalars can be interpreted as the expectation values of Wilson lines around the eleventh dimension associated with this symmetry breaking. At weak string coupling and large S^3 volume these states are very massive and the extreme low energy effective dynamics of the twisted sector states is described by $\mathcal{N}=2$ $U(1)^r$ super Yang-Mills in three dimensions. Clearly however, the underlying conformal field theory is not valid when the W -bosons become massless. The appropriate description is then the pure super Yang-Mills theory on $\mathbb{R}^{2,1} \times S^1$ which corresponds to a sector of M theory on $X \times S^1$. In this section however, our strategy will be to work at a generic point in the CFT moduli space which corresponds to being far out along the Coulomb branch of the gauge theory. We will attempt to calculate the superpotential there and then continue the result to four dimensions. This exactly mimics the strategy of [**21**] in field theory. Note that we are implicitly ignoring gravity here. More precisely, we are assuming that, in the absence of gravitational interactions with the twisted sector, the low energy physics of the twisted sectors of the CFT is described by the Coulomb branch of the gauge theory. This is natural since the twisted sector states are localised at the singularities of $X \times \mathbb{R}^{2,1}$ whereas the gravity propagates in bulk.

In this approximation, we can dualise the photons to obtain a theory of r chiral multiplets, each of whose bosonic components ($\boldsymbol{\varphi}$ and $\boldsymbol{\sigma}$) is periodic. But remembering that this theory arose from a non-Abelian one we learn that the moduli space of classical vacua is

$$\mathcal{M}_{cl} = \frac{\mathbb{C}^r}{\Lambda_W^{\mathbb{C}} \rtimes W_g} \tag{4.14}$$

where $\Lambda_W^{\mathbb{C}}$ is the complexified weight lattice of the ADE group and W_g is the Weyl group.

We can now ask about quantum effects. In particular, is there a non-trivial superpotential for these chiral multiplets? In a theory with four supercharges BPS instantons with only two chiral fermion zero modes can generate a superpotential. Are there instantons in Type IIA theory on J? BPS instantons come from branes wrapping supersymmetric cycles and Type IIA theory on a G_2 -holonomy space can have instantons corresponding to D6-branes wrapping the space itself or D2-brane instantons which wrap supersymmetric 3-cycles. For smooth G_2 -holonomy manifolds these were studied in [**14**]. In the case at hand the D6-branes would generate a superpotential for the dual of the graviphoton multiplet which lives in the gravity multiplet but, since we wish to ignore gravitational physics for the moment, we will ignore these. In any case, since X is non-compact, these configurations have infinite action. The D2-branes on the other hand are much more interesting. They can wrap the supersymmetric S^3 over which the singularities of X are fibered. We can describe the dynamics of a wrapped D2-brane as follows. At large volume, where the sphere becomes flatter and flatter the world-volume action is just the Kronheimer gauge theory that we described previously in section 3.1.1! Here we should describe this theory not just on S^3 but on a supersymmetric (or calibrated) S^3 embedded in a space with a non-trivial G_2 -holonomy structure. The upshot is that the world-volume theory is in fact a cohomological field theory [**8**] so we can trust it for any volume as long as the ambient space has G_2 -holonomy. This is because the supersymmetries on the world-volume are actually scalars on S^3 and so must square to zero. This "topological version" of the Kronheimer theory can be naturally obtained by "twisting" the usual gauge theory in the same way that one obtains the topological A-model and B-model by twisting and was described by Robbert Dijkgraaf in his lectures. In this case, the twisting is naturally incorporated by the normal bundle of the calibrated three-sphere.

Note that, since we are ignoring gravity, we are implicitly ignoring higher derivative corrections which could potentially also affect this claim. Another crucial point is that the S^3 which sits at the origin in $\mathbb{R}^4$ in the covering space of X is the supersymmetric cycle, and the spheres away from the origin are not supersymmetric, so that the BPS wrapped D2-brane is constrained to live on the singularities of X. In the quiver gauge theory, the origin is precisely the locus in moduli space at which the single D2-brane can fractionate (according to the quiver diagram) and this occurs by giving expectation values to the scalar fields which parameterise the Coulomb branch which corresponds to the position of our D2-brane in the dimensions normal to X.

What contribution to the superpotential do the fractional D2-branes make? To answer this we need to identify the configurations which possess only two fermionic zero modes. We will not give a precise string theory argument for this, but using the correspondence between this string theory and field theory will identify exactly which D-brane instantons we think are responsible for generating the superpotential. This may sound like a strong assumption, but, as we hope will become clear, the fact that the fractional D2-branes are wrapped D4-branes is actually anticipated by the field theory! This makes this assumption, in our opinion, somewhat weaker and adds credence to the overall picture being presented here.

In [**11**], it was shown that the fractionally charged D2-branes are actually D4-branes which wrap the 'vanishing' 2-cycles at the origin in $\mathbb{R}^4/\Gamma$. More precisely, each individual fractional D2-brane which originates from a single D2-brane possesses D4-brane charge, but the total configuration, since it began life as a single D2-brane, has zero D4-brane charge. The possible contributions to the superpotential are constrained by supersymmetry and must be given by a holomorphic function of the r chiral superfields and also of the holomorphic gauge coupling constant τ which corresponds to the complexified volume of the S^3 in eleven-dimensional M theory. We have identified above the bosonic components of the chiral superfields above. τ is given by

$$\tau = \int \varphi + iC \tag{4.15}$$

where φ is the G_2 -structure defining 3-form on X. The period of the M theory 3-form through S^3 plays the role of the theta angle.

The world-volume action of a D4-brane contains the couplings

$$L = B \wedge A_3 + A_5. \tag{4.16}$$

Holomorphy dictates that there is also a term

$$B \wedge \varphi \tag{4.17}$$

so that the combined terms are written as

$$B \wedge \tau + A_5 \tag{4.18}$$

Since the fourbranes wrap the 'vanishing cycles' and the S^3 we see that the contribution of the D4-brane corresponding to the k-th fractional D2 takes the form

$$S = -\boldsymbol{\beta}_k . \boldsymbol{z} \tag{4.19}$$

where we have defined

$$\boldsymbol{z} \ = \tau \boldsymbol{\varphi} \ + \boldsymbol{\sigma} \tag{4.20}$$

and the $\boldsymbol{\beta_k}$ are charge vectors. The r complex fields $\boldsymbol{z}$ are the natural holomorphic functions upon which the superpotential will depend.

The wrapped D4-branes are the magnetic duals of the massive D2-branes which we identified above as massive W -bosons. As such they are magnetic monopoles for the original ADE gauge symmetry. Their charges are therefore given by an element of the co-root lattice of the Lie algebra and thus each of the $r+1$ $\boldsymbol{\beta}$'s is a rank r vector in this space. Choosing a basis for this space corresponds to choosing a basis for the massless states in the twisted sector Hilbert space which intuitively we can think of as a basis for the cohomology groups Poincaré dual to the 'vanishing' 2-cycles. A natural basis is provided by the simple co-roots of the Lie algebra of ADE, which we denote by $\boldsymbol{\alpha}^*_k$ for $k = 1, ..., r$. This choice is natural, since these, from the field theory point of view, are the fundamental monopole charges.

At this point it is useful to mention that the r wrapped D4-branes whose magnetic charges are given by the simple co-roots of the Lie algebra correspond in field theory to monopoles with charges $\boldsymbol{\alpha}^*_k$ and each of these is known to possess precisely the right number of zero modes to contribute to the superpotential. Since we have argued that, in a limit of the Type IIA theory on X, the dynamics at low energies is governed by the field theory studied in [**21**] it is natural to expect that these wrapped fourbranes also contribute to the superpotential. Another striking feature of the field theory is that these monopoles also possess a fractional instanton number - the second Chern number of the gauge field on $\mathbb{R}^{2,1}\times S^1$. These are precisely in correspondence with the fractional D2-brane charges. Thus, in this sense, the field theory anticipates that fractional branes are wrapped branes.

In the field theory on $\mathbb{R}^{2,1}\times S^1$ it is also important to realise that there is precisely one additional BPS state which contributes to the superpotential. The key point is that this state, unlike the previously discussed monopoles, has dependence on the periodic direction in spacetime. This state is associated with the affine node of the Dynkin diagram. Its monopole charge is given by

$$-\Sigma_{k=1}^{r}\boldsymbol{\alpha}_k^*. \tag{4.21}$$

and it also carries one unit of instanton number.

The action for this state is

$$S = \Sigma_{k=1}^{r}\boldsymbol{\alpha}_k^*.\boldsymbol{z} - 2\pi i\tau \tag{4.22}$$

Together, these $r+1$ BPS states can be regarded as fundamental in the sense that all the other finite action BPS configurations can be thought of as bound states of them.

Thus, in the correspondence with string theory it is also natural in the same sense as alluded to above that a state with these corresponding quantum numbers also contributes to the superpotential. It may be regarded as a bound state of anti-D4-branes with a charge one D2-brane. In the case of $SU(n)$ this is extremely natural, since the total D4/D2-brane charge of the $r+1$ states is zero/one, and this is precisely the charge of the D2-brane configuration on S^3 whose world-volume action is the quiver gauge theory for the affine Dynkin diagram for $SU(n)$. In other words, the entire superpotential is generated by a single D2-brane which has fractionated.

In summary, we have seen that the correspondence between the Type IIA string theory on X and the super Yang-Mills theory on $\mathbb{R}^{2,1}\times S^1$ is quite striking. Within the context of this correspondence we considered a smooth point in the moduli space of the perturbative Type IIA CFT, where the spectrum matches that of the Yang-Mills theory along its Coulomb branch. On the string theory side we concluded that the possible instanton contributions to the superpotential are from wrapped D2-branes. Their world volume theory is essentially topological, from which we concluded that they can fractionate. As is well known, the fractional D2-branes are really wrapped fourbranes. In the correspondence with field theory, the wrapped fourbranes are magnetic monopoles, whereas the D2-branes are instantons. Thus, if these branes generate a superpotential they correspond in field theory to monopole-instantons. This is exactly how the field theory potential is known to be generated. We thus expect that the same occurs in the string theory on X.

Finally, the superpotential generated by these instantons is of affine-Toda type and is known to possess $c_2(\mathsf{ADE})$ minima corresponding to the vacua of the ADE super Yang-Mills theory on $\mathbb{R}^{3,1}$. The value of the superpotential in each of these vacua is of the form $e^{\frac{2\pi i\tau}{c_2}}$. As such it formally looks as though it was generated by fractional instantons, and in this context fractional M 2-brane instantons. This result holds in the four-dimensional M theory limit because of holomorphy.

Let us demonstrate the vacuum structure in the simple case when the gauge group is $SU(2)$. Then there is only one scalar field, z. There are two fractional

D2-brane instantons whose actions are

$$S_1 = -z \text{ and } S_2 = z - 2\pi i\tau \tag{4.23}$$

Both of these contribute to the superpotential as

$$W = e^{-S_1} + e^{-S_2}. \tag{4.24}$$

Defining $z = ln\, Y$ we have

$$W = Y + \frac{e^{2\pi i\tau}}{Y}. \tag{4.25}$$

The critical points of W are

$$Y = \pm e^{\frac{2\pi i\tau}{2}}. \tag{4.26}$$

This result about the superpotential suggests strongly that there is a limit of M theory near an ADE singularity in a G_2 -manifold which is precisely super Yang-Mills theory. We will now go on to explore other limits of this M theory background.

4.4. M theory Physics on ADE-singular G_2-manifolds. We saw previously that before taking the quotient by Γ, the M theory physics on $\mathbb{R}^4 \times S^3$ with its G_2-metric was smoothly varying as a function of τ. In fact the same is true in the case with ADE-singularities. One hint for this was that we explicitly saw just now that the superpotential is non-zero in the various vacua and this implies that the C-field is non-zero. This suggestion was concretely proven in [**7**].

Before orbifolding by Γ we saw there were three semiclassical limits of M theory in the space parameterised by τ. These were described by M theory on three large and smooth G_2-manifolds X_i, all three of which were of the form $\mathbb{R}^4 \times S^3$. There are also three semiclassical i.e. large volume G_2-manifolds when we orbifold by Γ. These are simply the quotients by Γ of the X_i. One of these is the G_2-manifold $\mathbb{R}^4/\Gamma_{\mathbb{ADE}} \times S^3$. The other two are both of the form $S^3/\Gamma_{\mathbb{ADE}} \times \mathbb{R}^4$. To see this, note that the three S^3's in the three G_2-manifolds X_i of the form $\mathbb{R}^4 \times S^3$ correspond to the three S^3 factors in $G = S^3 \times S^3 \times S^3$. $\Gamma_{\mathbb{ADE}}$ is a subgroup of one of these S^3's. If $\Gamma_{\mathbb{ADE}}$ acts on the $\mathbb{R}^4$ factor of X_1 in the standard way, then it must act on S^3 in X_2 - since X_2 can be thought of as the same manifold but with the two S^3's at infinity exchanged. Then, because of the permutation symmetry it also acts on the S^3 in X_3.

In the simple, crude, algebraic description in section 3.2.2, let X_1 be the manifold with a negative. Then define $\Gamma_{\mathbb{ADE}}$ to act on the $\mathbb{R}^4$ parameterised by x_i. Then $x_i = 0$ is an S^3 of fixed points parameterised by y_i. Thus $X_1/\Gamma_{\mathbb{ADE}}$ is isomorphic to our G_2-manifold with ADE-singularities. Consider now what happens when a is taken positive. This is our manifold X_2. Then, because $x_i = 0$ is not a point on X_2, there is no fixed point and $\Gamma_{\mathbb{ADE}}$ acts freely on the S^3 surrounding the origin in the x-space. Thus, $X_2/\Gamma_{\mathbb{ADE}}$ is isomorphic to $S^3/\Gamma_{\mathbb{ADE}} \times \mathbb{R}^4$, as is X_3.

On $X_1/\Gamma_{\mathbb{ADE}}$ in the large volume limit, we have a semi-classical description of the four-dimensional physics in terms of perturbative super Yang-Mills theory. But, at extremely low energies, this theory becomes strongly coupled, and is believed to confine and get a mass gap. So, apart from calculating the superpotential in each vacuum, as we did in section 4.3, we can't actually calculate the spectrum here.

What about the physics in the other two semiclassical limits, namely large $X_{2,3}/\Gamma_{\mathbb{ADE}}$? These G_2-manifolds are completely smooth. So supergravity is a good approximation to the M theory physics. What do we learn about the M theory physics in this approximation?

4.5. Confinement from G_2-manifolds. If it is true that the qualitative physics of M theory on $X_2/\Gamma_{\mathbb{ADE}}$ and $X_3/\Gamma_{\mathbb{ADE}}$ is the same as that of M theory on $X_1/\Gamma_{\mathbb{ADE}}$, then some of the properties of super Yang-Mills theory at low energies ought to be visible. The gauge theory is believed to confine ADE-charge at low energies. If a gauge theory confines in four dimensions, electrically charged confining flux tubes (confining strings) should be present. This is because the confining potential is linear. If the classical fields of the gauge theory contain only fields in the adjoint representation of the gauge group G, then these strings are charged with respect to the center of G, $Z(G)$. All other charges are screened by quantum fluctuations. Can we see these strings in M theory on $X_2/\Gamma_{\mathbb{ADE}}$? Indeed, the confining strings have a very simple description [**3**].

The natural candidates for such strings are $M2$-branes which wrap around 1-cycles in $X_2/\Gamma_{\mathbb{ADE}}$ or $M5$-branes which wrap 4-cycles in $X_2/\Gamma_{\mathbb{ADE}}$. Since $X_2/\Gamma_{\mathbb{ADE}}$ is homeomorphic to $S^3/\Gamma_{\mathbb{ADE}}\times\mathbb{R}^4$ which is contractible to $S^3/\Gamma_{\mathbb{ADE}}$, the homology groups of $X_2/\Gamma_{\mathbb{ADE}}$ are the same as those of the three-manifold $S^3/\Gamma_{\mathbb{ADE}}$. Thus, our space has no four-cycles to speak of, so the confining strings can only come from $M2$-branes wrapping one-cycles in $S^3/\Gamma_{\mathbb{ADE}}$. The string charges are classified by the first homology group $H_1(S^3/\Gamma_{\mathbb{ADE}},\mathbb{Z})$. For any manifold, the first homology group is isomorphic to the Abelianisation of its fundamental group, Π_1. The Abelianisation is obtained by setting all commutators in Π_1 to be trivial, i.e.

$$H_1(M,\mathbb{Z}) \cong \frac{\Pi_1(M)}{[\Pi_1(M),\Pi_1(M)]}. \tag{4.27}$$

The fundamental group of $S^3/\Gamma_{\mathbb{ADE}}$ is $\Gamma_{\mathbb{ADE}}$. Hence, in order to calculate the charges of our candidate confining strings we need to compute the Abelianisations of all of the finite subgroups of $SU(2)$.

$\Gamma_{A_{n-1}} \cong \mathbb{Z}_n$. The gauge group is locally $SU(n)$. Since $\mathbb{Z}_n$ is Abelian, its commutator subgroup is trivial and hence the charges of our strings take values in $\mathbb{Z}_n$. Since this is isomorphic to the center of $SU(n)$ this is the expected answer for a confining $SU(n)$ theory.

For $\Gamma \cong \mathbb{D}_{k-2}$, the binary dihedral group of order $4k-8$, the local gauge group of the Yang-Mills theory is $SO(2k)$. The binary dihedral group is generated by two elements α and β (see section 3.1.1) which obey the relations

$$\alpha^2 = \beta^{k-2} \tag{4.28}$$

$$\alpha\beta = \beta^{-1}\alpha \tag{4.29}$$

$$\alpha^4 = \beta^{2k-4} = 1. \tag{4.30}$$

To compute the Abelianisation of $\mathbb{D}_{k-2}$, we simply take these relations and impose that the commutators are trivial. From the second relation this implies that

$$\beta = \beta^{-1} \tag{4.31}$$

which in turn implies that

$$\alpha^2 = 1 \text{ for } k = 2p \tag{4.32}$$

and

$$\alpha^2 = \beta \text{ for } k = 2p + 1. \tag{4.33}$$

Thus, for $k = 2p$ we learn that the Abelianisation of $\mathbb{D}_{k-2}$ is isomorphic to $\mathbb{Z}_2 \times \mathbb{Z}_2$, whereas for $k = 2p + 1$ it is isomorphic to $\mathbb{Z}_4$. These groups are respectively the centers of $Spin(4p)$ and $Spin(4p+2)$. This is the expected answer for the confining strings in $SO(2k)$ super Yang-Mills which can be coupled to spinorial charges.

To compute the Abelianisations of the binary tetrahedral (denoted $\mathbb{T}$), octahedral ($\mathbb{O}$) and icosahedral ($\mathbb{I}$) groups which correspond respectively to E_6, E_7 and E_8 super Yang-Mills theory, we utilise the fact that the order of $F/[F, F]$ - with F a finite group - is the number of inequivalent one-dimensional representations of F. The representation theory of the finite subgroups of $SU(2)$ is described through the McKay correspondence by the representation theory of the corresponding Lie algebras. In particular the dimensions of the irreducible representations of $\mathbb{T}, \mathbb{O}$ and $\mathbb{I}$ are given by the coroot integers (or dual Kac labels) of the affine Lie algebras associated to E_6, E_7 or E_8 respectively. From this we learn that the respective orders of $\mathbb{T}/[\mathbb{T}, \mathbb{T}]$, $\mathbb{O}/[\mathbb{O}, \mathbb{O}]$ and $\mathbb{I}/[\mathbb{I}, \mathbb{I}]$ are three, two and one. Moreover, one can easily check that $\mathbb{T}/[\mathbb{T}, \mathbb{T}]$ and $\mathbb{O}/[\mathbb{O}, \mathbb{O}]$ are $\mathbb{Z}_3$ and $\mathbb{Z}_2$ respectively, by examining their group relations. Thus we learn that $\mathbb{T}/[\mathbb{T}, \mathbb{T}]$, $\mathbb{O}/[\mathbb{O}, \mathbb{O}]$ and $\mathbb{I}/[\mathbb{I}, \mathbb{I}]$ are, respectively isomorphic to the centers $Z(E_6)$, $Z(E_7)$ and $Z(E_8)$, in perfect agreement with the expectation that the super Yang-Mills theory confines. Note that the E_8-theory does not confine, since the strings are uncharged.

This result is also natural from the following point of view. In the singular $X_1/\Gamma_{\mathbb{ADE}}$ (where the actual gauge theory dynamics is) the gauge bosons correspond to $M2$-branes wrapped around zero-size cycles. When we vary τ away from the actual gauge theory limit until we reach M theory on a large and smooth $X_2/\Gamma_{\mathbb{ADE}}$ the confining strings are also wrapped $M2$-branes. In the gauge theory we expect the confining strings to be "built" from the excitations of the gauge fields themselves. In M theory, the central role played by the gauge fields is actually played by the $M2$-brane.

4.6. Mass Gap from G_2-manifolds. We can also see the mass gap expected of the gauge theory by studying the spectrum of M theory on the smooth G_2-manifolds $X_2/\Gamma_{\mathbb{ADE}}$ and $X_3/\Gamma_{\mathbb{ADE}}$. We already noted previously that the four-dimensional spectrum of M theory on the X_i was massive. For precisely the same reasons the spectrum of M theory on $X_{2,3}/\Gamma_{\mathbb{ADE}}$ is also massive.

5. Chiral Fermions from G_2-manifolds.

Thus far, we have seen that the simplest possible singularities of a G_2-manifold, namely ADE orbifold singularities, produce a convincing picture of how non-Abelian gauge groups emerge. However, for the purpose of obtaining a realistic model of particle physics this is not enough. To this end, a basic requisite is the presence of chiral fermions charged under these gauge symmetries. Chiral fermions are important in nature since they are massless as long as the gauge symmetries they are charged under are unbroken. This enables us to understand the lightness of the electron in terms of the Higgs vev.

What sort of singularity in a G_2-manifold X might we expect to give rise to a chiral fermion in M theory? If the singularity is "bigger than a point" then we don't expect chiral fermions. This is because if the codimension of the singularity is less than seven, the local structure of the singularity can actually be regarded as a singularity in a Calabi-Yau threefold or $K3$ and these singularities give rise to a spectrum of particles which form representations of $\mathcal{N} > 1$ supersymmetry. Such spectra are $\mathcal{C}PT$ self-conjugate. For instance, real codimension four singularities in G_2-manifolds are the ADE-singularities we discussed above and the corresponding four-dimensional spectra were not chiral. Similarly, if the singularity is codimension six, i.e. is along a one-dimensional curve Σ in X, then everywhere near Σ the tangent spaces of X naturally split into tangent and normal directions to X. Hence, the holonomy of X near Σ actually reduces to $SU(3)$ acting normally to Σ.

So we want to consider point-like singularities of X. The simplest such singularities are conical, for which the metric looks locally like

$$ds^2 = dr^2 + r^2 g(Y) \tag{5.1}$$

for some six-dimensional metric $g(Y)$ on a 6-manifold Y. This has a singularity at $r = 0$. We will argue that for many different choices of Y that chiral fermions are part of the M theory spectrum.

5.1. Hints from Anomaly Inflow. The basic strategy of this subsection will be to assume there is a G_2-manifold with a conical singularity of the above type and consider the variation of bulk terms in the M theory effective action under various gauge symmetries. These will be shown to be non-zero if Y obeys certain conditions. If the theory is to be consistent, these anomalous variations must be cancelled and this suggests the presence of chiral fermions in the spectrum. This is based upon [**24**] who showed that when X is compact all these variations add up to zero!

The gauge symmetries we will consider are the ones we have focussed on in these lectures: the $U(1)$ gauge symmetries from Kaluza-Klein reducing the C-field and the ADE symmetries from the ADE-singularities.

We begin with the former case. We take M theory on $X \times \mathbb{R}^{3,1}$ with X a cone on Y so that X with the vertex removed is $\mathbb{R} \times Y$. The Kaluza-Klein ansatz for C which gives gauge fields in four dimensions is

$$C = \Sigma_\alpha \, \beta^\alpha(x) \wedge A_\alpha(y) \tag{5.2}$$

where the β's are harmonic 2-forms on X. With this ansatz, consider the eleven-dimensional Chern-Simons interaction

$$S = \int_{X \times \mathbb{R}^{3,1}} \frac{1}{6} C \wedge G \wedge G. \tag{5.3}$$

Under a gauge transformation of C under which

$$C \longrightarrow C + d\epsilon \tag{5.4}$$

S changes by something of the form[4]

$$\delta S \sim \int_{X \times \mathbb{R}^{3,1}} d(\epsilon \wedge \; G \wedge G). \tag{5.5}$$

We can regard X as a manifold with boundary $\partial X = Y$ and hence

$$\delta S \sim \int_{Y \times \mathbb{R}^{3,1}} \epsilon \wedge \; G \wedge G. \tag{5.6}$$

[4]We will not be too careful about factors in this section.

If we now make the Kaluza-Klein ansatz for the 2-form ϵ

$$\epsilon = \Sigma_i \epsilon^\alpha \beta^\alpha \tag{5.7}$$

and use our ansatz for C, we find

$$\delta S \sim \int_Y \beta^\rho \wedge \beta^\sigma \wedge \beta^\delta \int_{\mathbb{R}^{3,1}} \epsilon^\rho dA^\sigma \wedge dA^\delta. \tag{5.8}$$

Thus if the integrals over Y (which are topological) are non-zero we obtain a non-zero four-dimensional interaction characteristic of an anomaly in an Abelian gauge theory. Thus, if the theory is to be consistent, it is natural to expect a spectrum of chiral fermions at the conical singularity which exactly cancels δS.

We now turn to non-Abelian gauge anomalies. We have seen that ADE gauge symmetries in M theory on a G_2-manifold X are supported along a three-manifold W in X. If additional conical singularities of X are to support chiral fermions charges under the ADE gauge group, then these singularities should surely also be points P_i on W. So let us assume that near such a point the metric on X assumes the conical form. In four-dimensional ADE gauge theories the triangle anomaly is only non-trivial for A_n-gauge groups. So we restrict ourselves to this case. In this situation, there is a seven-dimensional interaction of the form

$$S = \int_{W \times \mathbb{R}^{3,1}} K \wedge \Omega_5(A) \tag{5.9}$$

where A is the $SU(n)$ gauge field and

$$d\Omega_5(A) = trF \wedge F \wedge F. \tag{5.10}$$

K is a two-form which is the field strength of a $U(1)$ gauge field which is part of the normal bundle to W. K measures how the A_n-singularity twists around W. The $U(1)$ gauge group is the subgroup of $SU(2)$ which commutes with Γ_{A_n}.

Under a gauge transformation,

$$A \longrightarrow A + D_A \lambda \tag{5.11}$$

and

$$\delta S \sim \int_{W \times \mathbb{R}^{3,1}} (K \wedge dtr\lambda F \wedge F) \tag{5.12}$$

so if K is closed, $\delta S = 0$. This will be the case if the A_n-singularity is no worse at the conical singularity P than at any other point on W. If, however, the A_n-singularity actually increases rank at P, then

$$dK = 2\pi q \delta_P \tag{5.13}$$

and we have locally a Dirac monopole of charge q at P. The charge is an integer because of obvious quantisation conditions. In this situation we have that

$$\delta S \sim \int_{W \times \mathbb{R}^{3,1}} d(K \wedge tr\lambda F \wedge F) = -q \int_{\mathbb{R}^{3,1}} tr\, \lambda F \wedge F \tag{5.14}$$

which is precisely the triangle anomaly in an $SU(n)$ gauge theory. Thus, if we have this sort of situation in which the ADE-singularity along W degenerates further at P we also expect chiral fermions to be present.

We now go on to utilise the M theory heterotic duality of subsection (3.1) to construct explicitly conically singular manifolds for which we know the existence of chiral fermions.

5.2. Chiral Fermions via Duality With The Heterotic String. In section three we utilised duality with the heterotic string on $\mathbf{T}^3$ to learn about enhanced gauge symmetry in M theory. We applied this to G_2-manifolds quite successfully. In this section we will take a similar approach. The following is based upon [**5**].

We start by considering duality with the heterotic string. The heterotic string compactified on a Calabi-Yau three fold Z can readily give chiral fermions. On the other hand, most Calabi-Yau manifolds participate in mirror symmetry. For Z to participate in mirror symmetry means, according to Strominger, Yau and Zaslow, that, in a suitable limit of its moduli space, it is a $\mathbf{T}^3$ fibration (with singularities and monodromies) over a base W. Then, taking the $\mathbf{T}^3$'s to be small and using on each fiber the equivalence of the heterotic string on $\mathbf{T}^3$ with M theory on $K3$, it follows that the heterotic string on Z is dual to M theory on a seven-manifold X that is $K3$-fibered over W (again with singularities and monodromies). X depends on the gauge bundle on Z. Since the heterotic string on Z is supersymmetric, M theory on X is likewise supersymmetric, and hence X is a manifold of G_2 holonomy.

The heterotic string on Z will typically have a four-dimensional spectrum of chiral fermions. Since there are many Z's that could be used in this construction (with many possible classes of gauge bundles) it follows that there are many manifolds of G_2 holonomy with suitable singularities to give non-Abelian gauge symmetry with chiral fermions. The same conclusion can be reached using duality with Type IIA, as many six-dimensional Type IIA orientifolds that give chiral fermions are dual to M theory on a G_2 manifold [**10**]

Let us try to use this construction to determine what kind of singularity X must have. (The reasoning and the result are very similar to that given in [**17**] for engineering matter from Type II singularities. In [**17**] the Dirac equation is derived directly rather than being motivated – as we will – by using duality with the heterotic string.) Suppose that the heterotic string on Z has an unbroken gauge symmetry G, which we will suppose to be simply-laced (in other words, an A, D, or E group) and at level one. This means that each $K3$ fiber of X will have a singularity of type G. As one moves around in X one will get a family of G-singularities parameterized by W. If W is smooth and the normal space to W is a smoothly varying family of G-singularities, the low energy theory will be G gauge theory on $W \times \mathbb{R}^{3,1}$ without chiral multiplets. This was the situation studied in sections three and four. So chiral fermions will have to come from singularities of W or points where W passes through a worse-than-orbifold singularity of X.

We can use the duality with the heterotic string to determine what kind of singularities are required. The argument will probably be easier to follow if we begin with a specific example, so we will consider the case of the $\mathsf{E}_8 \times \mathsf{E}_8$ heterotic string with $G = SU(5)$ a subgroup of one of the E_8's. Such a model can very easily get chiral $\mathbf{5}$'s and $\mathbf{10}$'s of $SU(5)$; we want to see how this comes about, in the region of moduli space in which Z is $\mathbf{T}^3$-fibered over W with small fibers, and then we will translate this description to M theory on X.

Let us consider, for example, the $\mathbf{5}$'s. The commutant of $SU(5)$ in E_8 is a second copy of $SU(5)$, which we will denote as $SU(5)'$. Since $SU(5)$ is unbroken, the structure group of the gauge bundle E on Z reduces from E_8 to $SU(5)'$. Massless fermions in the heterotic string transform in the adjoint representation of E_8. The part of the adjoint representation of E_8 that transforms as $\mathbf{5}$ under $SU(5)$ transforms as **10** under $SU(5)'$. So to get massless chiral **5**'s of $SU(5)$, we must look at the Dirac equation $\mathcal{D}$ on Z with values in the **10** of $SU(5)'$; the zero modes of that Dirac equation will give us the massless **5**'s of the unbroken $SU(5)$.

We denote the generic radius of the $\mathbf{T}^3$ fibers as α, and we suppose that α is much less than the characteristic radius of W. This is the regime of validity of the argument for duality with M theory on X (and the analysis of mirror symmetry of SYZ). For small α, we can solve the Dirac equation on Z by first solving it along the fiber, and then along the base. In other words, we write $\mathcal{D} = \mathcal{D}_T + \mathcal{D}_W$, where $\mathcal{D}_T$ is the Dirac operator along the fiber and $\mathcal{D}_W$ is the Dirac operator along the base. The eigenvalue of $\mathcal{D}_T$ gives an effective "mass" term in the Dirac equation on W. For generic fibers of $Z \to W$, as we explain momentarily, the eigenvalues of $\mathcal{D}_T$ are all nonzero and of order $1/\alpha$. This is much too large to be canceled by the behavior of $\mathcal{D}_W$. So zero modes of $\mathcal{D}$ are localized near points in W above which $\mathcal{D}_T$ has a zero mode.

When restricted to a $\mathbf{T}^3$ fiber, the $SU(5)'$ bundle E can be described as a flat bundle with monodromies around the three directions in $\mathbf{T}^3$. In other words, as in section three, we have three Wilson lines on each fiber. For generic Wilson lines, every vector in the **10** of $SU(5)'$ undergoes non-trivial "twists" in going around some (or all) of the three directions in $\mathbf{T}^3$. When this is the case, the minimum eigenvalue of $\mathcal{D}_T$ is of order $1/\alpha$. This is simply because for a generic flat gauge field on the $\mathbf{T}^3$-fiber there will be no zero mode.

A zero mode of $\mathcal{D}_T$ above some point $P \in W$ arises precisely if for some vector in the **10**, the monodromies in the fiber are all trivial.

This means that the monodromies lie in the subgroup of $SU(5)'$ that leaves fixed that vector. If we represent the **10** by an antisymmetric 5×5 matrix B^{ij}, $i, j = 1, \ldots, 5$, then the monodromy-invariant vector corresponds to an antisymmetric matrix B that has some nonzero matrix element, say B^{12}; the subgroup of $SU(5)'$ that leaves B invariant is clearly then a subgroup of $SU(2) \times SU(3)$ (where in these coordinates $SU(2)$ acts on $i, j = 1, 2$ and $SU(3)$ on $i, j = 3, 4, 5$). Let us consider the case (which we will soon show to be generic) in which B^{12} is the only nonzero matrix element of B. If so, the subgroup of $SU(5)'$ that leaves B fixed is precisely $SU(2) \times SU(3)$. There is actually a distinguished basis in this problem – the one that diagonalizes the monodromies near P – and it is in this basis that B has only one nonzero matrix element.

The commutant of $SU(2) \times SU(3)$ in E_8 is $SU(6)$. So, over the point P, the monodromies commute not just with $SU(5)$ but with $SU(6)$. Everything of interest will happen inside this $SU(6)$. The reason for this is that the monodromies at P give large masses to all E_8 modes except those in the adjoint of $SU(6)$. So we will formulate the rest of the discussion as if the heterotic string gauge group were just $SU(6)$, rather than E_8. Away from P, the monodromies break $SU(6)$ to $SU(5) \times U(1)$ (the global structure is actually $U(5)$). Restricting the discussion from E_8 to $SU(6)$ will mean treating the vacuum gauge bundle as a $U(1)$ bundle (the $U(1)$ being the second factor in $SU(5) \times U(1) \subset SU(6)$) rather than an $SU(5)'$ bundle.

The fact that, over P, the heterotic string has unbroken $SU(6)$ means that, in the M theory description, the fiber over P has an $SU(6)$ singularity. Likewise, the fact that away from P the heterotic string has only $SU(5) \times U(1)$ unbroken means that the generic fiber, in the M theory description, must contain an $SU(5)$

singularity only, rather than an $SU(6)$ singularity. As for the unbroken $U(1)$, in the M theory description it must be carried by the C-field. Indeed, over generic points on W there is a non-zero size S^2 which shrinks to zero size at P in order that the gauge symmetry at that point increases. Kaluza-Klein reducing C along this S^2 gives a $U(1)$.

If we move away from the point P in the base, the vector B in the **10** of $SU(5)'$ is no longer invariant under the monodromies. Under parallel transport around the three directions in $\mathbf{T}^3$, it is transformed by phases $e^{2\pi i\theta_j}$, $j = 1, 2, 3$. Thus, the three θ_j must all vanish to make B invariant. As W is three-dimensional, we should expect generically that the point P above which the monodromies are trivial is isolated. (Now we can see why it is natural to consider the case that, in the basis given by the monodromies near P, only one matrix element of B is nonzero. Otherwise, the monodromies could act separately on the different matrix elements, and it would be necessary to adjust more than three parameters to make B invariant. This would be a less generic situation.) We will only consider the (presumably generic) case that P is disjoint from the singularities of the fibration $Z \to W$. Thus, the $\mathbf{T}^3$ fiber over P is smooth (as we have implicitly assumed in introducing the monodromies on $\mathbf{T}^3$).

In [**5**] we explicitly solved the Dirac equation in a local model for this situation. We found that the net number of chiral zero modes was one. We will not have time to describe the details of the solution here.

In summary, before we translate into the M theory language, the chiral fermions in the heterotic string theory on Z are localised at points on W over which the Wilson lines in the $\mathbf{T}^3$-fibers are trivial. In M theory this translates into the statement that the chiral fermions are localised at points in W over which the ADE-singularity "worsens". This is also consistent with what we found in the previous section.

5.3. M theory Description. So we have found a local structure in the heterotic string that gives a net chirality – the number of massless left-handed **5**'s minus right-handed **5**'s – of one. Let us see in more detail what it corresponds to in terms of M-theory on a manifold of G_2 holonomy.

Here it may help to review the case considered in [**17**] where the goal was geometric engineering of charged matter on a Calabi-Yau threefold in Type IIA. What was considered there was a Calabi-Yau three fold R, fibered by $K3$'s with a base W', such that over a distinguished point $P \in W'$ there is a singularity of type $\hat{G}$, and over the generic point in W' this singularity is replaced by one of type G – the rank of $\hat{G}$ being one greater than that of G. In our example, $\hat{G} = SU(6)$ and $G = SU(5)$. In the application to Type IIA, although R also has a Kähler metric, the focus is on the complex structure. For $\hat{G} = SU(6)$, $G = SU(5)$, let us describe the complex structure of R near the singularities. The $SU(6)$ singularity is described by an equation $xy = z^6$ - cf. section three. Its "unfolding" depends on five complex parameters and can be written $zy = z^6 + P_4(z)$, where $P_4(z)$ is a quartic polynomial in z. If – as in the present problem – we want to deform the $SU(6)$ singularity while maintaining an $SU(5)$ singularity, then we must pick P_4 so that the polynomial $z^6 + P_4$ has a fifth order root. This determines the deformation to be

$$xy = (z + 5\epsilon)(z - \epsilon)^5, \tag{5.15}$$

where we interpret ϵ as a complex parameter on the base W'. Thus, the above equation gives the complex structure of the total space R.

What is described above is the partial unfolding of the $SU(6)$ singularity, keeping an $SU(5)$ singularity. In our G_2 problem, we need a similar construction, but we must view the $SU(6)$ singularity as a hyper-Kähler manifold, not just a complex manifold. In unfolding the $SU(6)$ singularity as a hyper-Kähler manifold, each complex parameter in P_4 is accompanied by a real parameter that controls the area of an exceptional divisor in the resolution/deformation of the singularity. The parameters are thus not five complex parameters but five triplets of real parameters. (There is an $SO(3)$ symmetry that rotates each triplet. This is the $SO(3)$ rotating the three Kähler forms in section three.)

To get a G_2-manifold, we must combine the complex parameter seen in (5.15) with a corresponding real parameter. Altogether, this will give a three-parameter family of deformations of the $SU(6)$ singularity (understood as a hyper-Kähler manifold) to a hyper-Kähler manifold with an $SU(5)$ singularity. The parameter space of this deformation is what we have called W, and the total space is a seven-manifold that is our desired singular G_2-manifold X, with a singularity that produces the chiral fermions that we analyzed above in the heterotic string language.

To find the hyper-Kähler unfolding of the $SU(6)$ singularity that preserves an $SU(5)$ singularity is not difficult, using Kronheimer's description of the general unfolding via a hyper-Kähler quotient [**18**]. At this stage, we might as well generalize to $SU(N)$, so we consider a hyper-Kähler unfolding of the $SU(N+1)$ singularity to give an $SU(N)$ singularity. The unfolding of the $SU(N+1)$ singularity is obtained by taking a system of $N+1$ hypermultiplets $\Phi_0, \Phi_1, \ldots \Phi_N$ with an action of $K = U(1)^N$. Under the i^{th} $U(1)$ for $i = 1, \ldots, N$, Φ_i has charge 1, Φ_{i-1} has charge -1, and the others are neutral. This configuration of hypermultiplets and gauge fields is known as the quiver diagram of $SU(N+1)$ and appears in studying D-branes near the $SU(N+1)$ singularity We let $\mathbb{H}$ denote $\mathbb{R}^4$, so the hypermultiplets parameterize $\mathbb{H}^{N+1}$, the product of $N+1$ copies of $\mathbb{R}^4$. The hyper-Kähler quotient of $\mathbb{H}^{N+1}$ by K is obtained by setting the $\vec{D}$-field (or components of the hyper-Kähler moment map) to zero and dividing by K. It is denoted $\mathbb{H}^{N+1}//K$, and is isomorphic to the $SU(N+1)$ singularity $\mathbb{R}^4/\mathbb{Z}_{N+1}$. Its unfolding is described by setting the $\vec{D}$-fields equal to arbitrary constants, not necessarily zero. In all, there are $3N$ parameters in this unfolding – three times the dimension of K – since for each $U(1)$, $\vec{D}$ has three components, rotated by an $SO(3)$ group of R-symmetries.

We want a partial unfolding keeping an $SU(N)$ singularity. To describe this, we keep $3(N-1)$ of the parameters equal to zero and let only the remaining three vary; these three will be simply the values of $\vec{D}$ for one of the $U(1)$'s. The seven-manifold which we propose admits a natural G_2-holonomy metric is easy to describe. One picks a $U(1)$ subgroup of K - the gauge group of the Kronheimer construction. There are three D-terms $\vec{D}$ associated to this $U(1)$. Then one simply repeats Kronheimer's construction, but one does not set $\vec{D}$ to zero. This gives a 7-manifold which maps to $\mathbb{R}^3$ (parameterised by the space of values of $\vec{D}$) over which the generic fiber is the ADE-singularity obtained from the Kronheimer construction using K', the commutant of $U(1)$ in K. However, at the origin, i.e. when $\vec{D}$ is zero, the fiber degenerates further to an ADE-singularity of higher rank. This is exactly the sort of picture we expected from the heterotic string.

To carry out this procedure, we first write $K = K' \times U(1)'$ (where $U(1)'$ denotes a chosen $U(1)$ factor of $K = U(1)^N$). Then we take the hyper-Kähler quotient of $\mathbb{H}^{N+1}$ by K' to get a hyper-Kähler eight-manifold $\hat{X} = \mathbb{H}^{N+1}//K'$, after which we take the *ordinary* quotient, not the hyper-Kähler quotient, by $U(1)'$ to get a seven-manifold $X = \hat{X}/U(1)'$ that should admit a metric of G_2-holonomy. X has a natural map to $W = \mathbb{R}^3$ given by the value of the $\vec{D}$-field of $U(1)'$ – which was not set to zero – and this map gives the fibration of X by hyper-Kähler manifolds.

In the present example, we can easily make this explicit. We take $U(1)'$ to be the "last" $U(1)$ in $K = U(1)^N$, so $U(1)'$ only acts on Φ_{N-1} and Φ_N. K' is therefore the product of the first $N-1$ $U(1)$'s; it acts trivially on Φ_N, and acts on $\Phi_0, \ldots, \Phi_{N-1}$ according to the standard quiver diagram of $SU(N)$. So the hyper-Kähler quotient $\mathbb{H}^{N+1}//K'$ is just $(\mathbb{H}^N//K') \times \mathbb{H}'$, where $\mathbb{H}^N//K'$ is the $SU(N)$ singularity, isomorphic to $\mathbb{H}/\mathbb{Z}_N$ and $\mathbb{H}'$ is parameterized by Φ_N. So finally, X will be $(\mathbb{H}/\mathbb{Z}_N \times \mathbb{H}')/U(1)'$. To make this completely explicit, we just need to identify the group actions on $\mathbb{H}$ and $\mathbb{H}'$. If we parameterize $\mathbb{H}$ and $\mathbb{H}'$ respectively by pairs of complex variables (a, b) and (a', b') then the $\mathbb{Z}_N$ action on $\mathbb{H}$, such that the quotient $\mathbb{H}/\mathbb{Z}_N$ is the $SU(N)$ singularity, is given by

$$\begin{pmatrix} a \\ b \end{pmatrix} \to \begin{pmatrix} e^{2\pi ik/N} a \\ e^{-2\pi ik/N} b \end{pmatrix}, \tag{5.16}$$

while the $U(1)'$ action that commutes with this (and preserves the hyper-Kähler structure) is

$$\begin{pmatrix} a \\ b \end{pmatrix} \to \begin{pmatrix} e^{i\psi/N} a \\ e^{-i\psi/N} b \end{pmatrix}. \tag{5.17}$$

The $U(1)'$ action on $\mathbb{H}'$ is similarly

$$\begin{pmatrix} a' \\ b' \end{pmatrix} \to \begin{pmatrix} e^{i\psi/N} a' \\ e^{-i\psi/N} b' \end{pmatrix}. \tag{5.18}$$

In all, if we set $\lambda = e^{i\psi/N}$, $w_1 = \overline{a}'$, $w_2 = b'$, $w_3 = a$, $w_4 = \bar{b}$, then the quotient $\mathbb{H}/\mathbb{Z}_N \times \mathbb{H}'/U(1)$ can be described with four complex variables $w_1, \ldots, w_4$ modulo the equivalence

$$(w_1, w_2, w_3, w_4) \to (\lambda^N w_1, \lambda^N w_2, \lambda w_3, \lambda w_4), \quad |\lambda| = 1 \tag{5.19}$$

This quotient is a cone on a weighted projective space $\mathbb{WCP}^3_{N,N,1,1}$. In fact, if we impose the above equivalence relation for all nonzero complex λ, we would get the weighted projective space itself; by imposing this relation only for $|\lambda| = 1$, we get a cone on the weighted projective space. Note that the conical metric of G_2-holonomy on this space does not use usual Kähler metric on weighted projective space.

$\mathbb{WCP}^3_{N,N,1,1}$ has a family of A_{N-1}-singularities at points $(w_1, w_2, 0, 0)$. This is easily seen by setting λ to $e^{2\pi i/N}$. This set of points is a copy of $\mathbb{CP}^1 = \mathbb{S}^2$. Our proposed G_2-manifold is a cone over weighted projective space, so it has a family of A_{N-1}-singularities which are a cone over this $\mathbb{S}^2$. This is of course a copy of $\mathbb{R}^3$. Away from the origin in $\mathbb{R}^3$ the only singularities are these orbifold singularities. At the origin however, the whole manifold develops a conical singularity. There, the 2-sphere, which is noncontractible in the bulk of the manifold, shrinks to zero size. This is in keeping with the anomaly inflow arguments of the previous section. There we learned that an ADE-singularity which worsens over a point in W is a good candidate for the appearance of chiral fermions. Here, via duality with the heterotic string, we find that the conical singularity in this example supports one chiral fermion in the $\mathbf{N}$ of the $SU(N)$ gauge symmetry coming from the A_{N-1}-singularity. In fact, the $U(1)$ gauge symmetry from the C-field in this example combines with the $SU(N)$ to give a gauge group which is globally $U(N)$ and the fermion is in the fundamental representation.

Some extensions of this can be worked out in a similar fashion. Consider the case that away from P, the monodromies break $SU(N+1)$ to $SU(p)\times SU(q)\times U(1)$, where $p+q=N+1$. Analysis of the Dirac equation along the above lines shows that such a model will give chiral fermions transforming as $(\mathbf{p},\bar{\mathbf{q}})$ under $SU(p)\times SU(q)$ (and charged under the $U(1)$). To describe a dual in M theory on a manifold of G_2 holonomy, we let $K=K'\times U(1)'$, where now $K'=K_1\times K_2$, K_1 being the product of the first $p-1$ $U(1)$'s in K and K_2 the product of the last $q-1$, while $U(1)'$ is the p^{th} $U(1)$. Now we must define $\hat{X}=\mathbb{H}^{N+1}//K'$, and the manifold admitting a metric of G_2 holonomy should be $\hat{X}/U(1)'$.

We can compute $\hat{X}$ easily, since K_1 acts only on $\Phi_1,\ldots,\Phi_p$ and K_2 only on $\Phi_{p+1},\ldots,\Phi_{N+1}$. The hyper-Kähler quotients by K_1 and K_2 thus simply construct the $SU(p)$ and $SU(q)$ singularities, and hence $\hat{X}=\mathbb{H}/\mathbb{Z}_p\times\mathbb{H}/\mathbb{Z}_q$. $\hat{X}$ has planes of $\mathbb{Z}_p$ and $\mathbb{Z}_q$ singularities, which will persist in $X=\hat{X}/U(1)'$, which will also have a more severe singularity at the origin. So the model describes a theory with $SU(p)\times SU(q)$ gauge theory and chiral fermions supported at the origin. $U(1)'$ acts on $\mathbb{H}/\mathbb{Z}_p$ and $\mathbb{H}/\mathbb{Z}_q$ as the familiar global symmetry that preserves the hyper-Kähler structure of the $SU(p)$ and $SU(q)$ singularities. Representing those singularities by pairs (a,b) and (a',b') modulo the usual action of $\mathbb{Z}_p$ and $\mathbb{Z}_q$, $U(1)'$ acts by

$$\begin{pmatrix} a \\ b \end{pmatrix}\rightarrow\begin{pmatrix} e^{i\psi/p}a \\ e^{-i\psi/p}b \end{pmatrix}\quad and\quad \begin{pmatrix} a' \\ b' \end{pmatrix}\rightarrow\begin{pmatrix} e^{-i\psi/q}a' \\ e^{i\psi/q}b' \end{pmatrix}. \tag{5.20}$$

Now if p and q are relatively prime, we set $\lambda=e^{i\psi/pq}$, and we find that the $U(1)'$ action on the complex coordinates $w_1,\ldots,w_4$ (which are defined in terms of a,b,a',b' by the same formulas as before) is

$$(w_1,w_2,w_3,w_4)\rightarrow(\lambda^p w_1,\lambda^p w_2,\lambda^q w_3,\lambda^q w_4). \tag{5.21}$$

If p and q are relatively prime, then the $U(1)'$ action, upon taking λ to be a p^{th} or q^{th} root of 1, generates the $\mathbb{Z}_p\times\mathbb{Z}_q$ orbifolding that is part of the original definition of $\hat{X}$. Hence in forming the quotient $\hat{X}/U(1)'$, we need only to act on the w's by the equivalence relation. The quotient is therefore a cone on a weighted projective space $\mathbb{WCP}^3_{p,p,q,q}$. If p and q are not relatively prime, we let $(p,q)=r(n,m)$ where r is the greatest common divisor and n and m are relatively prime. Then we let $\lambda=\exp(ir\psi/pq)$, so the equivalence relation above is replaced with

$$(w_1,w_2,w_3,w_4)\rightarrow(\lambda^n w_1,\lambda^n w_2,\lambda^m w_3,\lambda^m w_4). \tag{5.22}$$

To reproduce $\hat{X}/U(1)$ we must now also divide by $\mathbb{Z}_r$, acting by

$$(w_1,w_2,w_3,w_4)\rightarrow(\zeta w_1,\zeta w_2,w_3,w_4), \tag{5.23}$$

where $\zeta^r=1$. So X is a cone on $\mathbb{WCP}^3_{n,n,m,m}/\mathbb{Z}_r$.

5.4. Other Gauge Groups and Matter Representations. We now explain how to generalise the above construction to obtain singularities with more general gauge groups and chiral fermion representations. Suppose that we want to get chiral fermions in the representation R of a simply-laced group G. This can be achieved for certain representations. We find a simply-laced group $\hat{G}$ of rank one more than the rank of G, such that $\hat{G}$ contains $G \times U(1)$ and the Lie algebra of $\hat{G}$ decomposes as $\mathbf{g} \oplus \mathbf{o} \oplus \mathbf{r} \oplus \bar{\mathbf{r}}$, where $\mathbf{g}$ and $\mathbf{o}$ are the Lie algebras of G and $U(1)$, $\mathbf{r}$ transforms as R under G and of charge 1 under $U(1)$, and $\bar{\mathbf{r}}$ transforms as the complex conjugate. Such a $\hat{G}$ exists only for special R's, and these are the R's that we will generate from G_2 singularities.

Given $\hat{G}$, we proceed as above on the heterotic string side. We consider a family of $\mathbf{T}^3$'s, parameterized by W, with monodromy that at a special point $P \in W$ leaves unbroken $\hat{G}$, and at a generic point breaks $\hat{G}$ to $G \times U(1)$. We moreover assume that near P, the monodromies have the same sort of generic behavior assumed above. Then the same computation as above will show that the heterotic string has, in this situation, a single multiplet of fermion zero modes (the actual chirality depends on solving the Dirac equation) in the representation R, with $U(1)$ charge 1.

Dualizing this to an M theory description, over P we want a $\hat{G}$ singularity, while over a generic point in W we should have a G singularity. Thus, we want to consider the unfolding of the $\hat{G}$ singularity (as a hyper-Kähler manifold) that preserves a G singularity. To do this is quite simple. We start with the Dynkin diagram of $\hat{G}$. The vertices are labeled with integers n_i, the Dynkin indices. In Kronheimer's construction, the $\hat{G}$ singularity is obtained as the hyper-Kähler quotient of $\mathbb{H}^k$ (for some k) by the action of a group $K = \prod_i U(n_i)$. Its unfolding is obtained by allowing the $\vec{D}$-fields of the $U(1)$ factors (the centers of the $U(n_i)$) to vary.

The G Dynkin diagram is obtained from that of $\hat{G}$ by omitting one node, corresponding to one of the $U(n_i)$ groups; we write the center of this group as $U(1)'$. Then we write K (locally) as $K = K' \times U(1)'$, where K' is defined by replacing the relevant $U(n_i)$ by $SU(n_i)$. We get a hyper-Kähler eight-manifold as the hyper-Kähler quotient $\hat{X} = \mathbb{H}^k//K'$, and then we get a seven-manifold X by taking the *ordinary* quotient $X = \hat{X}/U(1)'$. This maps to $W = \mathbb{R}^3$ by taking the value of the $U(1)'$ $\vec{D}$-field, which was not set to zero. The fiber over the origin is obtained by setting this $\vec{D}$-field to zero after all, and gives the original $\hat{G}$ singularity, while the generic fiber has a singularity of type G.

One can readily work out examples of pairs G, $\hat{G}$. We will just consider the cases most relevant for grand unification. For $G = SU(N)$, to get chiral fields in the antisymmetric tensor representation, $\hat{G}$ should be $SO(2N)$. For $G = SO(10)$, to get chiral fields in the $\mathbf{16}$, $\hat{G}$ should be E_6. For $G = SO(2k)$, to get chiral fields in the $\mathbf{2k}$, $\hat{G}$ should be $SO(2k+2)$. (Note in this case that $\mathbf{2k}$ is a real representation. However, the monodromies in the above construction break $SO(2k+2)$ to $SO(2k) \times U(1)$, and the massless $\mathbf{2k}$'s obtained from the construction are charged under the $U(1)$; under $SO(2k) \times U(1)$ the representation is complex.) For $2k = 10$, this example might be used in constructing $SO(10)$ GUT's. For $G = E_6$, to get $\mathbf{27}$'s, $\hat{G}$ should be E_7. A useful way to describe the topology of X in these examples is not clear.

In this construction, we emphasized, on the heterotic string side, the most generic special monodromies that give enhanced gauge symmetry, which corresponds on the M theory side to omitting from the hyper-Kähler quotient a rather special $U(1)$ that is related to a single node of the Dynkin diagram. We could also consider more general heterotic string monodromies; this would correspond in M theory to omitting a more general linear combination of the $U(1)$'s.

6. Outlook.

Having gathered all the necessary ingredients we can now briefly describe how one goes about building a model of particle physics from M theory on a G_2-manifold, X. First it is natural that X admits a map to a three-manifold W. The generic fibers of the map are all $K3$-surfaces which have an ADE singularity of some fixed type. $\mathsf{A}_4 = SU(5)$ is a promising possibility for particle physics. This plays the role of the GUT gauge group.

At a finite number of points on X which are also on W, there are conical singularities of the kind discussed in section five. These support chiral fermions in various representations of the ADE gauge group. For instance, in the case of $SU(5)$ we would like to obtain three $\mathbf{5}$'s and three $\overline{\mathbf{10}}$'s. The singularities of X should be of the required type.

We then take W to non-simply connected (e.g. W might be $S^3/\mathbb{Z}_n$). Wilson lines (or flat connections) of the $SU(5)$ gauge fields can then be used to break $SU(5)$ to $SU(3)\times SU(2)\times U(1)$ - the gauge group of the standard model.

An analysis of some of the basic properties of such models (assuming a suitable X exists) was carried out in [**25**]. It was found that one of the basic physical tests of such a model - namely the stability of the proton - was not problematic. This is because the various families of chiral fermions originate from different points on X so it is natural for them to be charged under different discrete symmetries. These symmetries prevent the existence of operators which would otherwise mediate the decay of the proton too quickly.

On the mathematical side, we still do not have examples of compact G_2-manifolds with these conical singularities. The physics suggests that they are natural spaces to construct and we hope that this will be done in the near future.

Of course, we are still a long way from having a realistic explanation of particle physics through M theory, since we do not understand in detail why supersymmetry is broken in nature and why the cosmological constant is so small.

References

1. B. S. Acharya, *M theory, Joyce orbifolds and super Yang-Mills,* Adv. Theor. Math. Phys. **3**, 227 (1999) [arXiv:hep-th/9812205].
2. B. S. Acharya, *On realising $N = 1$ super Yang-Mills in M theory,* arXiv:hep-th/0011089.
3. B. S. Acharya, *Confining strings from G_2-holonomy spacetimes,* arXiv:hep-th/0101206.
4. B. S. Acharya and C. Vafa, *On domain walls of $N = 1$ supersymmetric Yang-Mills in four dimensions,* arXiv:hep-th/0103011.
5. B. Acharya and E. Witten, *Chiral fermions from manifolds of G_2 holonomy,* arXiv:hep-th/0109152.
6. M. Atiyah, J. Maldacena and C. Vafa, *An M-theory flop as a large N duality,* J. Math. Phys. **42**, 3209 (2001) [arXiv:hep-th/0011256].
7. M. Atiyah and E. Witten, *M-theory dynamics on a manifold of G_2 holonomy,* Adv. Theor. Math. Phys. 6 (2002) **1**. arXiv:hep-th/0107177.
8. M. Bershadsky, C. Vafa and V. Sadov, *D-Branes and Topological Field Theories,* Nucl. Phys. B. **463**, 420 (1996) [arXiv:hep-th/9511222].
9. R. Bryant and S. Salamon, *On the Construction of Complete Metrics wtih Exceptional Holonomy,* Duke Math Journal,**58** (1989) 829. G. W. Gibbons, D. N. Page and C. N. Pope, *Einstein Metrics On S^3, $\mathbf{R}^3$ and $\mathbf{R}^4$ Bundles,* Commun. Math. Phys. **127**, 529 (1990).
10. M. Cvetic, G. Shiu and A. M. Uranga, *Chiral four-dimensional $N = 1$ supersymmetric type IIA orientifolds from intersecting D6-branes,* Nucl. Phys. B. **615**, 3 (2001) [arXiv:hep-th/0107166].
11. D. E. Diaconescu, M. R. Douglas and J. Gomis, *Fractional branes and wrapped branes,* JHEP **9802**, 013 (1998) [arXiv:hep-th/9712230].

12. M. R. Douglas and G. W. Moore, *D-branes, Quivers, and ALE Instantons*, arXiv:hep-th/9603167.
13. M. J. Duff, B. E. Nilsson and C. N. Pope, *Kaluza-Klein Supergravity,* Phys. Rept. **130**, 1 (1986).
14. J. A. Harvey and G. W. Moore, *Superpotentials and membrane instantons,* arXiv:hep-th/9907026.
15. P. Hořava and E. Witten, *Eleven-Dimensional Supergravity on a Manifold with Boundary,* Nucl. Phys. B **475**, 94 (1996) [arXiv:hep-th/9603142].
16. D. D. Joyce, *Compact Manifolds of Special Holonomy,* Oxford University Press, 2000.
17. S. Katz and C. Vafa, *Matter from geometry,* Nucl. Phys. B **497**, 146 (1997) [arXiv:hep-th/9606086].
18. P. Kronheimer, *Construction of ALE-spaces as hyper-Kähler Quotients,* J. Diff. Geom. **28** (1989) 665.
19. K. S. Narain, *New Heterotic String Theories In Uncompactified Dimensions* < 10*,* Phys. Lett. B **169**, 41 (1986).
20. G. Papadopoulos and P. K. Townsend, *Compactification of $D = 11$ supergravity on spaces of exceptional holonomy,* Phys. Lett. B **357**, 300 (1995) [arXiv:hep-th/9506150].
21. N. Seiberg and E. Witten, *Gauge dynamics and compactification to three dimensions,* The mathematical beauty of physics (Saclay 1996) 333 Adv Ser. Math. Phys. **24** World Scientific. arXiv:hep-th/9607163. N. M. Davies, T. J. Hollowood and V. V. Khoze, *Monopoles, affine algebras and the gluino condensate,* arXiv:hep-th/0006011.
22. E. Witten, *Fermion Quantum Numbers In Kaluza-Klein Theory,* In Appelquist, T. (ED.) et al: *Modern Kaluza-Klein Theories*, 438-511. *Shelter Island 1983, Procedings, Quantum Field Theory and the Fundamental Problems of Physics*, 227-277.
23. E. Witten, *String theory dynamics in various dimensions,* Nucl. Phys. B **443**, 85 (1995) [arXiv:hep-th/9503124].
24. E. Witten, *Anomaly cancellation on G_2-manifolds,* arXiv:hep-th/0108165.
25. E. Witten, *Deconstruction, G(2) holonomy, and doublet-triplet splitting,* arXiv:hep-ph/0201018.

Current address: Department of Physics, Rutgers University, 126 Frelinghuysen Road, Piscataway, NJ 08854

E-mail address: `bacharya@physics.rutgers.edu`

Clay Mathematics Proceedings
Volume 3, 2004

Conjectures in Kähler geometry

Simon K. Donaldson

Abstract. We state a general conjecture about the existence of Kähler metrics of constant scalar curvature, and discuss the background to the conjecture.

1. The equations

In this article we discuss some well-known problems in Kähler geometry. The general theme is to ask whether a complex manifold admits a preferred Kähler metric, distinguished by some natural differential-geometric criterion. A paradigm is the well-known fact that any Riemann surface admits a metric of constant Gauss curvature. Much of the interest of the subject comes from the interplay between, on the one hand, the differential geometry of metrics, curvature tensors *etc.* and, on the other hand, the complex analytic or algebraic geometry of the manifold. This is, of course, a very large field and we make no attempt at an exhaustive account, but it seems proper to emphasise at the outset that many of these questions have been instigated by seminal work of Calabi.

Let (V, ω_0) be a compact Kähler manifold of complex dimension n. The Kähler forms in the class $[\omega_0]$ can be written in terms of a Kähler potential $\omega_\phi = \omega_0 + i\overline{\partial}\partial\phi$. In the case when $2\pi[\omega_0]$ is an integral class, e^ϕ has a geometrical interpretation as the change of metric on a holomorphic line bundle $L \to V$. The *Ricci form* $\rho = \rho_\phi$ is $-i$ times the curvature form of K_V^{-1}, with the metric induced by ω_ϕ, so $[\rho] = 2\pi c_1(V) \in H^2(V)$. In the 1950's, Calabi [**3**] initiated the study of *Kähler-Einstein* metrics, with

$$\rho_\phi = \lambda\omega_\phi, \tag{1}$$

for constant λ. For these to exist we need the topological condition $2\pi c_1(V) = \lambda[\omega]$. When this condition holds we can write (by the $\overline{\partial}\partial$ Lemma)

$$\rho_0 - \lambda\omega_0 = i\overline{\partial}\partial f,$$

for some function f. The Kähler-Einstein equation becomes the second order, fully nonlinear, equation

$$(\omega_0 + i\overline{\partial}\partial\phi)^n = e^{f-\lambda\phi}\omega_0^n. \tag{2}$$

2000 *Mathematics Subject Classification.* 32Q15, 53C55, 53C56.

More explicitly, in local coordinates z_α and in the case when the metric ω_0 is Euclidean, the equation is

$$\det\left(\delta_{\alpha\beta} + \frac{\partial^2\phi}{\partial z_\alpha \partial \overline{z}_\beta}\right) = e^{f-\lambda\phi}. \tag{3}$$

This is a complex Monge-Ampère equation and the analysis is very much related to that of real Monge-Ampère equations of the general shape

$$\det\left(\frac{\partial^2 u}{\partial x_i \partial x_j}\right) = F(x,u), \tag{4}$$

where u is a convex function on an open set in $\mathbf{R}^n$. There is a tremendous body of work on these real and complex Monge-Ampère equations. In the Kähler setting, the decisive contributions, dating back to the 1970's, are due to Yau [**9**] and Aubin [**2**]. The conclusion is, roughly stated, that PDE techniques reduce the problem of finding a solution to that of finding a priori bounds for $\|\phi\|_{L^\infty}$. In the case when $\lambda < 0$ this bound is easily obtained from the maximum principle; in the case when $\lambda = 0$ the bound follows from a more sophisticated argument of Yau. This leads, of course, to the renowned Calabi-Yau metrics on manifolds with vanishing first Chern class. When λ is positive, so in algebro-geometric language we are considering a *Fano manifold* V, the bound on $\|\phi\|_{L^\infty}$, and hence the existence of a Kähler-Einstein metric, may hold or may not, depending on more subtle properties of the geometry of the manifold V and, in a long series of papers, Tian has made enormous progress towards understanding precisely when a solution exists. Notably, Tian made a general conjecture in [**8**], which we will return to in the next section.

In the early 1980's, Calabi initiated another problem [**4**]. His starting point was to consider the L^2 norm of the curvature tensor as a functional on the metrics and seek critical points, called *extremal Kähler metrics*. The Euler-Lagrange equations involve the *scalar curvature*

$$S = (\rho \wedge \omega^{n-1})/\omega^n.$$

The extremal condition is the equation

$$\overline{\partial}(\text{grad } S) = 0. \tag{5}$$

On the face of it this is a very intractable partial differential equation, combining the full nonlinearity of the Monge-Ampère operator, which is embedded in the definition of the curvature tensor, with high order: the equation being of order six in the derivatives of the Kähler potential ϕ. Things are not, however, quiet as bad as they may seem. The extremal equation asserts that the vector field grad S on V is holomorphic so if, for example, there are no non-trivial holomorphic vector fields on V the equation reduces to the *constant scalar curvature equation*

$$S = \sigma, \tag{6}$$

where the constant σ is determined by V through Chern-Weil theory. This reduction still leaves us with an equation of order four and, from the point of view of partial differential equations, the difficulty which permeates the theory is that one cannot directly apply the maximum principle to equations of this order. From the point of view of Riemannian geometry, the difficulty which permeates the theory is that control of the scalar curvature—in contrast to the Ricci tensor—does not give much control of the metric.

In the case when $[\rho] = \lambda[\omega]$ the constant scalar curvature and Kähler-Einstein conditions are equivalent. Of course this is a global phenomenon: locally the equations are quite different. Obviously Kähler-Einstein implies constant scalar curvature. Conversely, one has an identity

$$\overline{\partial} S = \partial^* \rho,$$

so if the scalar curvature is constant the Ricci form ρ is *harmonic*. But $\lambda\omega$ is also a harmonic form so if ρ and $\lambda\omega$ are in the same cohomology class they must be equal, by the uniqueness of harmonic representatives.

There are two parabolic evolution equations associated to these problems. The Ricci flow

$$\frac{\partial \omega}{\partial t} = \rho - \lambda\omega. \tag{7}$$

and the Calabi flow

$$\frac{\partial \omega}{\partial t} = i\overline{\partial}\partial S. \tag{8}$$

Starting with an extremal metric, the Calabi flow evolves the metric by diffeomeorphims (the one-parameter group generated by the vector field grad S): the geometry is essentially unchanged. The analogues of general extremal metrics (nonconstant scalar curvature) for the Ricci flow are the "Ricci solitons"

$$\rho - \lambda\omega = L_v \omega, \tag{9}$$

where L_v is the Lie derivative along a holomorphic vector field v.

2. Conjectural picture

We present a precise algebro-geometric condition which we expect to be equivalent to the existence of a constant scalar curvature Kähler metric. This conjecture is formulated in [**6**]; in the Kähler-Einstein/Fano case the conjecture is essentially the same as that made by Tian in [**8**]. An essential ingredient is the notion of the "Futaki invariant". Suppose $L \to V$ is a holomorphic line bundle with $c_1(L) = 2\pi[\omega_0]$ and with a Hermitian metric whose induced connection has curvature $-i\omega$. Suppose we have a $\mathbf{C}^*$-action α on the pair V, L. Then we get a complex-valued function H on V by comparing the horizontal lift of the vector field generating the action on V with that generating the action on L. In the case when $S^1 \subset \mathbf{C}^*$ acts by isometries H is real-valued and is just the Hamiltonian in the usual sense of symplectic geometry. The Futaki invariant of the $\mathbf{C}^*$-action is

$$\int_V (S - \sigma) H,$$

where σ is the average value of the scalar curvature S (a topological invariant). There is another, more algebro-geometric, way of describing this involving determinant lines. For large k we consider the line

$$\Lambda^{max} H^0(V; L^k).$$

The $\mathbf{C}^*$-action on (V, L) induces an action on this line, with some integer weight w_k. Let d_k be the dimension of $H^0(V; L^k)$ and $F(k) = w_k / k d_k$. By standard theory, this has an expansion for large k:

$$F(k) = F_0 + F_1 k^{-1} + F_2 k^{-2} + \dots.$$

The equivariant Riemann-Roch formula shows that the Futaki invariant is just the coefficient F_1 in this expansion. Turning things around, we can *define* the Futaki invariant to be F_1, the advantage being that this algebro-geometric point of view extends immediately to singular varieties, or indeed general schemes.

Given (V, L), we define a "test configuration" of exponent r to consist of

(1) a scheme $\mathcal{V}$ with a line bundle $\mathcal{L} \to \mathcal{V}$;
(2) a map $\pi : \mathcal{V} \to \mathbf{C}$ with smooth fibres $V_t = \pi^{-1}(t)$ for non-zero t, such that V_t is isomorphic to V and the restriction of $\mathcal{L}$ is isomorphic to L^r;
(3) a $\mathbf{C}^*$-action on $\mathcal{L} \to \mathcal{V}$ covering the standard action on $\mathbf{C}$.

We define the Futaki invariant of such a configuration to be the invariant of the action on the central fibre $\pi^{-1}(0)$, (with the restriction of $\mathcal{L}$) - noting that this may not be smooth. We say that the configuration is "destabilising" if the Futaki invariant is bigger than or equal to zero and, in the case of invariant zero the configuration is not a product $V \times \mathbf{C}$. Finally we say that (V, L) is "K-stable" (Tian's terminology) if there are no destabilising configurations. Then our conjecture is:

Conjecture 1. *Suppose* (V, ω_0) *is a compact Kähler manifold and* $[\omega_0] = 2\pi c_1(L)$. *Then there is a metric of constant scalar curvature in the class* $[\omega_0]$ *if and only if* (V, L) *is K-stable.*

The direct evidence for the truth of this conjecture is rather slim, but we will attempt to explain briefly why one might hope that it is true.

The first point to make is that "K-stability", as defined above, is related to the standard notion of "Hilbert-Mumford stability" in algebraic geometry. That is, we consider, for fixed large k, the embedding $V \to \mathbf{CP}^N$ defined by the sections of L^k which gives a point $[V, L]_k$ in the appropriate Hilbert scheme of subschemes of $\mathbf{CP}^N$. The group $SL(N+1, \mathbf{C})$ acts on this Hilbert scheme, with a natural linearisation, so we have a standard notion of Geometric Invariant Theory stability of $[V, L]_k$. Then K-stability of (V, L) is closely related to the stability of $[V, L]_k$ for all sufficiently large k. (The notions are not quite the same: the distinction between them is analogous to the distinction between "Mumford stability" and "Gieseker stability" of vector bundles.)

The second point to make is that there is a "moment map" interpretation of the differential geometric set-up. This is explained in more detail in [**5**] , although the main idea seems to be due originally to Fujiki [**7**]. For this, we change our point of view and instead of considering different metrics (i.e. symplectic forms) on a fixed complex manifold we fix a symplectic manifold (M, ω) and consider the set $\mathcal{J}$ of compatible complex structures on M. Thus a point J in $\mathcal{J}$ gives the same data–a complex manifold with a Kähler metric–which we denoted previously by (V, ω). The group $\mathcal{G}$ of "exact" symplectomorphisms of (M, ω) acts on $\mathcal{J}$ and one finds that the map $J \mapsto S - \sigma$ is a *moment map* for the action. In this way, the moduli space of constant scalar curvature Kähler metrics appears as the standard symplectic quotient of $\mathcal{J}$. In such situations, one anticipates that the symplectic quotient will be identified with a complex quotient, involving the complexification of the relevant group. In the case at hand, the group $\mathcal{G}$ does not have a *bona fide* complexification but one can still identify infinite-dimensional submanifolds which play the role of the orbits of the complexification: these are just the equivalence classes under the equivalence relation $J_1 \sim J_2$ if (M, J_1) and (M, J_2) are isomorphic as complex manifolds. With this identification, and modulo the detailed notion of stability, Conjecture 1 becomes the familiar statement that a stable orbit for the complexified group contains a zero of the moment map. Of course all of this is a formal picture and does not lead by itself to any kind of proof, in this infinite-dimensional setting. We do however get a helpful and detailed analogy with the better understood theory of Hermitian Yang-Mills connections. In this analogy the constant scalar curvature equation corresponds to the Hermitian Yang-Mills equation, for a connection A on a holomorphic bundle over a *fixed* Kähler manifold,

$$F_A.\omega = \text{constant}.$$

The extremal equation $\overline{\partial}(\text{grad } S) = 0$ corresponds to the Yang-Mills equation

$$d_A^* F_A = 0,$$

whose solutions, in the framework of holomorphic bundles, are just direct sums of Hermitian-Yang-Mills connections.

Leaving aside these larger conceptual pictures, let us explain in a down-to-earth way why one might expect Conjecture 1 to be true. Let us imagine that we can solve the Calabi flow equation (8) with some arbitrary initial metric ω_0. Then, roughly, the conjecture asserts that one of four things should happen in the limit as $t \to \infty$. (We are discussing this flow, here, mainly for expository purposes. One would expect similar phenomena to appear in other procedures, such as the continuity method. But it should be stressed that, in reality, there are very few rigorous results about this flow in complex dimension $n > 1$: even the long time existence has not been proved.)

(1) The flow converges, as $t \to \infty$, to the desired constant scalar curvature metric on V.
(2) The flow is asymptotic to a one-parameter family of extremal metrics on the same complex manifold V, evolving by diffeomeorphisms. Thus in this case V admits an extremal metric. Transforming to the other setting, of a fixed symplectic form, the flow converges to a point in the equivalence class defined by V. In this case V cannot be K-stable, since the diffeomorphisms arise from a $\mathbf{C}^*$-action on V with non-trivial Futaki invariant and we get a destabilising configuration by taking $\mathcal{V} = V \times \mathbf{C}$ with this action.
(3) The manifold V does not admit an extremal metric but the transformed flow J_t on $\mathcal{J}$ converges. In this case the limit of the transformed flow lies in another equivalence class, corresponding to another complex structure V' on the same underlying differentiable manifold. The manifold V' admits an extremal metric. The original manifold V is not K-stable because there is a destabilising configuration where $\mathcal{V}$ is *diffeomorphic* to $V \times \mathbf{C}$ but the central fibre has a different complex structure V' ("jumping" of complex structure).
(4) The transformed flow J_t on $\mathcal{J}$ does not converge to any complex structure on the given underlying manifold but some kind of singularities develop. However, one can still make sufficient sense of the limit of J_t to extract a scheme from it, and this scheme can be fitted in as the central fibre of a destabilising configuration, similar to case (3).

We stress again that this is more of a programme of what one might hope eventually to prove, rather than a summary of what is really known. In the Kähler-Einstein/Fano situation one can develop a parallel programme (as sketched by Tian in [**8**]) for the Ricci flow (about which much more is known), where Ricci solitons take the place of extremal metrics. In any event we hope this brings out the point that one can approach two kinds of geometric questions, which on the face of it seem quite different.

(1) ALGEBRAIC GEOMETRY PROBLEM. Describe the possible destabilising configurations and in particular the nature of the singularities of the central fibre (e.g. does one need schemes as opposed to varieties?).
(2) PDE/DIFFERENTIAL GEOMETRY PROBLEM. Describe the possible behaviour of the Calabi flow/Ricci flow (or other continuity methods), and the nature of the singularities that can develop.

The essence of Conjecture 1 is that these different questions should have the same answer.

3. Toric varieties and a toy model

One can make some progress towards the verification of Conjecture 1 in the case when V is a toric variety [**6**]. Such a variety corresponds to an integral polytope P in $\mathbf{R}^n$ and the metric can be encoded in a convex function u on P. The constant scalar curvature condition becomes the equation (due to Abreu [**1**])

$$\sum_{i,j} \frac{\partial^2 u^{ij}}{\partial x_i \partial x_j} = -\sigma, \tag{10}$$

where (u^{ij}) is the inverse of the Hessian matrix (u_{ij}) of second derivatives of u. This formulation displays very well the way in which the equation is an analogue, of order 4, of the real Monge-Ampère equation (4). The equation is supplemented by boundary conditions which can be summarised by saying that the desired solution should be an absolute minimum of the functional

$$\mathcal{F}(u) = \int_P -\log\det(u_{ij}) + \mathcal{L}(u), \tag{11}$$

where

$$\mathcal{L}(u) = \int_{\partial P} u \; d\rho - \sigma \int_P u \; d\mu. \tag{12}$$

Here $d\mu$ is Lebesgue measure on P and $d\rho$ is a natural measure on ∂P. (Each codimension-1 face of ∂P is defined by a linear form, which we can normalise to have coprime integer coefficients. This linear form and the volume element on $\mathbf{R}^n$ induce a volume element on the face.) We wish to draw attention to one interesting point, which can be seen as a very small part of Conjecture 1. Suppose there is a non-trivial convex function g on P such that $\mathcal{L}(g) \leq 0$. Then one can show that $\mathcal{F}$ does not attain a minimum so there is no constant scalar curvature metric. Suppose on the other hand that f is a *piecewise linear, rational* convex function (that is, the maximum of a finite set of rational affine linear functions). Then one can associate a canonical test configuration to f and show that this is destabilising if $\mathcal{L}(f) \leq 0$. Thus we have

CONJECTURE 2. *If there is a non-trivial convex function g on P with $\mathcal{L}(g) \leq 0$ then there is a non-trivial piecewise-linear, rational convex function f with $\mathcal{L}(f) \leq 0$.*

This is a problem of an elementary nature, which was solved in [**6**] in the case when the dimension n is 2, but which seems quite difficult in higher dimensions. (And one can also ask for a more conceptual proof than that in [**6**] for dimension 2.) On the other hand if this Conjecture 2 is false then very likely the same is true for Conjecture 1: in that event one probably has to move outside algebraic geometry to capture the meaning of constant scalar curvature.

Even in dimension 2 the partial differential equation (10) is formidable. We can still see some interesting things if we go right down to dimension 1. Thus is this case V is the Riemann sphere and P is the interval $[-1,1]$ in $\mathbf{R}$. The equation (10) becomes

$$\frac{d^2}{dx^2}((u'')^{-1}) = -\sigma, \tag{13}$$

which one can readily solve explicitly. This is no surprise since we just get a description of the standard round metric on the 2-sphere. To make things more interesting we can consider the equation

$$\frac{d^2}{dx^2}((u'')^{-1}) = -A, \tag{14}$$

where A is a function on $(-1,1)$. This equation has some geometric meaning, corresponding to a rotationally invariant metric on the sphere whose scalar curvature is a given function $A(h)$ of the Hamiltonian h for the circle action. The boundary conditions we want are, in this case, $u'' \sim (1 \pm x)^{-1}$ as $x \to \pm 1$. But if we have a solution with

$$u''(x) \to \infty \tag{15}$$

as $x \to \pm 1$, these are equivalent to the normalisations

$$\int_{-1}^{1} A(x)dx = 1, \ \int_{-1}^{1} xA(x)dx = 0, \tag{16}$$

which we suppose hold.

We now consider the linear functional

$$\mathcal{L}_A(u) = u(1) + u(-1) - \int_{-1}^{1} u(x)A(x)dx. \tag{17}$$

and

$$\mathcal{F}_A(u) = \int_{-1}^{1} -\log(u''(x))dx + \mathcal{L}_A(u). \tag{18}$$

Then we have

THEOREM 1. *There is a solution to equations (14), (15) if and only if* $\mathcal{L}_A(f) > 0$ *for all (non-affine) convex functions* g *on* $[-1,1]$. *In this case the solution is an absolute minimum of the functional* $\mathcal{F}_A$.

To prove this we consider the function $\phi(x) = 1/u''(x)$. This should satisfy the equation $\phi'' = -A$ with $\phi \to 0$ at ± 1. Thus the function ϕ is given via the usual Green's function

$$\phi(x) = \int_{-1}^{1} g_x(y)A(y)dy,$$

where $g_x(y)$ is a linear function of y on the intervals $(-1,x)$ and $(x,1)$, vanishing on the endpoints ± 1 and with a negative jump in its derivative at $y=x$. Thus $-g_x(y)$ is a convex function on $[-1,1]$ and

$$\mathcal{L}_A(-g_x) = \int_{-1}^{1} A(y) g_x(y) dy = \phi(x).$$

Thus our hypothesis ($\mathcal{L}_A(g) > 0$ on convex g) implies that the solution ϕ is positive throughout $(-1,1)$ so we can form ϕ^{-1} and integrate twice to solve the equation

$$u'' = \phi^{-1}$$

thus finding the desired solution u. The converse is similar. The fact that the solution is an absolute minimum follows from the convexity of the functional $\mathcal{F}_A$.

References

1. M. Abreu *Kähler geometry of toric varieties and extremal metrics* International J. Math. **9** (1998) 641-51
2. T. Aubin *Equations du type Monge-Ampère sur les variétés kähleriennes compactes* Bul. Sc. Math. **102** (1978) 63-95
3. E. Calabi *The space of Kähler metrics* Proc. Int. Congress Math. Amsterdam **2** (1954) 206-7
4. E. Calabi *Extremal Kähler metrics* In: Seminar in Differential geometry (Yau,ed.) Annals of Math. Study 102 Princeton U.P 1982 259-290
5. S. Donaldson *Remarks on gauge theory, complex geometry and 4-manifold topology* Fields Medallists' Lectures (Atiyah, Iagolnitzer eds.) World Scientific 1997 13-33
6. S. Donaldson *Scalar curvature and stability of toric varieties* Jour. Differential Geometry, To appear.
7. A. Fujiki *The moduli spaces and Kähler metrics of polarised algebraic varieties* Sugaku **42** (1990) 173-91
8. G. Tian *Kähler-Einstein metrics with positive scalar curvature* Inv. Math. **130** (1997) 1-37
9. S-T. Yau *On the Ricci curvature of a compact Kähler manifold and the complex Monge-Ampère equation. I.*, Comm. Pure Appl. Math. **31** (1978) 339-411

Department of Mathematics,Imperial College, London
E-mail address: `s.donaldson@ic.ac.uk`

Clay Mathematics Proceedings
Volume 3, 2004

Branes, Calibrations and Supergravity

Jerome P. Gauntlett

Abstract. We attempt to provide an elementary and somewhat self contained discussion of the construction of supergravity solutions describing branes wrapping calibrated cycles, emphasising the geometrical aspects and focusing on D=11 supergravity. Following a discussion of the role of special holonomy backgrounds in D=11 supergravity, the basic membrane and fivebrane solutions are reviewed and the connection with the AdS/CFT correspondence is made. The world-volume description of branes is introduced and used to argue that branes wrapping calibrated cycles in special holonomy manifolds preserve supersymmetry. The corresponding supergravity solutions are constructed first in an auxiliary gauged supergravity theory which is obtained via Kaluza-Klein reduction.

1. Introduction

Supergravity theories in D=10 and D=11 spacetime dimensions play an important role in string/M-theory since they describe the low-energy dynamics. There are five different supersymmetric string theories, all in D=10. At low energies the type IIA and type IIB string theories give rise to type IIA and type IIB supergravity, respectively, while the type I, and the two heterotic string theories all give rise to type I supergravities. The five string theories are all related to each other, possibly after compactification, by various dualities. There are also dualities which relate string theory to M-theory, which resides in D=11. M-theory is much less understood than string theory, but one of the most important things that is known about it is that its low-energy effective action is given by D=11 supergravity.

Solutions to the supergravity equations of motion, particularly those that preserve some supersymmetry, are of interest for many reasons. One reason is that they are useful in studying compactifications from D=10 or 11 down to a lower dimensional spacetime. By compactifying down to four spacetime dimensions, for example, one might hope to make contact with particle physics phenomenology. In addition to strings, it is known that string theory has a rich spectrum of other extended objects or "branes". Indeed, supergravity solutions can be constructed describing the geometries around such branes, and these provide a very important description of the branes. Similarly, there are membrane and fivebrane solutions

---

2000 *Mathematics Subject Classification.* 83E30, 83E50, 53C38, 81T30.

The author thanks Bobby Acharya, Nakwoo Kim, Dario Martelli, Stathis Pakis and Daniel Waldram for enjoyable collaborations upon which some of these notes are based.

of D=11 supergravity, which will be reviewed later, which implies that M-theory contains such branes. An important application of brane and more general intersecting brane solutions is that they can be used to effectively study the quantum properties of black holes.

Supergravity solutions also provide powerful tools to study quantum field theories. The most significant example is Maldacena's celebrated AdS/CFT correspondence [**108**], which conjectures that string/M-theory on certain supergravity geometries that include anti-de Sitter (AdS) space factors is equivalent to certain conformally invariant quantum field theories (CFTs). The supergravity approximation to string/M-theory allows one to calculate highly non-trivial information about the conformal field theories.

The AdS/CFT correspondence is truly remarkable. On the one hand it states that certain quantum field theories, that *a priori* have nothing to do with gravity, are actually described by theories of quantum gravity (string/M-theory). Similarly, and equally surprising, it also states that quantum gravity on certain geometries is actually quantum field theory. As a consequence much effort has been devoted to further understanding and generalising the correspondence.

The basic AdS/CFT examples arise from studying the supergravity solutions describing planar branes in flat space, in the "near horizon limit". Roughly speaking, this is the limit close to the location of the brane. Here we shall discuss more general supergravity solutions that describe branes that are partially wrapped on various calibrated cycles within special holonomy manifolds. We will construct explicit solutions in the near horizon limit, which is sufficient for applications to the AdS/CFT correspondence.

To keep the presentation simple, we will mostly restrict our discussion to D=11 supergravity. In an attempt to make the lectures accessible to both maths and physics students we will emphasise the geometrical aspects and de-emphasise the quantum field theory aspects. To make the discussion somewhat self contained, we begin with some basic material; it is hoped that the discussion is not too pedestrian for the physics student and not too vague for the maths student!

We start with an introduction to D=11 supergravity, defining the notion of a bosonic solution of D=11 supergravity that preserves supersymmetry. We then describe why manifolds with covariantly constant spinors, and hence with special holonomy, are important. Following this we present the geometries describing planar membranes and fivebranes. These geometries have horizons, and in the near horizon limit we obtain geometries that are products of AdS spaces with spheres, which leads to a discussion of the basic AdS/CFT examples.

To motivate the search for new AdS/CFT examples we first introduce the world-volume description of branes. Essentially, this is an approximation that treats the branes as "probes" propagating in a fixed background geometry. We define calibrations and calibrated cycles, and explain why such probe-branes wrapping calibrated cycles in special holonomy manifolds preserve supersymmetry. The aim is then to construct supergravity solutions describing such wrapped branes after including the back-reaction of the branes on the special holonomy geometry.

The construction of these supergravity solutions is a little subtle. In particular, the solutions are first constructed in an auxiliary gauged supergravity theory. We will focus, for illustration, on the geometries corresponding to wrapped fivebranes. For this case the solutions are first found in $SO(5)$ gauged supergravity in D=7.

This theory arises from the consistent truncation of the dimensional reduction of D=11 supergravity on a four-sphere, as we shall discuss. This means that any solution of the D=7 supergravity theory automatically gives a solution of D=11 supergravity. We will present several details of the construction of the solutions describing fivebranes wrapping SLAG 3-cycles, and summarise more briefly the other cases. We also comment on some aspects of the construction of the solutions for wrapped membranes and D3-branes of type IIB supergravity.

We conclude with a discussion section that outlines some open problems as well as a brief discussion of other related work on the construction of wrapped NS-fivebranes of type IIB supergravity.

2. D=11 supergravity

The bosonic field content of D=11 supergravity [**36**] consists of a metric, g, and a three-form C with four-form field strength $G = dC$ which live on a D=11 manifold which we take to be spin. The signature is taken to be mostly plus, $(-, +, \ldots, +)$. In addition the theory has a fermionic gravitino, ψ_μ. The action with $\psi_\mu = 0$ is given by

$$S = \frac{1}{2\kappa^2}\int d^{11}x\sqrt{-g}R - \frac{1}{2}G \wedge *G - \frac{1}{6}C \wedge G \wedge G, \tag{2.1}$$

and thus the bosonic equations of motion, including the Bianchi identity for the four-form, are

$$\begin{aligned} R_{\mu\nu} &= \frac{1}{12}(G^2_{\mu\nu} - \frac{1}{12}g_{\mu\nu}G^2) \\ d * G + \frac{1}{2}G \wedge G &= 0 \\ dG &= 0\ , \end{aligned} \tag{2.2}$$

where $G^2_{\mu\nu} = G_{\mu\sigma_1\sigma_2\sigma_3}G_\nu{}^{\sigma_1\sigma_2\sigma_3}$ and $G^2 = G_{\sigma_1\sigma_2\sigma_3\sigma_4}G^{\sigma_1\sigma_2\sigma_3\sigma_4}$, with $\mu, \nu, \sigma = 0, 1, \ldots, 10$. The theory is invariant under supersymmetry transformations whose infinitesimal form is given schematically by

$$\begin{aligned} \delta g &\sim \epsilon\psi \\ \delta C &\sim \epsilon\psi \\ \delta\psi &\sim \hat{\nabla}\epsilon + \epsilon\psi\psi\ , \end{aligned} \tag{2.3}$$

where the spinor ϵ parameterises the variation and the connection $\hat{\nabla}$ will be given shortly.

Of primary interest are bosonic solutions to the equations of motion that preserve at least one supersymmetry. These are solutions to the equations of motion with $\psi = 0$ which are left inert under a supersymmetry variation. From (2.3) we see that $\delta g = \delta C = 0$ trivially, and hence we seek solutions to the equations of motion that admit non-trivial solutions to the equation $\hat{\nabla}\epsilon = 0$.

As somewhat of an aside we mention a potentially confusing point. For the theory to be supersymmetric, it is necessary that all of the fermions are Grassmann odd (anti-commuting) spinors. However, since the only place that fermions enter into bosonic supersymmetric solutions is via $\hat{\nabla}\epsilon = 0$ and since this is linear in ϵ we can, and will, take ϵ to be a commuting (i.e. ordinary) spinor from now on. The Grassmann odd character of the fermions is certainly important in the quantum theory, but this will not concern us here.

To be more precise about the connection $\hat{\nabla}$ let us introduce some further notation. We will use the convention that $\mu, \nu, \ldots$ are coordinate indices and $\alpha, \beta, \ldots$ are tangent space indices, i.e., indices with respect to an orthonormal frame. The D=11 Clifford algebra, $Cliff(10,1)$, is generated by gamma-matrices Γ^α satisfying the algebra

$$\Gamma^\alpha\Gamma^\beta + \Gamma^\beta\Gamma^\alpha = 2\eta^{\alpha\beta} \ , \tag{2.4}$$

with $\eta = diag(-1, 1, \ldots, 1)$. We will work in a representation where the gamma-matrices are real 32×32 matrices acting on real 32 component spinors, with $\Gamma_0\Gamma_1 \ldots \Gamma_{10} = +1$. Recall that $Spin(10,1)$ is generated by

$$\frac{1}{4}\Gamma^{\alpha\beta} \equiv \frac{1}{8}\left(\Gamma^\alpha\Gamma^\beta - \Gamma^\beta\Gamma^\alpha\right) \ , \tag{2.5}$$

and here we have introduced the notation that $\Gamma^{\alpha_1\ldots\alpha_p}$ is an anti-symmetrised product of p gamma-matrices. The charge conjugation matrix is defined to be Γ_0 and $\bar{\epsilon} \equiv \epsilon^{\mathrm{T}}\Gamma_0$.

We can now write the condition for a bosonic configuration to preserve supersymmetry as

$$\hat{\nabla}_\mu\epsilon \equiv \nabla_\mu\epsilon + \frac{1}{288}[\Gamma_\mu{}^{\nu_1\nu_2\nu_3\nu_4} - 8\delta_\mu^{\nu_1}\Gamma^{\nu_2\nu_3\nu_4}]G_{\nu_1\nu_2\nu_3\nu_4}\epsilon = 0 \ , \tag{2.6}$$

where $\nabla_\mu\epsilon$ is the usual covariant derivative on the spin bundle

$$\nabla_\mu\epsilon = (\partial_\mu + \frac{1}{4}\omega_{\mu\alpha\beta}\Gamma^{\alpha\beta})\epsilon \ . \tag{2.7}$$

Observe that the terms involving the four-form in (2.6) imply that $\hat{\nabla}$ takes values in the Clifford algebra and not just the Spin subalgebra. This is the typical situation in supergravity theories but there are exceptions, such as type I supergravity, where the connection takes values in the spin subalgebra and has totally anti-symmetric torsion [**135**].

Non-trivial solutions to (2.6) are called Killing spinors. The nomenclature is appropriate since if ϵ^i, ϵ^j are Killing spinors then $K^{ij\mu} \equiv \bar{\epsilon}^i\Gamma^\mu\epsilon^j$ are Killing vectors. To see this, first define $\Omega^{ij}_{\mu\nu} = \bar{\epsilon}^i\Gamma_{\mu\nu}\epsilon^j$ and $\Sigma^{ij}_{\mu_1\ldots\mu_5} = \bar{\epsilon}^i\Gamma_{\mu_1\ldots\mu_5}\epsilon^j$. Then use (2.6) to show that [**76**]

$$\nabla_\mu K^{ij}_\nu = \frac{1}{6}\Omega^{ij\sigma_1\sigma_2}G_{\sigma_1\sigma_2\mu\nu} + \frac{1}{6!}\Sigma^{ij\sigma_1\sigma_2\sigma_3\sigma_4\sigma_5} * G_{\sigma_1\sigma_2\sigma_3\sigma_4\sigma_5\mu\nu} \ , \tag{2.8}$$

and hence in particular $\nabla_{(\mu}K^{ij}_{\nu)} = 0$. It can also be shown that the "diagonal" Killing vectors K^{ii}, for each Killing spinor ϵ^i, are either time-like or null [**27**]. The zeroth components of these vectors in an orthonormal frame are given by $(\epsilon^i)^{\mathrm{T}}\epsilon^i$, and are clearly non-vanishing if and only if ϵ^i is, and hence so is K itself.

It is useful to know under what conditions a geometry admitting a Killing spinor will also solve the equations of motion. In the case when there is a time-like Killing spinor, i.e. a Killing spinor whose corresponding Killing vector is time-like, it was proved in [**76**] that the geometry will solve all of the equations of motion providing that G satisfies the Bianchi identity $dG = 0$ and the four-form equation of motion $d * G + (1/2)G \wedge G = 0$. If all of the Killing spinors are null, it is necessary, in addition, to demand that just one component of the Einstein equations is satisfied [**76**].

Note that given the value of a Killing spinor at a point, the connection defines a Killing spinor everywhere, via parallel transport. Also, as the Killing spinor equation is linear, the Killing spinors form a vector space whose dimension n can, in principle, be from $1, \ldots, 32$. The fraction of preserved supersymmetry is then $n/32$. Although solutions are known preserving many fractions of supersymmetry, it is not yet known if all fractions can occur (for some recent speculations on this issue see [**51**]). A general characterisation of the most general supersymmetric geometries preserving one time-like Killing spinor is presented in [**76**]. It was shown that the geometry is mostly determined by a ten-dimensional manifold orthogonal to the orbits of the Killing vector that admits an $SU(5)$-structure with rather weakly constrained intrinsic torsion. The analogous analysis for null Killing spinors has not yet been carried out. A complete classification of maximally supersymmetric solutions preserving all 32 supersymmetries is presented in [**58**].

Most of our considerations will be in the context of D=11 supergravity, but it is worth commenting on some features of M-theory that embellish D=11 supergravity. Firstly, the flux G, which is unconstrained in D=11 supergravity, satisfies a quantisation condition in M-theory. Introducing the Planck length l_p, via

$$2\kappa^2 \equiv (2\pi)^8 l_p^9 \ , \tag{2.9}$$

for M-theory on a D=11 spin manifold Y we have [**138**]

$$\frac{1}{(2\pi l_p)^3} G - \frac{\lambda}{2} \in H^4(Y, \mathbb{Z}) \ , \tag{2.10}$$

where $\lambda(Y) = p_1(Y)/2$ with $p_1(Y)$ the first Pontryagin class of Y, given below. Note that since Y is a spin manifold, $p_1(Y)$ is divisible by two. Actually, more generally it is possible to consider M-theory on unorientable manifolds admitting pinors and some discussion can be found in [**138**].

A second point is that the low-energy effective action of M-theory is given by D=11 supergravity supplemented by an infinite number of higher order corrections. It is not yet known how to determine almost all of these corrections, but there is one important exception. Based on anomaly considerations it has been shown that the equation of motion for the four-form G is modified, at next order, by [**137, 52**]

$$d * G + \frac{1}{2} G \wedge G = -\frac{(2\pi l_p)^6}{192} \left(p_1^2 - 4 p_2 \right) \ , \tag{2.11}$$

where the first and second Pontryagin forms are given by

$$p_1 = -\frac{1}{8\pi^2} tr \ R^2 \ , \qquad p_2 = -\frac{1}{64\pi^4} tr R^4 + \frac{1}{128\pi^4} (tr R^2)^2 \ . \tag{2.12}$$

At the same order there are other corrections to the equations of motion and also to the supersymmetry variations, but these have not yet been determined. Thus it is not yet known how to fully incorporate this correction consistently with supersymmetry but nevertheless it does have important consequences (see e.g. [**132**]). As this correction will not play a role in the subsequent discussion, we will ignore it.

In the next sub-sections we will review two basic classes of supersymmetric solutions to D=11 supergravity. The first class are special holonomy manifolds and the second class are the membrane and fivebrane solutions.

2.1. Special Holonomy. First consider supersymmetric solutions that have vanishing four-form flux G, where things simplify considerably. The equations of motion and the Killing spinor equation then become

$$\begin{aligned} R_{\mu\nu} &= 0 \\ \nabla_\mu \epsilon &= 0 \ . \end{aligned} \tag{2.13}$$

That is, Ricci-flat manifolds with covariantly constant spinors. The second condition implies that the manifolds have special holonomy. To see this, observe that it implies the integrability condition

$$[\nabla_\mu, \nabla_\nu]\epsilon = \frac{1}{4} R_{\mu\nu\alpha\beta}\Gamma^{\alpha\beta}\epsilon = 0 \ . \tag{2.14}$$

The subgroup of $Spin(10,1)$ generated by $R_{\mu\nu\alpha\beta}\Gamma^{\alpha\beta}$ gives the restricted holonomy group H. Thus (2.14) implies that a Killing spinor must be invariant under H. i.e. it must be a singlet under the decomposition of the **32** spinor representation of $Spin(10,1)$ into H representations, and this constrains the possible holonomy groups H that can arise.

Of most interest to us here are geometries $\mathbb{R}^{1,10-d} \times M_d$, which are the direct product of $(11-d)$-dimensional Minkowski space, $\mathbb{R}^{1,10-d}$, with a d-dimensional Riemannian manifold M_d, which we mostly take to be simply connected. (For a discussion of supersymmetric solutions with Lorentzian special holonomy, see [**27, 57**]). The possible holonomy groups of the Levi-Civita connection on manifolds M_d admitting covariantly constant spinors is well known, and we now briefly summarise the different cases.

$Spin(7)$**-Holonomy**: In $d = 8$ there are Riemannian manifolds with $Spin(7)$ holonomy. These have a nowhere vanishing self-dual Cayley four-form Ψ whose components in an orthonormal frame can be taken as

$$\begin{aligned} \Psi = {} & e^{1234} + e^{1256} + e^{1278} + e^{3456} + e^{3478} + e^{5678} \\ & + e^{1357} - e^{1368} - e^{1458} - e^{1467} - e^{2358} - e^{2367} - e^{2457} + e^{2468} \ , \end{aligned} \tag{2.15}$$

where e.g. $e^{1234} = e^1 \wedge e^2 \wedge e^3 \wedge e^4$. The Cayley four-form is covariantly constant for $Spin(7)$ manifolds and this is equivalent to Ψ being closed:

$$d\Psi = 0 \ . \tag{2.16}$$

$Spin(7)$ holonomy manifolds have a single covariantly constant chiral $Spin(8)$ spinor, which we denote by ρ. Moreover, the Cayley four-form can be constructed as a bilinear in ρ:

$$\Psi_{mnpq} = -\bar{\rho}\gamma_{mnpq}\rho \ , \tag{2.17}$$

where here $m, n, p, q = 1, \ldots, 8$. For more discussion on the spinor conventions for this case and those below, see appendix B of [**75**].

G_2**-Holonomy**: In $d = 7$ there are Riemannian manifolds with G_2 holonomy. These have a nowhere vanishing associative three-form ϕ whose components in an orthonormal frame can be taken as

$$\phi = e^{246} - e^{235} - e^{145} - e^{136} + e^{127} + e^{347} + e^{567} \ . \tag{2.18}$$

The three-form is covariantly constant and this is in fact equivalent to the conditions

$$d\phi = d * \phi = 0 \ . \tag{2.19}$$

These geometries possess a single covariantly constant minimal $d = 7$ spinor ρ. The associative three-form can be constructed from ρ via

$$\phi_{mnp} = -\mathrm{i}\bar{\rho}\gamma_{mnp}\rho \ . \tag{2.20}$$

$SU(n)$-**Holonomy**: In $d = 2n$ there are Calabi-Yau n-folds (CY_n) with $SU(n)$ holonomy. The cases relevant for D=11 supergravity have $n = 2, 3, 4, 5$. Calabi-Yau manifolds are complex manifolds, with complex structure J, and admit a nowhere vanishing holomorphic $(n, 0)$-form Ω. The Kähler form, which we also denote by J, is obtained by lowering an index on the complex structure. In an orthonormal frame we can take

$$\begin{aligned} J &= e^{12} + e^{34} + \cdots + e^{(2n-1)(2n)} \\ \Omega &= (e^1 + ie^2)(e^3 + ie^4)\ldots(e^{2n-1} + ie^{2n}) \ . \end{aligned} \tag{2.21}$$

Both J and Ω are covariantly constant and this is equivalent to the vanishing of the exterior derivative of the Kähler form and the holomorphic $(n, 0)$-form:

$$dJ = d\Omega = 0 \ . \tag{2.22}$$

These manifolds have a covariantly constant complex chiral spinor ρ. The complex conjugate of this spinor is also covariantly constant. For $n = 2, 4$ the conjugate spinor has the same chirality, while for $n = 3, 5$ it has the opposite chirality. J and Ω can be written in terms of the spinor ρ as

$$\begin{aligned} J_{mn} &= \mathrm{i}\rho^\dagger\gamma_{mn}\rho \\ \Omega_{m_1\ldots m_{2n}} &= \rho^{\mathrm{T}}\gamma_{m_1\ldots m_{2n}}\rho \ . \end{aligned} \tag{2.23}$$

$Sp(n)$-**Holonomy**: In $d = 4n$ there are hyper-Kähler n-manifolds (HK_n) with $Sp(n)$ holonomy. The cases relevant for D=11 supergravity have $n = 1, 2$. These admit three covariantly constant complex structures J^a satisfying the algebra of the imaginary quaternions

$$J^a \cdot J^b = -\delta^{ab} + \epsilon^{abc}J^c \ . \tag{2.24}$$

If we lower an index on the J^a we obtain three Kähler forms and the condition for $Sp(n)$-holonomy is equivalent to them being closed:

$$dJ^a = 0 \ . \tag{2.25}$$

Note that when $n = 1$, since $Sp(1) \cong SU(2)$, four-dimensional hyper-Kähler manifolds are equivalent to Calabi-Yau two-folds. From the CY_2 side, the extra two complex structures are obtained from the holomorphic two-form via $\Omega = J^2 + iJ^1$. The remaining case of interest for D=11 supergravity is eight-dimensional hyper-Kähler manifolds when $n = 2$. In this case, in an orthonormal frame we can take the three Kähler forms to be given by

$$\begin{aligned} J^1 &= e^{12} + e^{34} + e^{56} + e^{78} \\ J^2 &= e^{14} + e^{23} + e^{58} + e^{67} \\ J^3 &= e^{13} + e^{42} + e^{57} + e^{86} \ . \end{aligned} \tag{2.26}$$

Each complex structure J^a has a corresponding holomorphic $(4, 0)$-form given by

$$\begin{aligned} \Omega^1 &= \tfrac{1}{2}J^3 \wedge J^3 - \tfrac{1}{2}J^2 \wedge J^2 + \mathrm{i}J^2 \wedge J^3 \\ \Omega^2 &= \tfrac{1}{2}J^1 \wedge J^1 - \tfrac{1}{2}J^3 \wedge J^3 + \mathrm{i}J^3 \wedge J^1 \\ \Omega^3 &= \tfrac{1}{2}J^2 \wedge J^2 - \tfrac{1}{2}J^1 \wedge J^1 + \mathrm{i}J^1 \wedge J^2 \ . \end{aligned} \tag{2.27}$$

These manifolds have three covariantly constant $Spin(8)$ spinors of the same chirality ρ_a, $a = 1, 2, 3$. The three Kähler forms can be constructed as

$$\begin{aligned} J^1_{mn} &= -\bar{\rho}_2 \gamma_{mn} \rho_3 \\ J^2_{mn} &= -\bar{\rho}_3 \gamma_{mn} \rho_1 \\ J^3_{mn} &= -\bar{\rho}_1 \gamma_{mn} \rho_2 \ . \end{aligned} \tag{2.28}$$

In addition to these basic irreducible examples we can also consider M_d to be the direct product of two manifolds. A rather trivial possibility is to consider the product of one of the above manifolds with a number of flat directions. Two non-trivial possibilities are to consider the product $CY_3 \times CY_2$ with $SU(3) \times SU(2)$ holonomy, or the product $CY_2 \times CY_2'$ with $SU(2) \times SU(2)$ holonomy.

We have summarised the possibilities in table 1. We have also recorded the amount of D=11 supersymmetry preserved by geometries of the form $\mathbb{R}^{1,10-d} \times M_d$. As we noted, this corresponds to the total number of singlets in the decomposition of $\mathbf{32}$ of $Spin(10, 1)$ into representations of H. Let us illustrate the counting for the $d = 8$ cases. The spinor representation $\mathbf{32}$ of $Spin(10, 1)$ decomposes into $Spin(2, 1) \times Spin(8)$ representations as

$$\mathbf{32} \to (\mathbf{2}, \mathbf{8}_+) + (\mathbf{2}, \mathbf{8}_-) \ , \tag{2.29}$$

where the subscripts refer to the chirality of the two spinor representations of $Spin(8)$. When M_8 is a $Spin(7)$-manifold we have the further decomposition under $Spin(7) \subset Spin(8)$

$$\mathbf{8}_+ \to \mathbf{7} + \mathbf{1}, \qquad \mathbf{8}_- \to \mathbf{8} \ . \tag{2.30}$$

The singlet corresponds to the single covariantly constant, $Spin(7)$ invariant, spinor on the $Spin(7)$-manifold discussed above. From (2.29) we see that this gives rise to two preserved supersymmetries and that they transform as a minimal two real component spinor of $Spin(2, 1)$. This is also described as preserving $N = 1$ supersymmetry in D=3 spacetime dimensions corresponding to the $\mathbb{R}^{1,2}$ factor. When M_8 is Calabi-Yau under $SU(4) \subset Spin(8)$ we have

$$\mathbf{8}_+ \to \mathbf{6} + \mathbf{1} + \mathbf{1}, \qquad \mathbf{8}_- \to \mathbf{4} + \bar{\mathbf{4}} \ . \tag{2.31}$$

The two singlets combine to form the complex covariantly constant spinor on CY_4 mentioned above. In this case four supersymmetries are preserved, transforming as two minimal spinors of $Spin(2, 1)$, or $N = 2$ supersymmetry in D=3. When M_8 is hyper-Kähler, under $Sp(2) \subset Spin(8)$ we have

$$\mathbf{8}_+ \to \mathbf{5} + \mathbf{1} + \mathbf{1} + \mathbf{1}, \qquad \mathbf{8}_- \to \mathbf{4} + \mathbf{4} \ , \tag{2.32}$$

and six supersymmetries are preserved, or $N = 3$ in D=3. Similarly, when M_8 is the product of two Calabi-Yau two-folds eight supersymmetries are preserved, or $N = 4$ in D=3. If we also allow tori, then when M_8 is the product of a Calabi-Yau two-fold with T^4, sixteen supersymmetries are preserved, or $N = 8$ in D=3, while the simple case of T^8 preserves all thirty-two supersymmetries, or $N = 16$ in D=3.

An important way to make contact with four-dimensional physics is to consider geometries of the form $\mathbb{R}^{1,3} \times M_7$ with M_7 compact. If we choose M_7 to be T^7 then it preserves all 32 supersymmetries or $N = 8$ supersymmetry in D=4 spacetime dimensions. $T^3 \times CY_2$ preserves 16 supersymmetries or $N = 4$ in D=4, $S^1 \times CY_3$ preserves 8 supersymmetries or $N = 2$ in D=4 and G_2 preserves four supersymmetries or $N = 1$ in D=4.

$\dim(M)$	Holonomy	Supersymmetry
10	$SU(5)$	2
10	$SU(3)\times SU(2)$	4
8	$Spin(7)$	2
8	$SU(4)$	4
8	$Sp(2)$	6
8	$SU(2)\times SU(2)$	8
7	G_2	4
6	$SU(3)$	8
4	$SU(2)$	16

TABLE 1. Manifolds of special holonomy and the corresponding amount of preserved supersymmetry.

$N=1$ supersymmetry in four spacetime dimensions has many attractive phenomenological features and this is the key reason for the recent interest in manifolds with G_2 holonomy. As discussed in Acharya's lectures at this school, it is important to emphasise that the most interesting examples from the physics point of view are not complete. In addition one can use non-compact G_2 holonomy manifolds very effectively to study various quantum field theories in four spacetime dimensions (see e.g. [**9**]).

Another important class of examples is to consider $d=7$ manifolds of the form $S^1/Z_2\times CY_3$ where the Z_2 action has two fixed planes. It can be shown that the orbifold breaks a further one-half of the supersymmetries and one is again left with four supersymmetries in four spacetime dimensions. These configurations are related to the the strongly coupled limit of heterotic string theory compactified on CY_3 [**93, 92**].

In summary, when $G=0$, geometries of the form $\mathbb{R}^{1,10-d}\times M_d$ preserve supersymmetry when M_d admits covariantly constant spinors and hence has special holonomy. For physical applications, M_d need be neither compact nor complete. In the next section we will consider the basic solutions with $G\neq 0$, the fivebrane and the membrane solutions.

2.2. Membranes and Fivebranes. The simplest, and arguably the most important supersymmetric solutions with non-vanishing four-form are the membrane and fivebrane solutions. Further discussion can be found in e.g. [**134**].

The fivebrane geometry is given by

$$\begin{aligned} ds^2 &= H^{-1/3}\left[d\xi^i d\xi^j\eta_{ij}\right]+H^{2/3}\left[dx^I dx^I\right] \\ G_{I_1I_2I_3I_4} &= -c\epsilon_{I_1I_2I_3I_4J}\partial_J H\ , \qquad c=\pm1\ , \end{aligned} \tag{2.33}$$

where $i,j=0,1,\dots 5$, $I=1,\dots,5$ and $H=H(x^I)$. This geometry preserves 1/2 of the supersymmetry. In the obvious orthonormal frame $\{H^{-1/6}d\xi^i, H^{1/3}dx^I\}$, the 16 Killing spinors are given by

$$\epsilon = H^{-1/12}\epsilon_0\ , \tag{2.34}$$

where ϵ_0 is a constant spinor, and satisfy

$$\Gamma^{012345}\epsilon = c\epsilon\ . \tag{2.35}$$

Since Γ^{012345} squares to unity and is traceless, we conclude that the geometry admits 16 independent Killing spinors.

This geometry satisfies the equations of motion providing that we impose the Bianchi identity for G which implies that H is harmonic. If we take H to have a single centre

$$H = 1 + \frac{\alpha_5 N}{r^3}, \qquad r^2 = x^I x^I \ , \tag{2.36}$$

with N positive and $\alpha_5 = \pi l_p^3$, then the solution carries cN units of quantised magnetic four-form flux

$$\frac{1}{(2\pi l_p)^3} \int_{S^4} G = cN \ , \tag{2.37}$$

with N a positive integer, consistent with (2.10). When $c = +1$ the solution describes N coincident fivebranes, that are oriented along the $d\xi^0 \wedge d\xi^1 \wedge \dots d\xi^5$ plane. When $c = -1$ the solution describes N coincident anti-fivebranes. Roughly speaking, the fivebranes can be thought of as being located at $r = 0$, where the solution appears singular. However, this is in fact a regular horizon and moreover, it is possible to analytically continue to obtain a completely non-singular geometry [**78**]. Thus it is not possible to say exactly where the fivebranes are located.

In the directions transverse to the fivebrane the metric becomes asymptotically flat. We can thus calculate the ADM mass per unit volume, or tension, and we find

$$Tension = NT_5 \ , \qquad T_5 = \frac{1}{(2\pi)^5 l_p^6} \ , \tag{2.38}$$

where T_5 is the tension of a single fivebrane (for a careful discussion of numerical coefficients appearing in T_5 and the membrane tension T_2 below, see [**46**]). It is possible to show that the supersymmetry algebra actually implies that the tension of the fivebranes is fixed by the magnetic charge. This "BPS" condition is equivalent to the geometry preserving 1/2 of the supersymmetry. Note also that if H is taken to be a multi-centred harmonic function then we obtain a solution with the N coincident fivebranes separated.

It is interesting to examine the near horizon limit of the geometry of N coincident fivebranes, when $r \approx 0$. By dropping the one from the harmonic function in (2.36) we get

$$ds^2 = \frac{r}{(\alpha_5 N)^{1/3}} \left[d\xi^i d\xi^j \eta_{ij} \right] + \frac{(\alpha_5 N)^{2/3}}{r^2} \left[dr^2 + r^2 d\Omega_4 \right] \ . \tag{2.39}$$

where $d\Omega_4$ is the metric on the round four-sphere. After a coordinate transformation we can rewrite this as

$$ds^2 = R^2 \left[\frac{d\xi^i d\xi^j \eta_{ij} + d\rho^2}{\rho^2} \right] + \frac{R^2}{4} d\Omega_4 \ , \tag{2.40}$$

which is just $AdS_7 \times S^4$, in Poincaré coordinates, with the radius of the AdS_7 given by

$$R = 2(\pi N)^{1/3} l_p \ . \tag{2.41}$$

There are still N units of flux on the four-sphere. This geometry is in fact a solution to the equations of motion that preserves all 32 supersymmetries. A closely related fact is that the Lorentz symmetry $SO(5,1)$ of the fivebrane solution has been enhanced to the conformal group $SO(6,2)$. The interpretation of this fact and the

$SO(5)$ isometries of the four-sphere will be discussed in the next section. Before doing so, we introduce the membrane solution.

The membrane geometry is given by

$$\begin{aligned} ds^2 &= H^{-2/3}\left[d\xi^i d\xi^j \eta_{ij}\right] + H^{1/3}\left[dx^I dx^I\right] \\ C &= cH^{-1} d\xi^0 \wedge d\xi^1 \wedge d\xi^3, \qquad c = \pm 1 \ , \end{aligned} \tag{2.42}$$

with, here, $i,j = 0,1,2$, $I = 1,\ldots 8$ and $H = H(x^I)$. This geometry preserves one half of the supersymmetry. Using the orthonormal frame $\{H^{-1/3}d\xi^i, H^{1/6}dx^I\}$ we find that the Killing spinors are given by

$$\epsilon = H^{-1/6}\epsilon_0 \ , \tag{2.43}$$

where ϵ_0 is a constant spinor, and satisfy the constraint

$$\Gamma^{012}\epsilon = c\epsilon \ . \tag{2.44}$$

Since Γ^{012} squares to unity and is traceless, we conclude that the geometry admits 16 independent Killing spinors.

This geometry solves all of the equations of motion providing that we impose the four-form equation of motion. This implies that the function H is harmonic. Now take H to be

$$H = 1 + \frac{\alpha_2 N}{r^6}, \qquad r^2 = x^I x^I \ , \tag{2.45}$$

with N a positive integer and $\alpha_2 = 32\pi^2 l_p^6$. The solution carries cN units of quantised electric four-form charge:

$$\frac{1}{(2\pi l_p)^6}\int_{S^7} *G = cN \ . \tag{2.46}$$

When $c = \pm 1$, the solution describes N coincident (anti-)membranes oriented along the $d\xi^0 \wedge d\xi^1 \wedge d\xi^3$ plane. Transverse to the membrane the solution tends to flat space, and we can thus determine the ADM tension of the membranes. We again find that it is related to the charge as dictated by supersymmetry

$$Tension = NT_2 \ , \qquad T_2 = \frac{1}{(2\pi)^2 l_p^3} \ , \tag{2.47}$$

where T_2 is the tension of a single membrane. If the harmonic function is replaced with a multi-centre harmonic function we obtain a solution with the membranes separated.

The solution describing N coincident membanes appears singular at $r \approx 0$, but one can in fact show that this is a horizon. The solution can be extended across the horizon and one finds a timelike singularity inside the horizon (see e.g. [**134**]), which can be mapped onto a membrane source with tension T_2. To obtain the near horizon geometry, $r \approx 0$, we drop the one in the harmonic function (2.45), to find, after a coordinate transformation,

$$ds^2 = R^2\left[\frac{d\xi^i d\xi^j \eta_{ij} + d\rho^2}{\rho^2}\right] + 4R^2 d\Omega_7 \ , \tag{2.48}$$

which is simply the direct product $AdS_4 \times S^7$ with the radius of AdS_4 given by

$$R = \left(\frac{N\pi^2}{2}\right)^{1/6} l_p \ . \tag{2.49}$$

There are still N units of flux on the seven-sphere. This configuration is itself a supersymmetric solution preserving all 32 supersymmetries. The $SO(2,1)$ Lorentz symmetry of the membrane solution has been enhanced to the conformal group $SO(3,2)$ and the seven-sphere admits an $SO(8)$ group of isometries.

This concludes our brief review of the basic planar membrane and fivebrane solutions. There is a whole range of more general solutions describing the intersection of planar membranes and fivebranes, and we refer to the reviews [**64, 133**] for further details.

3. AdS/CFT Correspondence

In the last section we saw that D=11 supergravity admits supersymmetric solutions corresponding to N coincident membranes or coincident fivebranes, and that in the near horizon limit the metrics become $AdS_4 \times S^7$ or $AdS_7 \times S^4$, respectively. The famous conjecture of Maldacena [**108**] states that M-theory on these backgrounds is equivalent to certain conformal field theories in three or six spacetime dimensions, respectively. For a comprehensive review of this topic, we refer to [**6**], but we would like to make a few comments in order to motivate the construction of the supersymmetric solutions of D=11 supergravity presented in later sections.

The best understood example of the AdS/CFT correspondence actually arises in type IIB string theory, so we first pause to introduce it. The low-energy limit of type IIB string theory is the chiral type IIB supergravity [**131, 97**]. The bosonic field content of the supergravity theory consists of a metric, a complex scalar, a complex three-form field strength and a self-dual five-form field strength. The theory admits a 1/2 supersymmetric three-brane, called a D3-brane. The metric of the corresponding supergravity solution is given by

$$ds^2 = H^{-1}\left[d\xi^i d\xi^j \eta_{ij}\right] + H\left[dx^I dx^I\right] \ , \tag{3.1}$$

where $i,j = 0,1,2,3$, $I,J = 1,\dots 6$ and $H = H(x^I)$ is a harmonic function. If we choose

$$H = 1 + \frac{\alpha_3 N}{r^2} \ , \tag{3.2}$$

with N a positive integer, and α_3 some constant with dimensions of length squared, then the solution corresponds to N coincident D3-branes. The only other non-vanishing field is the self-dual five-form and the solution, for suitably chosen α_3, carries N units of flux when integrated around a five-sphere surrounding the D3-branes. In the near horizon limit, $r \approx 0$, we get $AdS_5 \times S^5$, with equal radii.

The boundary of AdS_5 is the conformal compactification of four-dimensional Minkowski space, M_4. The AdS/CFT conjecture states that type IIB string theory on $AdS_5 \times S^5$ is equivalent (dual) to $\mathcal{N} = 4$ supersymmetric Yang-Mills theory with gauge group $SU(N)$ on M_4. This quantum field theory is very special as it has the maximal amount of supersymmetry that a quantum field theory can have. Moreover, it is a conformal field theory (CFT), i.e. invariant under the conformal group. The AdS/CFT correspondence relates parameters in the field theory with those of the string theory on $AdS_5 \times S^5$. It turns out that perturbative Yang-Mills theory can be a good description only when the radius R of the AdS_5 is small, while supergravity is a good approximation only when R is large and N is large. The fact that these different regimes don't overlap is a key reason why such seemingly different theories could be equivalent at all.

The natural objects to consider in a conformal field theory are correlation functions of operators. A precise dictionary between operators in the conformal field theory and fields (string modes) propagating in AdS_5 is given in [**84, 139**]. Moreover, in the supergravity approximation, correlation functions of the operators are determined by the dependence of the supergravity action on the asymptotic behaviour of the fields on the boundary. For example, the conformal dimensions of the operators is determined by the mass of the fields.

It remains very unclear how to prove the AdS/CFT conjecture. Nevertheless, it has now passed an enormous number of tests. Amongst the simplest is to compare the symmetries on the two sides. $\mathcal{N} = 4$ super Yang-Mills theory has an internal $SO(6)$ "R-symmetry" and is invariant under the conformal group in four dimensions, $SO(4,2)$. But $SO(4,2) \times SO(6)$ are precisely the isometries of $AdS_5 \times S^5$. Moreover, after including the supersymmetry, we find that both sides are invariant under the action of the supergroup $SU(2,2|4)$ whose bosonic subgroup is $SO(4,2) \times SO(6)$.

Let us now return to the near horizon geometries of the membrane and fivebrane. For the membrane, it is conjectured that M-theory on $AdS_4 \times S^7$ with N units of flux on the seven-sphere is equivalent to a maximally supersymmetric conformal field theory on the conformal compactification of three-dimensional Minkowski space, the boundary of AdS_4. More precisely this conformal field theory is the infrared (low energy) limit of $N = 8$ super-Yang-Mills theory with gauge group $SU(N)$ in three dimensions. It is known that this theory has an $SO(8)$ R-symmetry. For this case, the $SO(3,2) \times SO(8)$ isometries of $AdS_4 \times S^7$ correspond to the conformal invariance and the R-symmetry of the conformal field theory. After including supersymmetry we find that both sides are invariant under the supergroup $OSp(8|4)$.

For the fivebrane, it is conjectured that M-theory on $AdS_7 \times S^4$ with N units of flux on the four-sphere is equivalent to a maximally supersymmetric chiral conformal field theory on the conformal compactification of six-dimensional Minkowski space, the boundary of AdS_7. This conformal field theory is still rather mysterious and the AdS/CFT correspondence actually provides a lot of useful information about it (assuming the correspondence is valid!). The $SO(6,2) \times SO(5)$ isometries of $AdS_7 \times S^4$ correspond to the conformal invariance and the $SO(5)$ R-symmetry of the field theory. After including supersymmetry we find that both sides are invariant under the supergroup $OSp(6,2|4)$. For the membrane and fivebrane examples, when N is large, the radius of the AdS spaces are large and M-theory is well approximated by D=11 supergravity.

Much effort has been devoted to further understanding and generalising the AdS/CFT correspondence. Let us briefly discuss some of the generalisations that have been pursued, partly to put the solutions we will construct later into some kind of context, and partly as a rough guide to some of the vast literature on the subject.

The three basic examples of the AdS/CFT correspondence relate string/M-theory on $AdS_{d+1} \times \mathit{sphere}$ geometries to conformally invariant quantum field theories in $d = 3, 4$ and 6 with maximal supersymmetry. One direction is to find new supersymmetric solutions of supergravity theories that are the products, possibly warped[1], of AdS_{d+1} with other compact spaces, preserving less than maximal supersymmetry. Since the isometry group of AdS_{d+1} is the conformal group, this indicates that these would be dual to new superconformal field theories in d spacetime dimensions. Non-supersymmetric solutions with AdS factors are also of interest, as they could be dual to non-supersymmetric CFTs. However, one has to check whether the solutions are stable at both the perturbative and non-perturbative level, which is very difficult in general. By contrast, in the supersymmetric case, stability is guaranteed from the supersymmetry algebra. The focus has thus been on supersymmetric geometries with AdS factors.

One class of examples, discussed in [**103, 2, 114**], is to start with the fivebrane, membrane or D3-brane geometry (2.33), (2.42) or (3.1), respectively, and observe that if the flat space transverse to the brane is replaced by a manifold with special holonomy as in table 1, and the associated flux left unchanged, then the resulting solution will still preserve supersymmetry but will preserve, in general, a reduced amount. Now let the special holonomy manifold be a cone over a base X, i.e. let the metric of the transverse space be

$$dr^2 + r^2 ds^2(X) \ , \tag{3.3}$$

X must be Einstein and have additional well known properties to ensure that the metric has special holonomy. For example a five-dimensional X should be Einstein-Sasaki in order that the six-dimensional cone is Calabi-Yau. Apart from the special case when X is the round sphere these spaces have a conical singularity at $r = 0$. To illustrate this construction explicitly for the membrane, one replaces the eight-dimensional flat space transverse to the membrane in (2.42) with an eight-dimensional cone with special holonomy:

$$ds^2 = H^{-2/3}\left[d\xi^i d\xi^j \eta_{ij}\right] + H^{1/3}\left[dr^2 + r^2 ds^2(X)\right] \ , \tag{3.4}$$

where $H = 1 + \alpha_2 N/r^6$, as before. Clearly this can be interpreted as N coincident membranes sitting at the conical singularity. By considering the near horizon limit of (3.4), $r \approx 0$, one finds that it is now the direct product $AdS_4 \times X$ and this provides a rich class of new AdS/CFT examples.

Another generalisation is to exploit the fact that the maximally supersymmetric conformal field theories can be perturbed by certain operators. In some cases, under renormalisation group flow, these quantum theories will flow in the infrared (low energies) to new superconformal field theories, with less supersymmetry. It is remarkable that corresponding dual supergravity solutions can be found. Given the dictionary between operators in the conformal field theory in d dimensions and fields in AdS_{d+1} mentioned above, the perturbation of the conformal field theory should correspond to dual supergravity solutions that asymptotically tend to AdS_{d+1} in a prescribed way. Now on rather general grounds it can be argued that this AdS_{d+1} boundary corresponds to the ultraviolet (UV) of the dual perturbed conformal field theory, and that going away from the boundary into the interior corresponds to going to the infrared (IR) in the dual quantum field theory [**136**]. This can be seen, for example, by studying the action of the conformal group on AdS_{d+1} and on the correlation functions in the d-dimensional conformal field theory. Thus if the perturbed conformal field theory is flowing to another conformal field theory

[1] A warped product of two spaces with coordinates x and y corresponds to a metric of the form $f(y)ds^2(x) + ds^2(y)$, for some function $f(y)$.

in the IR, we expect that there should be supergravity solutions that interpolate from the perturbed AdS boundary to another AdS region in the interior. Indeed such solutions can be found (see for example [**102, 59, 126**]).

The above examples concern gravity duals of superconformal field theories or flows between superconformal field theories. Another way to generalise the correspondence is to find supergravity solutions that are dual to non-conformal field theories. For example, one might study perturbed superconformal field theories that flow in the infrared to non-conformal phases, such as Coulomb, Higgs and confining phases. The corresponding dual supergravity solutions should still have an asymptotic AdS_{d+1} boundary, describing the perturbed conformal field theory, but they will no longer interpolate to another AdS_{d+1} region but to different kinds of geometry dual to the different phases. For an example of these kinds of solutions see [**125**]. Often the geometries found in the IR are singular and further analysis is required to determine the physical interpretation.

The generalisation we will discuss in the rest of the lectures was initiated by Maldacena and Nuñez [**110**]. The idea is to construct supergravity solutions describing branes wrapping calibrated cycles in manifolds of special holonomy, in the near horizon limit. The next section will explain the background for attempting this, and in particular why such configurations preserve supersymmetry. The subsequent section will describe the construction of the solutions using the technical tool of gauged supergravity. The D=11 solutions describing wrapped membranes and fivebranes and the D=10 solutions describing wrapped D3-branes provide a large class of solutions with dual field theory interpretations. The simplest, and perhaps most important, solutions are warped products of AdS spaces, cycles with Einstein metrics and spheres. The presence of the AdS factor for these solutions implies that they provide a large class of new AdS/CFT examples. In addition, as we shall discuss, there are more complicated solutions that describe flows from a perturbed AdS boundary, describing the UV, to both conformal and non-conformal behaviour in the IR.

Type IIB string theory also contains NS fivebranes. In the discussion section we will briefly discuss how supergravity solutions describing NS fivebranes wrapped on calibrated cycles can be used to study non-conformal field theories. A particularly interesting solution [**111**] (see also [**33**]) gives rise to a dual quantum field theory that has many features of $\mathcal{N} = 1$ supersymmetric Yang-Mills theory in four dimensions. This is a very interesting theory as it has many features that are similar to QCD. Two other interesting ways of studying $\mathcal{N} = 1$ supersymmetric Yang-Mills theory in four dimensions can be found in [**104**] and [**128**]. Related constructions with $\mathcal{N} = 2$ supersymmetry can be found in [**69, 22, 20, 127**] (for reviews see [**5, 19, 21**]). The explicit regular solutions found in [**104**] are examples of a more general construction discussed in [**42**] (see [**40**] for a review).

4. Brane worldvolumes and calibrations

The supergravity brane solutions that were presented in section 2 describe static planar branes of finite tension and infinite extent. Physical intuition suggests that these branes should become dynamical objects if they are perturbed. Moreover, we also expect that branes with different topologies should exist. On the other hand it is extremely difficult to study these aspects of branes purely from the supergravity point of view. Luckily, there are alternative descriptions of branes which can be used. In this section we will describe the low-energy world-volume description of branes. Essentially, this is a probe-approximation in which the branes are taken to be very light and hence propagate in a fixed background geometry with no back-reaction. We will use this description to argue that branes can wrap calibrated cycles in manifolds of special holonomy while preserving supersymmetry [**12, 13, 73, 79**]. This then provides the motivation to seek D=11 supergravity solutions that describe a large number of such wrapped branes, when the back-reaction on the geometry will be very significant. We will construct the solutions in the next section in the near horizon limit, which is the limit relevant for AdS/CFT applications.

It will be useful to first review some background material concerning calibrations on manifolds of special holonomy, before turning to the brane world-volume theories.

4.1. Calibrations. A calibration [**90**] on a Riemannian manifold M is a p-form φ satisfying two conditions:

$$\begin{aligned} d\varphi &= 0 \\ \varphi|_{\xi^p} &\leq Vol|_{\xi^p}, \qquad \forall \xi^p \ , \end{aligned} \tag{4.1}$$

where ξ_p is any tangent p-plane, and Vol is the volume form on the cycle induced from the metric on M. A p-cycle Σ_p is calibrated by φ if it satisfies

$$\varphi|_{\Sigma_p} = Vol|_{\Sigma_p} \ . \tag{4.2}$$

A key feature of calibrated cycles is that they are minimal surfaces in their homology class. The proof is very simple. Consider another cycle Σ' such that $\Sigma - \Sigma'$ is the boundary of a $(p+1)$-dimensional manifold Ξ. We then have

$$Vol(\Sigma) = \int_\Sigma \varphi = \int_\Xi d\varphi + \int_{\Sigma'} \varphi = \int_{\Sigma'} \varphi \leq Vol(\Sigma') \ . \tag{4.3}$$

The first equality is due to Σ being calibrated. The second equality uses Stokes' theorem. The remaining steps use the closure of φ and the second part of the definition of a calibration.

We will only be interested in calibrations that can be constructed as bilinears of spinors, for reasons that will soon become clear. The general procedure for such a construction was first discussed in [**44, 89**]. In fact all of the special holonomy manifolds that we discussed earlier have such calibrations. We now summarise the various cases, noting that the the spinorial construction and the closure of the calibrations was already presented in section 2. That the calibrations also satisfy the second condition in (4.1) was shown, for almost all cases, in [**90**]; it is also straightforward to establish using the spinorial construction.

On $Spin(7)$-holonomy manifolds the Cayley four-form Ψ is a calibration and the 4-cycles calibrated by Ψ are called Cayley 4-cycles.

G_2-holonomy manifolds have two types of calibrations, ϕ and $*\phi$. The former calibrates associative 3-cycles, while the latter calibrates co-associative 4-cycles.

Calabi-Yau n-folds generically have two classes of calibrations. The first class is the Kähler calibrations given by $\frac{1}{n!}J^n$, where the wedge product is used. These calibrate even $2n$-dimensional cycles and this is equivalent to the cycles being holomorphic. The second type of calibration is the special Lagrangian (SLAG) calibration given by the real part of the holomorphic n-form $e^{i\theta}\Omega$, where the constant $\theta \in S^1$, and these calibrate special Lagrangian n-cycles. Recall that for our purposes

$n = 2, 3, 4, 5$. When $n = 2$, there is no real distinction between SLAG and Kähler 2-cycles since the cycles that are Kähler with respect to one complex structure are SLAG with respect to another (recall that $CY_2 = HK_1$). When $n = 4$ there are also 4-cycles that are calibrated by $\frac{1}{2}J^2 + Re(e^{i\theta}\Omega)$ – these are in fact Cayley 4-cycles if we view the Calabi-Yau four fold as a special example of a $Spin(7)$-manifold.

Hyper-Kähler manifolds in eight dimensions are special cases of Calabi-Yau four-folds. They thus admit Kähler and special Lagrangian calibrations with respect to each complex structure. They also admit Cayley calibrations as just described. In addition there are quaternionic calibrations [**45**] that calibrate quaternionic 4-cycles which are Kähler with respect to all three complex structures: $Vol = \frac{1}{2}(J^1)^2 = \frac{1}{2}(J^2)^2 = \frac{1}{2}(J^3)^2$, when restricted to the cycle. For example, with respect to the hyper-Kähler structure (2.26), we see that e^{1234} is a quaternionic 4-cycle[2]. Of more interest to us will be the complex-Lagrangian ($\mathbb{C}$-Lag) calibrations [**45**] which calibrate 4-cycles that are Kähler with respect to one complex structure and special Lagrangian with respect to the other two: for example, $Vol = \frac{1}{2}(J^1)^2 = Re(\Omega^2) = -Re(\Omega^3)$ when restricted to the cycle. Referring to (2.26) and (2.27) we see that e^{1256} is an example of such a $\mathbb{C}$-Lag 4-cycle.

In constructing supergravity solutions describing branes wrapping calibrated cycles in the next section, it will be very important to understand the structure of the normal bundle of calibrated cycles. Let us summarise some results of Mclean [**113**]. The tangent bundle of the special holonomy manifold restricted to the cycle splits into the tangent bundle of the cycle plus the normal bundle

$$T(M)|_\Sigma = T(\Sigma) \oplus N(\Sigma) \ . \tag{4.4}$$

In some, but not all cases, the normal bundle, $N(\Sigma)$, is intrinsic to Σ. Given a calibrated cycle, one can also ask which normal deformation, if any, is a normal deformation through a family of calibrated cycles.

A simple case to describe are the special Lagrangian cycles, where $N(\Sigma)$ is intrinsic to Σ. It is not difficult to show that on a special Lagrangian cycle the Kähler form, J, restricted to Σ vanishes. Thus, for any vector field V on Σ we have that $J_{ij}V^j$ are the components of a one-form on Σ which is orthogonal to all vectors on Σ. In other words, $J^i{}_jV^j$ defines a normal vector field. Hence $N(\Sigma)$ is isomorphic to $T(\Sigma)$.

In addition, the normal deformation described by the vector V is a normal deformation through the space of special Lagrangian submanifolds if and only if the one-form with components $J_{ij}V^j$ is harmonic. Thus if Σ is compact, the dimension of the moduli space of special Lagrangian manifolds near Σ is given by the first Betti number, $\beta^1(\Sigma) = dim[H^1(\Sigma, \mathbb{R})]$. In particular if $\beta^1 = 0$, then Σ has no harmonic one-forms and hence it is rigid as a special Lagrangian submanifold.

Next consider co-associative 4-cycles in manifolds of G_2 holonomy, for which $N(\Sigma)$ is also intrinsic to Σ. In fact $N(\Sigma)$ is isomorphic to the bundle of anti-self-dual two-forms on Σ. A normal vector field is a deformation through a family of co-associative 4-cycles if and only if the corresponding anti-self-dual two-form is closed and hence harmonic. Thus if Σ is compact the dimension of the moduli space of co-associative 4-cycles near Σ is given by the Betti number, $\beta^2_-(\Sigma) = dim[H^2_-(\Sigma, \mathbb{R})]$. In particular if $\beta^2_- = 0$, then Σ is rigid as a co-associative submanifold.

[2]Supergravity solutions describing fivebranes wrapping quaternionic 4-cycles in $\mathbb{R}^8$, which are necessarily linear [**45**], were constructed in [**120, 121**].

The normal bundles of associative 3-cycles in manifolds of G_2 holonomy are not intrinsic to Σ in general. The normal bundle is given by $S \otimes V$ where S is the spin bundle of Σ (oriented three-manifolds are always spin) and V is a rank two $SU(2)$ bundle. In other words the normal directions are specified by two-dimensional spinors on Σ that carry an additional $SU(2)$ index. A normal vector field gives a deformation through a family of associative 3-cycles if and only if the corresponding twisted spinor is harmonic, i.e. in the kernel of the twisted Dirac operator. In the special case that the bundle V is trivial, the spinor must be harmonic. For example, if Σ is an associative three-sphere and V is trivial, as in the G_2 manifold constructed in [**28**], then it is rigid.

The deformation theory of Cayley 4-cycles in manifolds of $Spin(7)$ holonomy has a similar flavour to the associative 3-cycles. The normal bundle is given by $S_- \otimes V$ where S_- is the bundle of spinors of negative chirality on Σ and V is a rank two $SU(2)$ bundle. Although not all 4-cycles admit a spin structure, all Cayley 4-cycles admit such twisted spinors. A normal vector field gives a deformation through a family of Cayley 4-cycles if and only if the corresponding twisted spinor is harmonic, i.e. in the kernel of the twisted Dirac operator. In the special case that the bundle V is trivial the spinor must be harmonic. For example, if Σ is a Cayley four-sphere and V is trivial, as in the $Spin(7)$ manifold constructed in [**28**], then it is rigid.

Finally, let us make some comments concerning the Kähler cycles. These cycles reside in Calabi-Yau manifolds M which have vanishing first Chern class, $c_1[T(M)] = 0$. Since $c_1[T(M)|_\Sigma] = c_1[T(\Sigma)] + c_1[N(\Sigma)]$ we conclude that in general

$$c_1[N(\Sigma)] = -c_1[T(\Sigma)] \ . \tag{4.5}$$

If one considers the special case that Σ is a divisor i.e. a complex hypersurface (i.e. real codimension two), then $N(\Sigma)$ is intrinsic to Σ. Indeed one can show that $N(\Sigma) \cong K(\Sigma)$ where $K(\Sigma)$ is the canonical bundle of Σ.

4.2. Membrane world-volume theory. Let us now turn to the world-volume theory of branes, beginning with membranes [**15, 16**]. We will consider the membranes to be propagating in a fixed D=11 geometry which is taken to be a bosonic solution to the equations of motion of D=11 supergravity, with metric g and three-form C. The bosonic dynamical fields are maps $x^\mu(\sigma)$ from the world-volume of the membrane, W, to the D=11 target space geometry. If we let σ^i be coordinates on W with $i = 0, 1, 2$, and x^μ be coordinates on the D=11 target geometry with $\mu = 0, 1, \ldots, 10$, the reparametrisation invariant action is given by

$$S = T_2 \int_W d^3\sigma \left[-det\ \partial_i x^\mu \partial_j x^\nu g_{\mu\nu}(x)\right]^{1/2} + \frac{1}{3!}\epsilon^{ijk}\partial_i x^{\mu_1}\partial_j x^{\mu_2}\partial_k x^{\mu_3} C_{\mu_1\mu_2\mu_3} \ . \tag{4.6}$$

The first term is just the volume element of the pullback of the metric to the world-volume and is called the Nambu-Goto action. The second term arises because the membrane carries electric four-form charge; it generalises the coupling of an electrically charged particle to a vector potential. The full action also includes fermions and is invariant under supersymmetry when the D=11 target admits Killing spinors. The supersymmetry of brane-world volume theories is actually quite intricate, but luckily we will not need many of the details. The reason is similar to the reason that we didn't need to discuss such details for D=11 supergravity. Once again, our interest is bosonic solutions to the equations of motion that preserve some supersymmetry. For such configurations, the supersymmetry variation of the bosonic fields automatically vanishes, and hence one only needs to know the supersymmetry variation of the fermions, and this will be mentioned later.

To get some further insight, consider static bosonic D=11 backgrounds with vanishing three-form, $C = 0$,and write the metric as

$$ds^2 = -dt^2 + g_{MN} dx^M dx^N \ , \tag{4.7}$$

where $M, N = 1, 2, \ldots, 10$. If we substitute this into (4.6) and partially fix the reparametrisation invariance by choosing $\sigma^0 = t$ the membrane action gives rise to the energy functional

$$E = T_2 \int_{W'} d^2\sigma \, [m_{ab}]^{1/2} \ , \tag{4.8}$$

where $a, b = 1, 2$, W' is the spatial part of the world-volume, and m_{ab} is the spatial part of the induced world-volume metric given by

$$m_{ab} = \partial_a x^M \partial_b x^N g_{MN} \ . \tag{4.9}$$

In other words, the energy is just given by the tension of the membrane times the spatial area of the membrane. Now, static solutions to the equations of motion minimise the energy functional. Thus static configurations minimise the area of the membrane, which implies that the spatial part of the membrane is a minimal surface. This is entirely in accord with expectations: the tension of the membrane tends to make it shrink. It should be noted that the minimal surfaces can be of infinite extent: the simplest example is an infinite flat membrane in D=11 Minkowski space. Of most interest to us will be membranes wrapping compact minimal surfaces.

Let us further restrict to background geometries of the form $\mathbb{R}^{1,10-d} \times M_d$ with vanishing three-form that preserve supersymmetry. In other words M_d has special holonomy as discussed in section 2. The membrane world-volume theory is supersymmetric with the number of supersymmetries determined by the number of Killing spinors. Static membrane configurations that preserve supersymmetry wrap cycles called supersymmetric cycles. We now argue that supersymmetric cycles are equivalent to calibrated cycles with the associated calibration being constructed from the Killing spinors.

In order that a bosonic world-volume configuration be supersymmetric the supersymmetry variation of the fermions must vanish. Given the explicit supersymmetry variations, it is simple to show this implies that [**12**]

$$(1 - \Gamma)\epsilon = 0 \ , \tag{4.10}$$

where ϵ is a D=11 Killing spinor and the matrix Γ is given by

$$\begin{aligned} \Gamma &= \frac{1}{\sqrt{\det m}} \Gamma^0 \gamma \\ \gamma &= \left(\frac{1}{2!} \epsilon^{ab} \partial_a x^M \partial_b x^N \Gamma_{MN} \right) \ , \end{aligned} \tag{4.11}$$

where $[\Gamma_M, \Gamma_N]_+ = 2g_{MN}$. The matrix Γ satisfies $\Gamma^2 = 1$ and is Hermitian, $\Gamma^\dagger = \Gamma$. We now calculate

$$\epsilon^\dagger \frac{(1-\Gamma)}{2} \epsilon = \epsilon^\dagger \frac{(1-\Gamma)}{2} \frac{(1-\Gamma)}{2} \epsilon = || \frac{(1-\Gamma)}{2} \epsilon ||^2 \geq 0 \ . \tag{4.12}$$

We thus conclude that $\epsilon^\dagger\epsilon \geq \epsilon^\dagger\Gamma\epsilon$ with equality if and only $(1-\Gamma)\epsilon = 0$, which is equivalent to the configuration being supersymmetric. The inequality can be rewritten

$$\sqrt{\det m} \geq \epsilon^\dagger\Gamma^0\gamma\epsilon = -\bar{\epsilon}\gamma\epsilon \; . \tag{4.13}$$

Thus the two-form defined by

$$\varphi = -\frac{1}{2!}\bar{\epsilon}\Gamma_{MN}\epsilon dx^M \wedge dx^N \;, \tag{4.14}$$

satisfies the second condition in (4.1) required for a calibration. One can argue that the supersymmetry algebra [**88**] implies that it is closed and hence is in fact a calibration (we will verify this directly in a moment). Moreover, the inequality is saturated if and only if the membrane is wrapping a supersymmetric cycle, and we see that this is equivalent to the cycle being calibrated by (4.14).

The only two-form calibrations on special holonomy backgrounds are Kähler calibrations, and indeed φ is in fact equal to a Kähler two-form on the background. To see this very explicitly and to see how much supersymmetry is preserved when a membrane wraps a Kähler 2-cycle, first consider the D=11 background to be $\mathbb{R}\times CY_5$. We noted earlier that this background preserves two D=11 supersymmetries: in a suitable orthonormal frame, the two covariantly constant D=11 spinors can be taken to satisfy the projections (see, for example, the discussion in appendix B of [**75**]):

$$\Gamma^{1234}\epsilon = \Gamma^{3456}\epsilon = \Gamma^{5678}\epsilon = -\Gamma^{78910}\epsilon = -\epsilon \; . \tag{4.15}$$

Note that these imply that $\Gamma^{012}\epsilon = \epsilon$. Substituting either of these spinors into (4.14) we find that φ is precisely the Kähler calibration on CY_5:

$$\varphi = J = e^{12} + e^{34} + e^{56} + e^{78} - e^{910} \; . \tag{4.16}$$

Consider now a membrane wrapping a Kähler 2-cycle in CY_5, i.e. its worldvolume is $\mathbb{R}\times\Sigma$ with $\Sigma \subset CY_5$. To be concrete, consider $Vol(\Sigma) = e^{12}|_\Sigma$. We then find that the supersymmetry condition (4.10) implies that $\Gamma^{012}\epsilon = \epsilon$, which is precisely the projection on the spinors that we saw in the supergravity solution for the membrane (2.44). For this case we see that this projection does not constrain the two supersymmetries satisfying (4.15) further and thus a membrane can wrap a Kähler 2-cycle in a CY_5 "for free". Clearly if we wrapped an anti-membrane, satisfying $\Gamma^{012}\epsilon = -\epsilon$, there would be no surviving supersymmetry[3]. Let us now consider the background to be $\mathbb{R}\times CY_4\times\mathbb{R}^2$. This preserves four supersymmetries satisfying projections which we can take to be

$$\Gamma^{1234}\epsilon = \Gamma^{3456}\epsilon = \Gamma^{5678}\epsilon = \mp\Gamma^{78910}\epsilon = -\epsilon \; . \tag{4.17}$$

Two of these satisfy $\Gamma^{012}\epsilon = \epsilon$ and two satisfy $\Gamma^{012}\epsilon = -\epsilon$. After substituting into (4.14) they give rise to two Kähler forms on $CY_4\times\mathbb{R}^2$:

$$\varphi = J = e^{12} + e^{34} + e^{56} + e^{78} \mp e^{910} \; . \tag{4.18}$$

If we now wrap the membrane on a Kähler 2-cycle with $Vol(\Sigma) = e^{12}|_\Sigma$, then we see that the supersymmetry condition (4.10), $\Gamma^{012}\epsilon = \epsilon$, preserves two of the supersymmetries. Similarly, if we wrapped an anti-membrane satisfying $\Gamma^{012}\epsilon = -\epsilon$ it would also preserve two supersymmetries.

[3] Note that if we change the orientation by switching $e^{10} \to -e^{10}$, then (4.15) would assume a more symmetric form and we would find that we could wrap an anti-membrane along Σ for free.

The amount of supersymmetry preserved by any brane wrapping a calibrated cycle in a special holonomy background can be worked out in a similar way: one considers a convenient set of projections for the background geometry and then supplements them with those of the wrapped brane (or anti-brane). In almost all cases, wrapping the brane breaks 1/2 of the supersymmetries preserved by the special holonomy background. We have summarised the possibilities for the membrane in table 2.

Calibration	World-Volume	Supersymmetry
Kähler	$\mathbb{R} \times (\Sigma_2 \subset CY_2)$	8
	$\mathbb{R} \times (\Sigma_2 \subset CY_3)$	4
	$\mathbb{R} \times (\Sigma_2 \subset CY_4)$	2
	$\mathbb{R} \times (\Sigma_2 \subset CY_5)$	2

TABLE 2. The different ways in which membranes can wrap calibrated cycles and the amount of supersymmetry preserved.

The action (4.6) describes the dynamics of a membrane propagating in a fixed D=11 supergravity background. Such a membrane is often called a "probe membrane". Of course, the dynamics of the membrane will back-react on the geometry, and so one should really supplement the D=11 supergravity action with the world-volume action:

$$S = S_{D=11} + S_{WV} \ . \tag{4.19}$$

If there are many coincident membranes then this back-reaction could be large.

We have been emphasising the geometric aspects of the membrane world-volume theory. The world-volume theory is also a quantum field theory. To gain some insight into this aspect, let us restrict the target geometry to be D=11 Minkowski space and the world-volume to be $\mathbb{R}^{1,2}$. Now fix the reparametrisation invariance completely by setting $\sigma^0 = t, \sigma^1 = x^1, \sigma^2 = x^2$. We can then expand the determinant to get

$$S = T_2 \int d^3\sigma (-\frac{1}{2}\partial_a x^I \partial^a x^I + fermions + \dots) \ , \tag{4.20}$$

where we have dropped an infinite constant and the dots refer to higher derivative terms. The eight scalar fields describe the eight transverse fluctuations of the membrane. After quantisation, this action gives a three-dimensional quantum field theory, with eight scalar fields plus fermions, that preserves 16 supersymmetries or $N = 8$ supersymmetry in three dimensions. This quantum field theory is interacting with gravity, via (4.19), but if we take the limit, $l_p \to 0$, it decouples from gravity. In other words, in this decoupling limit we get a three-dimensional quantum field theory living on the world-volume of the membrane. When there are N coincident branes, the quantum field theory is much more complicated. There is a piece describing the centre of mass dynamics of the branes given by (4.20) with $T_2 \to NT_2$ and there is another piece describing the interactions between the membranes. This latter theory is known to be a superconformal field theory that arises as the IR limit of $N = 8$ supersymmetric Yang-Mills theory in three dimensions. Recall that this is precisely the superconformal field theory that is conjectured to be dual to M-theory on $AdS_4 \times S^7$.

The important message here is that the supergravity solution describing the membranes in the near horizon limit, $AdS_4 \times S^7$, is conjectured to be equivalent to the quantum field theory arising on the membrane world-volume theory, in a limit which decouples gravity.

Now consider a more complicated example. Take the D=11 background to be of the form $\mathbb{R}^{1,6} \times CY_2$ with a probe membrane wrapping a Kähler 2-cycle $\Sigma \subset CY_2$. i.e. the world-volume of the membrane is $\mathbb{R} \times (\Sigma \subset CY_2)$. There is again a quantum field theory living on the brane interacting with gravity. In the decoupling limit, $l_p \to 0$ and keeping the volume of Σ fixed, we get a supersymmetric quantum field theory on $\mathbb{R} \times \Sigma$. If Σ is compact, the low-energy infrared (IR) limit of this quantum field theory corresponds to length scales much larger than the size of Σ. In this IR limit the quantum field theory on $\mathbb{R} \times \Sigma$ behaves like a quantum field theory on the time direction $\mathbb{R}$, which is just a quantum mechanical model.

If we could construct a supergravity solution describing membranes wrapping such Kähler 2-cycles, in the near horizon limit, we would have an excellent candidate for an M-theory dual for this quantum field theory on $\mathbb{R} \times \Sigma$. Moreover, if the supergravity solution has an AdS_2 factor, it would strongly indicate that the corresponding dual quantum mechanics, arising in the IR limit, is a superconformal quantum mechanics. These kinds of supergravity solutions have been found [**70**] and the construction will be described in the next section.

It is worth making some further comments about the quantum field theory on $\mathbb{R} \times \Sigma$. For a single membrane the physical bosonic degrees of freedom describe the transverse deformations of the membrane. In the case of a membrane with world-volume $\mathbb{R}^{1,2}$ in $\mathbb{R}^{1,10}$ we saw above in (4.20) that there are eight scalar fields describing these deformations. Geometrically, they are sections of the normal bundle, which is trivial in this case. Now consider, for example, a membrane with world-volume $\mathbb{R} \times (\Sigma \subset CY_2)$. There are six directions transverse to the membrane that are also transverse to the CY_2 and these lead to six scalar fields. There are also two directions transverse to the membrane that are tangent to the CY_2: these give rise to a section of the normal bundle. As we discussed earlier the normal deformations of a Kähler 2-cycle $\Sigma \subset CY_2$ (which are also SLAG 2-cycles with respect to another complex structure) are specified by one-forms on Σ.

This "transition" from scalars to one-forms is intimately connected with the way in which the field theory on $\mathbb{R} \times \Sigma$ realises supersymmetry. In particular it arises because the theory is coupled to external R-symmetry gauge fields. We will discuss this issue again in the context of fivebranes wrapping SLAG 3-cycles, as this is the example we will focus on when we construct the supergravity solutions in the next section.

4.3. D3-brane and fivebrane world-volume theories. Let us now briefly discuss the world-volume theories of the type IIB D3-brane and the M-theory fivebrane. The D3-brane action is given by a Dirac-Born-Infeld type action that includes a coupling to a four-form potential whose field strength is the self-dual five-form (see Myers' lectures for further discussion). If we consider, for simplicity, a bosonic type IIB background with all fields vanishing except for the metric, the world-volume action for a single D3-brane, with fermions set to zero, is given by

$$S = T_3 \int_W d^4\sigma \left[-det(\partial_i x^\mu \partial_j x^\nu g_{\mu\nu}(x) + F_{ij})\right]^{1/2} \ . \tag{4.21}$$

The main new feature is that, in addition to the world-volume fields x^μ, there is now a $U(1)$ gauge field with field strength F.

If we set $F = 0$, the action (4.21) reduces to the Nambu-Goto action. Following a similar analysis to that of the membrane, we again find in static supersymmetric backgrounds of the form $\mathbb{R}^{1,10-d} \times M_d$ that supersymmetric cycles with $F = 0$ are calibrated cycles. As the D3-brane has three spatial world-volume directions, there are now more possibilities. A D3-brane can either wrap a calibrated 3-cycle in M_d, with world-volume $\mathbb{R} \times (\Sigma_3 \subset M_d)$ or a Kähler 2-cycle in CY_n with world-volume $\mathbb{R}^{1,1} \times (\Sigma_2 \subset CY_n)$. The possibilities with the amount of supersymmetry preserved are presented in table 3. For the Kähler cases, the world-volume has an $\mathbb{R}^{1,1}$ factor and we have also denoted by (n_+, n_-) the amount of $d = 2$ supersymmetry preserved on the $\mathbb{R}^{1,1}$ factor, where n_+ is the number of chiral supersymmetries and n_- the number of anti-chiral supersymmetries.

Calibration	World-Volume	Supersymmetry
Kähler	$\mathbb{R}^{1,1} \times (\Sigma_2 \subset CY_2)$	8, (4,4) d=2
	$\mathbb{R}^{1,1} \times (\Sigma_2 \subset CY_3)$	4, (2,2) d=2
	$\mathbb{R}^{1,1} \times (\Sigma_2 \subset CY_4)$	2, (1,1) d=2
SLAG	$\mathbb{R} \times (\Sigma_3 \subset CY_3)$	4
Associative	$\mathbb{R} \times (\Sigma_3 \subset G_2)$	2

Table 3. The different ways in which D3-branes can wrap calibrated cycles and the amount of supersymmetry preserved.

In order to get some insight into the field theory living on D3-branes, consider the target to be D=10 Minkowski space-time, the world-volume to be $\mathbb{R}^{1,3}$ and fix the reparametrisation invariance by setting $\sigma^0 = t, \sigma^1 = x^1, \sigma^2 = x^2, \sigma^3 = x^3$. After expanding the determinant and dropping a constant term, we get

$$S = T_3 \int d^4\sigma \left(-\frac{1}{2}\partial_i x^I \partial^i x^I - \frac{1}{4} F_{ij} F^{ij} + fermions + \dots\right) \ . \tag{4.22}$$

In addition to F there are six scalar fields that describe the transverse displacement of the D3-brane. This action is simply $N = 4$ super-Yang-Mills theory with gauge group $U(1)$. When there are N D3-branes, it is known that the DBI action should be replaced by a non-Abelian generalisation but its precise form is not yet known. However, it is known that after decoupling gravity, the leading terms give $U(N)$ $\mathcal{N} = 4$ super-Yang-Mills theory. After dropping the $U(1)$ centre of mass piece, we find $\mathcal{N} = 4$ $SU(N)$ Yang-Mills theory. Recall that this is the theory that is conjectured to be dual to type IIB string theory on $AdS_5 \times S^5$, which is the near horizon limit of the type IIB supergravity solution describing a planar D3-brane. Once again, the near horizon limit of the supergravity solution is dual to the field theory arising on the brane, and the same should apply to the near horizon limits of the supergravity solutions describing the wrapped D3-branes in table 3.

The world-volume theory of M-theory fivebranes is arguably the most intricate of all branes [**95, 96, 122, 10, 17**]. The bosonic dynamical fields are maps $x^\mu(\sigma)$ along with a three-form field strength H_{ijk} that satisfies a non-linear self-duality condition. If we set $H = 0$ the dynamics is described by the Nambu-Goto action and we again find in a static supersymmetric background of the form $\mathbb{R}^{1,10-d} \times M_d$ that supersymmetric cycles with $H = 0$ are calibrated cycles. As the fivebrane has five spatial world-volume directions, there are many possibilities, which are summarised in table 4. We have included the amount of supersymmetry preserved including the number of supersymmetries counted with respect to the flat part of the world-volume $\mathbb{R}^{1,q}$ when $q \geq 1$.

Calibration	World-Volume	Supersymmetry
SLAG	$\mathbb{R}^{1,3} \times (\Sigma_2 \subset CY_2)$	8, $\mathcal{N}$=2 d=4
	$\mathbb{R}^{1,2} \times (\Sigma_3 \subset CY_3)$	4, $\mathcal{N}$=2 d=3
	$\mathbb{R}^{1,1} \times (\Sigma_4 \subset CY_4)$	2, (1,1) d=2
	$\mathbb{R} \times (\Sigma_5 \subset CY_5)$	1
	$\mathbb{R}^{1,1} \times (\Sigma_2 \subset CY_2) \times (\Sigma'_2 \subset CY'_2)$	4, (2,2) d=2
	$\mathbb{R} \times (\Sigma_2 \subset CY_2) \times (\Sigma_3 \subset CY_3)$	2
Kähler	$\mathbb{R}^{1,3} \times (\Sigma_2 \subset CY_3)$	4, $\mathcal{N}$=1 d=4
	$\mathbb{R}^{1,1} \times (\Sigma_4 \subset CY_3)$	4, (4,0) d=2
	$\mathbb{R}^{1,1} \times (\Sigma_4 \subset CY_4)$	2, (2,0) d=2
$\mathbb{C}$-Lag	$\mathbb{R}^{1,1} \times (\Sigma_4 \subset HK_2)$	3, (2,1) d=2
Associative	$\mathbb{R}^{1,2} \times (\Sigma_3 \subset G_2)$	2, $\mathcal{N}$=1 d=3
Co-associative	$\mathbb{R}^{1,1} \times (\Sigma_4 \subset G_2)$	2, (2,0) d=2
Cayley	$\mathbb{R}^{1,1} \times (\Sigma_4 \subset Spin(7))$	1, (1,0) d=2

TABLE 4. The different ways in which fivebranes can wrap calibrated cycles and the amount of supersymmetry preserved.

The six-dimensional field theory living on a single planar fivebrane has five scalar fields, describing the transverse fluctuations, the three-form H and fermions, and has a chiral $(2, 0)$ supersymmetry. The field theory when there are N coincident fivebranes is not yet well understood. The AdS/CFT conjecture states that it is dual to M-theory propagating on $AdS_7 \times S^4$. Recall that the field theory has an $SO(5)$ R-symmetry.

In order to construct supergravity solutions describing wrapped branes, it is very helpful to understand how supersymmetry is realised in the field theory living on a probe-brane world-volume. The details depend on which calibrated cycle is being wrapped and they are intimately connected to the structure of the normal bundle of the calibrated cycle. Let us concentrate on the case of fivebranes wrapping SLAG 3-cycles, as this will be the focus of the next section. The field theory on the probe fivebrane world-volume lives on $\mathbb{R}^{1,2} \times \Sigma_3$. In order for this field theory to be supersymmetric it is necessary that there be some notion of a constant spinor on Σ_3. It is not immediately clear what this notion is, since, in general, Σ_3 will not have a covariantly constant spinor. However, the field theory living on the fivebrane with world-volume $\mathbb{R}^{1,5} \subset \mathbb{R}^{1,10}$ has an internal $SO(5)$ R-symmetry, coming from the five flat transverse directions, under which the fermions transform. The covariant derivative of the spinors is schematically of the form

$$(\partial_\mu + \omega_\mu - A_\mu)\epsilon \ , \tag{4.23}$$

where ω is the spin connection and A is the $SO(5)$ gauge connection. Now consider the fivebrane theory on $\mathbb{R}^{1,2} \times \Sigma_3$. If we decompose $SO(5) \to SO(3) \times SO(2)$ and choose the $SO(3)$ gauge fields to be given by the $SO(3)$ spin connection on Σ_3, $A = \omega$, then clearly we can have constant spinors on Σ_3 that could parameterise the supersymmetry. This is exactly the way supersymmetry is realised for wrapped branes [**18**]. It is sometimes said that the field theory is "twisted" because of the similarities with the construction of topological field theories.

Geometrically, the identification of the $SO(3) \subset SO(5)$ gauge fields with the spin connection on Σ_3 corresponds to the structure of the normal bundle of a SLAG 3-cycle. The five directions transverse to the fivebrane wrapping the SLAG 3-cycle consist of three directions that are tangent to the CY_3 and two flat directions that are normal to the CY_3. This is responsible for breaking the $SO(5)$ symmetry of the flat fivebrane down to $SO(3) \times SO(2)$. We expect an $SO(2)$ R-symmetry to survive corresponding to the two flat directions, and thus the $SO(2)$ gauge fields are zero in the vacuum state. In section 4.1 we argued that for SLAG 3-cycles $N(\Sigma_3) \cong T(\Sigma_3)$ and this is responsible for the fact that there are non-zero $SO(3)$ gauge fields in the vacuum state and moreover, $A = \omega$.

We can determine which external $SO(5)$ gauge fields are excited for fivebranes wrapping different supersymmetric cycles from our previous discussion of the normal bundles of calibrated cycles. We shall mention this again in the next section in the context of constructing the corresponding supergravity solutions.

4.4. Generalised Calibrations. As somewhat of an aside, we comment that there are more general supersymmetric cycles than those we have discussed above.

Firstly, we only considered background geometries when the background fluxes (e.g. the four-form field strength G for D=11 supergravity) are all set to zero. By analysing supersymmetric brane configurations when the fluxes are non-zero, one is naturally lead to the notion of "generalised calibrations" [**87, 88**] (see also [**11**]). The key new feature is that the exterior derivative of the generalised calibration is no longer zero and is related to the flux. It is interesting to note that generalised calibrations play an important role in characterising the most general classes of supersymmetric supergravity solutions [**67, 74, 76, 75**]. They have also been discussed in [**86**].

A second generalisation is to determine the conditions for supersymmetric branes when non-trivial world-volume fluxes (F for D3-branes and H for the fivebranes) are switched on. This is related to the possibility of branes ending on branes and is discussed in [**14, 72, 65, 112**].

5. Supergravity solutions for fivebranes wrapping calibrated cycles

At this point, we have established that D=11 supergravity has supersymmetric membrane and fivebrane solutions. By considering the world-volume approximation to the dynamics of these branes we concluded that supersymmetric solutions of D=11 supergravity describing branes wrapping calibrated cycles in special holonomy manifolds should also exist. In this section we will explain the explicit construction of such solutions, in the near horizon limit, focusing on the richest case of fivebranes.

At first sight it is not at all clear how to construct these solutions. One might imagine that one should start with an explicit special holonomy metric, which are rather rare, and then "switch on the brane". In fact the procedure we adopt [**110**] is more indirect and subtle. A key point is that we aim to find the solutions in the near horizon limit, i.e. near to the brane wrapping the cycle, and this simplifies things in two important ways. Firstly, we expect that only the local geometry of the calibrated cycle in the special holonomy manifold, including the structure of its normal bundle, enters into the construction. Secondly, we will be able to employ a very useful technical procedure of first finding the solutions in D=7 $SO(5)$ gauged supergravity. This theory arises from the consistent truncation of the Kaluza-Klein reduction on a four-sphere of D=11 supergravity, as we shall describe. In particular any supersymmetric solution of the D=7 theory gives rise to a supersymmetric solution of D=11 supergravity. Although the converse is certainly not true, the D=7 gauged supergravity does include many interesting solutions corresponding to the near horizon limit of wrapped fivebrane geometries.

We first discuss Kaluza-Klein reduction starting with the simplest case of reduction on a circle. We then describe the reduction on a four-sphere leading to D=7 gauged supergravity. Following this we will describe the construction of fivebranes wrapping calibrated cycles. We focus on the case of fivebranes wrapping SLAG 3-cycles to illustrate some details and then summarise some aspects of the other cases.

5.1. Consistency of Kaluza-Klein reduction. The basic example of Kaluza-Klein dimensional reduction is to start with pure gravity in five spacetime dimensions and then reduce on a circle to get a theory of gravity in four spacetime dimensions coupled to a $U(1)$ gauge field and a scalar field. The procedure is to first expand the five-dimensional metric in harmonics on the circle. One obtains an infinite tower of modes whose four-dimensional mass is proportional to the inverse of the radius of the circle, as well as some massless modes consisting of the four-dimensional metric, gauge field and scalar field just mentioned. Finally one truncates the theory to the massless mode sector.

This truncation is said to be "consistent" in the sense that any solution of the four-dimensional theory is automatically a solution to the five-dimensional theory. The reason for this consistency is simply that the massless modes being kept are independent of the coordinate on the circle, while the massive modes, which have non-trivial dependence on the coordinate on the circle, are all set to zero. Note that this is an exact statement that does not rely on the radius of the circle being small, where one might argue that the massive modes are decoupling because they are all getting very heavy. Similarly, as we shall shortly illustrate in more detail, one can consistently truncate the reduction of theories with additional matter fields on a circle, and more generally on tori.

The D=7 gauged supergravity that we shall be interested in arises from the dimensional reduction of D=11 supergravity on a four-sphere. In general there are no consistent truncations of the dimensional reductions of gravity theories on spheres with dimension greater than one. The reason is that all of the harmonics on the sphere, including those associated with the lowest mass modes, typically depend on the coordinates of the sphere. Indeed, generically, if one reduces a theory of gravity on a sphere and attempts to truncate to the lowest mass modes, one will find that it is not consistent. That is, solutions of the truncated theory will not correspond to exact solutions of the higher-dimensional theory. Of course if the radius of the sphere were taken to be very small the truncated solutions could provide very good approximations to solutions of the higher-dimensional theory. However, in some special cases, including the reduction of D=11 supergravity on a four-sphere, it has been shown that there is in fact a consistent truncation. We will exploit this fact to construct exact solutions of D=11 supergravity by "uplifting" solutions that we first find in the D=7 gauged supergravity.

Before we present the Kaluza-Klein reduction formulae for D=7 gauged supergravity, which are rather involved, let us first present them in the much simpler setting of type IIA supergravity.

5.2. Reduction on S^1 to type IIA supergravity. Type IIA supergravity in ten dimensions can be obtained from the Kaluza-Klein reduction of D=11 supergravity on S^1 [**77, 30, 100**]. To see this, we construct an ansatz for the D=11 supergravity fields that just maintains the lowest massless modes. For the bosonic fields we let

$$\begin{aligned} ds^2 &= e^{-2\Phi/3} ds_{10}^2 + e^{4\Phi/3}(dy + C^{(1)})^2 \\ C &= C^{(3)} + B \wedge dy \ . \end{aligned} \tag{5.1}$$

Here the ten-dimensional line element ds_{10}^2, the scalar dilaton Φ, the "Ramond-Ramond" one-form $C^{(1)}$ and three-form $C^{(3)}$, and the the Neveu-Schwarz two-form B are all independent of y. The field strengths of the forms will be denoted $F^{(2)} = dC^{(1)}$, $F^{(4)} = dC^{(3)}$ and $H = dB$. If we substitute this ansatz into the D=11 equations of motion we find equations of motion for the ten-dimensional fields which are derivable from the action

$$S = \int d^{10}x\sqrt{-g}\left(e^{-2\Phi}[R + 4\partial\Phi^2 - \frac{1}{12}H^2] - \frac{1}{48}F_{(4)}^2 - \frac{1}{4}F_{(2)}^2\right) - \frac{1}{2}B \wedge F_4 \wedge F_4 \ . \tag{5.2}$$

This is precisely the bosonic part of the action of type IIA supergravity. After similarly including the fermions we find the full supersymmetric type IIA action, which preserves 32 supersymmetries (two D=10 Majorana-Weyl spinors of opposite chirality). Note that the isometries of S^1 give rise to the $U(1)$ gauge field with field strength $F_{(2)}$.

The key point to emphasise is that, by construction, any solution of type II supergravity automatically can be uplifted to give a solution of D=11 supergravity that admits a $U(1)$ isometry using the formulae in (5.1). Moreover, the D=11 supergravity solution will preserve at least the same amount of supersymmetry as the type IIA solution[4].

To illustrate with a simple example, consider the following supersymmetric solution of type IIA supergravity:

$$\begin{aligned} ds^2 &= H^{-1}(d\xi^i d\xi^j \eta_{ij}) + (dx^I dx^I) \\ B &= H^{-1} d\xi^0 \wedge d\xi^1 \\ e^{2\Phi} &= H^{-1} \ , \end{aligned} \tag{5.3}$$

[4]Note that there are D=11 solutions with $U(1)$ isometries that have supersymmetries that will not survive the dimensional reduction because the Killing spinors have a non-trivial dependence on the coordinate on the circle.

with $i,j=0,1$ and $I,J=1,\ldots,8$. This is a solution to IIA supergravity providing that H is harmonic in the transverse space. Choosing the simple single centre solution $H=1+\alpha_2 N/r^6$ we find that this solution carries N units of quantised electric H-flux. In fact this solution describes the fields around N coincident fundamental IIA strings of infinite extent. It preserves one-half of the supersymmetry. If we now uplift this solution to get a solution of D=11 supergravity using (5.1) we obtain the planar membrane solution (2.42).

5.3. Reduction on S^4 to D=7 $SO(5)$ gauged supergravity. The dimensional reduction of D=11 supergravity on a four-sphere can be consistently truncated to give D=7 $SO(5)$ gauged supergravity [**116, 117**]. The origin of the $SO(5)$ gauge symmetry is the $SO(5)$ isometries of the four-sphere.

The explicit formulae for the D=11 bosonic fields is given by

$$\begin{aligned} ds^2 &= \Delta^{-2/5} ds_7^2 + \frac{\Delta^{4/5}}{m^2} DY^A (T^{-1})^{AB} DY^B \\ 8G &= \epsilon_{A_1\ldots A_5}\Bigg[-\frac{1}{3m^3} DY^{A_1} DY^{A_2} DY^{A_3} DY^{A_4} \frac{(T\cdot Y)^{A_5}}{Y\cdot T\cdot Y} \\ &\quad + \frac{4}{3m^3} DY^{A_1} DY^{A_2} DY^{A_3} D\left(\frac{(T\cdot Y)^{A_4}}{Y\cdot T\cdot Y}\right) Y^{A_5} \\ &\quad + \frac{2}{m^2} F^{A_1A_2} DY^{A_3} DY^{A_4} \frac{(T\cdot Y)^{A_5}}{Y\cdot T\cdot Y} + \frac{1}{m} F^{A_1A_2} F^{A_3A_4} Y^{A_5}\Bigg] \\ &\quad + d(S_B Y^B)\ , \end{aligned} \tag{5.4}$$

and the wedge product of forms is to be understood in the expression for G. Here Y^A, $A,B=1,\ldots,5$, are constrained coordinates parametrising a four-sphere, satisfying $Y^AY^A=1$. In this section x^μ, $\mu,\nu=0,1,\ldots,6$ are D=7 coordinates and ds_7^2 is the D=7 line element associated with the D=7 metric $g_7(x)$. The $SO(5)$ isometries of the round four-sphere lead to the introduction of $SO(5)$ gauge fields $B^A{}_B(x)$, with field strength $F^A{}_B(x)$, that appear in the covariant derivative DY^A:

$$DY^A = dY^A + 2mB^A{}_BY^B\ . \tag{5.5}$$

The matrix T is defined by

$$T^{AB}(x) = (\Pi^{-1})_i{}^A(x)(\Pi^{-1})_j{}^B(x)\delta^{ij}\ , \tag{5.6}$$

where $i,j=1,\ldots,5$ and $\Pi_A{}^i(x)$ are 14 scalar fields that parameterise the coset $SL(5,\mathbb{R})/SO(5)$. The warp factor Δ is defined via

$$\Delta^{-6/5} = Y^A T^{AB}(x) Y^B \tag{5.7}$$

and $S_A(x)$ are five three-forms.

If we substitute this into the D=11 supergravity equations of motion we get equations of motion for $g_7(x), B(x), \Pi(x)$ and $S(x)$. The resulting equations of motion can be derived from the D=7 action:

$$\begin{aligned} S = \int d^7x\sqrt{-g}\Bigg[& R + \frac{1}{2}m^2\left(T^2 - 2T_{ij}T^{ij}\right) - P_{\mu ij}P^{\mu ij} - \frac{1}{2}\left(\Pi_A{}^i\Pi_B{}^jF^{AB}_{\mu\nu}\right)^2 \\ & - m^2\left(\Pi^{-1}{}_i{}^A S_{\mu\nu\rho,A}\right)^2\Bigg] - 6mS_A\wedge F_A \\ & + \sqrt{3}\epsilon_{ABCDE}\delta^{AG}S_G\wedge F^{BC}\wedge F^{DE} + \frac{1}{8m}\left(2\Omega_5[B]-\Omega_3[B]\right)\ , \end{aligned} \tag{5.8}$$

where we have dropped the 7 subscript on g_7 here and below, for clarity. This is precisely the bosonic part of the action of D=7 $SO(5)$ gauged supergravity [**123**]. In particular, the kinetic energy terms for the scalar fields are determined by $P_{\mu ij}$ which is defined to be the part symmetric in i and j of

$$(\Pi^{-1})_i{}^A \left(\delta_A{}^B \partial_\mu + 2mB_{\mu\, A}{}^B\right) \Pi_B{}^k \delta_{kj} \ . \tag{5.9}$$

The potential terms for the scalar fields are defined in terms of

$$T_{ij} = (\Pi^{-1})_i{}^A (\Pi^{-1})_j{}^A, \qquad T = T_{ii} \ . \tag{5.10}$$

Finally, $\Omega_3[B]$ and $\Omega_5[B]$ are Chern-Simons forms for the gauge fields, whose explicit form will not be needed.

It is also possible to explicitly construct an ansatz for the D=11 fermions to recover the fermions and supersymmetry of the D=7 gauged supergravity. In particular the construction implies that any bosonic (supersymmetric) solution of D=7 $SO(5)$ gauged supergravity will uplift via (5.4) to a bosonic (supersymmetric) solution of D=11 supergravity.

In order to find bosonic supersymmetric solutions of the D=7 gauged supergravity, we need the supersymmetry variations of the fermions. The fermions consist of gravitini ψ_μ and dilatini λ and, in a bosonic background, their supersymmetry variations are given by

$$\begin{aligned}
\delta\psi_\mu &= \nabla_\mu \epsilon - \frac{1}{40}(\gamma_\mu{}^{\nu\rho} - 8\delta_\mu^\nu \gamma^\rho)\Gamma_{ij}\epsilon\, \Pi_A{}^i \Pi_B{}^j F_{\nu\rho}^{AB} \\
&\quad + \frac{1}{20} m T \gamma_\mu \epsilon + \frac{m}{10\sqrt{3}}(\gamma_\mu{}^{\nu\rho\sigma} - \frac{9}{2}\delta_\mu{}^\nu \gamma^{\rho\sigma})\Gamma^i \epsilon\, \Pi^{-1}{}_i{}^A S_{\nu\rho\sigma,A} \\
\delta\lambda_i &= \frac{1}{2}\gamma^\mu \Gamma^j \epsilon\, P_{\mu ij} + \frac{1}{16}\gamma^{\mu\nu}(\Gamma_{kl}\Gamma_i - \frac{1}{5}\Gamma_i \Gamma_{kl})\epsilon\, \Pi_A{}^k \Pi_B{}^l F_{\mu\nu}^{AB} \\
&\quad + \frac{1}{2} m (T_{ij} - \frac{1}{5} T \delta_{ij})\Gamma^j \epsilon + \frac{m}{20\sqrt{3}}\gamma^{\mu\nu\rho}(\Gamma_i{}^j - 4\delta_i^j)\epsilon\, \Pi^{-1}{}_j{}^A S_{\mu\nu\rho,A} \ .
\end{aligned} \tag{5.11}$$

Here γ^μ are the D=7 gamma matrices of $Cliff(6,1)$, while Γ^i are those of $Cliff(5)$, and these act on ϵ which is a spinor with respect to both $Spin(6,1)$ and $Spin(5)$. Note that γ^μ and Γ^i commute. The covariant derivative appearing in the supersymmetry variation for the gravitini is given by

$$\nabla_\mu \epsilon = \left(\partial_\mu + \frac{1}{4}\omega_\mu^{ab}\gamma_{ab} + \frac{1}{4} Q_{\mu ij}\Gamma^{ij}\right)\epsilon \ , \tag{5.12}$$

where $Q_{\mu ij}$ is the part of (5.9) anti-symmetric in i and j. To obtain supersymmetric configurations we set $\delta\psi = \delta\lambda^i = 0$ and let ϵ be a commuting spinor. Note that setting the variations of the dilatini to zero leads to algebraic constraints on ϵ.

Although still very complicated, D=7 gauged supergravity is a simpler theory than D=11 supergravity as many degrees of freedom have been truncated. It is clear from (5.4) that the "breathing mode", corresponding to uniformly scaling the round four-sphere, is one of the modes that has been truncated. This means that there are no solutions of the D=7 gauged supergravity which give rise to the full fivebrane solution of D=11 supergravity (2.33) with (2.36). However, the near horizon limit of the fivebrane solution, $AdS_7 \times S^4$, is easily found. Indeed it arises as the simplest "vacuum" solution of the theory where the gauge fields, the three-forms and the scalars are all set to zero: $B = S = 0$, $\Pi = \delta$. In this case the equations of motion reduce to solving

$$R_{\mu\nu} = -\frac{3}{2}m^2 g_{\mu\nu} \ , \tag{5.13}$$

and the unique solution preserving all supersymmetry is AdS_7, which we can write in Poincaré coordinates as

$$ds^2 = \frac{4}{m^2}\left[\frac{d\xi^i d\xi^j \eta_{ij} + d\rho^2}{\rho^2}\right] \ . \tag{5.14}$$

If we uplift this solution to D=11 using (5.4) we recover the $AdS_7 \times S^4$ solution (2.40), with the radius of the AdS_7 given by $2/m$. Note that, in these coordinates, the D=7 metric clearly displays the flat planar world-volume of the fivebrane.

5.4. Fivebranes wrapping SLAG 3-cycles. Supergravity solutions describing fivebranes wrapping different calibrated cycles have been constructed in D=7 gauged supergravity and then uplifted to D=11 in [**110, 3, 71, 68**]. We will illustrate in some detail the construction of fivebranes wrapping SLAG 3-cycles [**71**] and then comment more briefly on the other cases.

Consider the D=11 supersymmetric geometry $\mathbb{R}^{1,2} \times CY_3 \times \mathbb{R}^2$ and $G = 0$, with a probe fivebrane wrapping a SLAG 3-cycle inside the Calabi-Yau three-fold. i.e. the world-volume of the fivebrane is $\mathbb{R}^{1,2} \times \Sigma_3$ with $\Sigma_3 \subset CY_3$. The five directions transverse to the fivebrane world-volume consist of three that are tangent to the CY_3 and two flat directions that are normal to the CY_3. If we wrap many fivebranes, the back-reaction on the geometry will be significant and we aim to find the corresponding supergravity solutions with $G \neq 0$.

To construct these solutions, we seek a good ansatz for D=7 gauged supergravity. By analogy with the flat planar fivebrane solution, we think of the D=7 coordinates as being those of the world-volume of the fivebrane plus an additional radial direction which, when the solution is uplifted to D=11, should correspond to a kind of radial distance away from the fivebrane in the five transverse directions. Thus an obvious ansatz for the D=7 metric is

$$ds^2 = e^{2f}[ds^2(\mathbb{R}^{1,2}) + dr^2] + e^{2g} d\bar{s}^2(\Sigma_3) \ , \tag{5.15}$$

where f, g are functions of r only and $d\bar{s}^2(\Sigma_3)$ is some metric on the 3-cycle. Now before the back-reaction is taken into account, three of the directions transverse to the fivebrane were tangent to CY_3 and two were flat. We thus decompose the $SO(5)$-symmetry of the D=7 theory into $SO(3) \times SO(2)$ and only switch on $SO(3)$ gauge fields. In order to preserve supersymmetry, as we shall elaborate on shortly, the $SO(3)$ gauge fields are chosen to be proportional to the spin connection of Σ_3:

$$2mB^a{}_b = \bar{\omega}^a{}_b \ , \tag{5.16}$$

for $a, b = 1, 2, 3$. This is a key part of the ansatz and it precisely corresponds to the fact that the normal bundle of SLAG-cycles is isomorphic to the tangent bundle of the cycle, as we discussed in section 4.1 and also at the end of section 4.3. An ansatz for the scalar fields respecting $SO(3) \times SO(2)$ symmetry is given by

$$\Pi_A{}^i = diag(e^{2\lambda}, e^{2\lambda}, e^{2\lambda}, e^{-3\lambda}, e^{-3\lambda}) \ , \tag{5.17}$$

where λ is a third function of r. It is consistent with the equations of motion to set the three-forms S to zero, which we do to complete the ansatz.

For the configuration to preserve supersymmetry, we require that there exist spinors ϵ such that the supersymmetry variations (5.11) vanish. Given the above ansatz, the composite gauge fields Q appearing in (5.12) are given by the $SO(3)$ gauge fields $Q^{ab} = 2mB^{ab}$. As a consequence, in demanding the vanishing of the variation of the gravitini in the directions along Σ_3, we find that

$$(\partial + \frac{1}{4}\bar{\omega}^{ab}\gamma_{ab} + \frac{m}{2}B^{ab}\Gamma_{ab})\epsilon = 0 \ , \tag{5.18}$$

where $\bar{\omega}$ is the $SO(3)$ spin connection on Σ_3. We now begin to see the significance of the assumption (5.16). In particular, in the obvious orthonormal frame, (5.18) can be satisfied by spinors independent of the coordinates along Σ_3, if we impose the following projections on ϵ:

$$\gamma^{ab}\epsilon = -\Gamma^{ab}\epsilon \ . \tag{5.19}$$

Note that this precisely parallels our discussion on the preservation of supersymmetry in the context of the world-volume of the probe fivebrane at the end of section 4.3. To ensure that the variation of all components of the gravitini and dilatini vanish we also need to impose

$$\gamma^r\epsilon = \epsilon \tag{5.20}$$

and we find that the the only dependence of the Killing spinors on the coordinates is radial: $\epsilon = e^{f/2}\epsilon_0$ where ϵ_0 is a constant spinor. Note that only two of the conditions (5.19) are independent and they break 1/4 of the supersymmetry. When combined with (5.20) we see that 1/8 of the supersymmetry is preserved, in agreement with table 4. Actually, if one follows through these preserved supersymmetries to D=11 one finds that they are precisely those that one expects when a fivebrane wraps a SLAG 3-cycle: there are two projections corresponding to the CY_3 and another for the fivebrane.

A more detailed analysis of the conditions for supersymmetry implies that the metric on Σ_3 must in fact be Einstein. Given the factor e^{2g}, we can always normalise so that the Ricci tensor and the metric on Σ_3 are related by

$$\bar{R}ic(\Sigma_3) = l\bar{g}(\Sigma_3) \ , \tag{5.21}$$

with $l = 0, \pm 1$. In three dimensions the Riemann curvature tensor is determined by the Ricci tensor. When $l = 0$ (5.21) implies that Σ_3 is flat: the resulting D=11 solution, after uplifting, corresponds to a fivebrane with a flat planar world-volume and is thus not of primary interest. Indeed the solution turns out to be just a special case of the fivebrane solution presented in (2.33) with a special harmonic function with $SO(3) \times SO(2)$ symmetry.

The cases of most interest are thus when $l = \pm 1$. When $l = 1$ the Einstein condition (5.21) implies that Σ_3 is the three-sphere, S^3, or a quotient by a discrete subgroup of isometries of the isometry group $SO(4)$, while for $l = -1$ it is hyperbolic three-space, H^3, or a quotient by a discrete subgroup of $SO(3,1)$. Note in particular that when $l = -1$ it is possible that the resulting geometry is compact. We mentioned above that the Killing spinors are independent of the coordinates on the cycle, and hence the quotients S^3/Γ and H^3/Γ are also supersymmetric.

Finally, in addition, supersymmetry implies the following first order BPS equations on the three radial functions:

$$\begin{aligned} e^{-f} f' &= -\frac{m}{10}\left[3e^{-4\lambda}+2e^{6\lambda}\right]+\frac{3l}{20m}e^{4\lambda-2g} \\ e^{-f} g' &= -\frac{m}{10}\left[3e^{-4\lambda}+2e^{6\lambda}\right]-\frac{7l}{20m}e^{4\lambda-2g} \\ e^{-f} \lambda' &= \frac{m}{5}\left[e^{6\lambda}-e^{-4\lambda}\right]+\frac{l}{10m}e^{4\lambda-2g}\ . \end{aligned} \tag{5.22}$$

If these equations are satisfied then all of the D=7 equations of motion are satisfied and we have found a supersymmetric solution.

Using (5.4) we can now uplift any solution of these BPS equations to obtain supersymmetric solutions of D=11 supergravity. The metric is given by

$$ds^2_{11}=\Delta^{-2/5}ds^2_7+\frac{1}{m^2}\Delta^{4/5}\left[e^{4\lambda}DY^aDY^a+e^{-6\lambda}dY^\alpha dY^\alpha\right]\ , \tag{5.23}$$

where

$$\begin{aligned} DY^a &= dY^a+\bar{\omega}^a{}_bY^b \\ \Delta^{-6/5} &= e^{-4\lambda}Y^aY^a+e^{6\lambda}Y^\alpha Y^\alpha\ , \end{aligned} \tag{5.24}$$

with $a,b=1,2,3$, $\alpha=4,5$ and (Y^a,Y^α) are constrained coordinates on S^4 satisfying $Y^aY^a+Y^\alpha Y^\alpha=1$. The expression for the four-form is given by substituting into (5.4). Clearly, the four-sphere is no longer round and it is non-trivially fibred over the three-dimensional Einstein space Σ_3.

It is illuminating[5] to change coordinates from the constrained coordinates (r,Y^A) to unconstrained coordinates (ρ^a,ρ^α) via

$$\begin{aligned} \rho^a &= -\frac{1}{m}e^{f+g+2\lambda}Y^a \\ \rho^\alpha &= -\frac{1}{m}e^{2f-3\lambda}Y^\alpha\ . \end{aligned} \tag{5.25}$$

The metric then takes the form:

$$\begin{aligned} ds^2 &= (\Delta^{-2/5}e^{2f})\left[ds^2(\mathbb{R}^{1,2})+e^{2g-2f}d\bar{s}^2(\Sigma_3)\right] \\ &\quad+(\Delta^{4/5}e^{-4f})\left[e^{2f-2g}(d\rho^a+\bar{\omega}^a{}_b\rho^b)^2+d\rho^\alpha d\rho^\alpha\right]\ . \end{aligned} \tag{5.26}$$

In these coordinates the warp factors have a similar form to the simple planar fivebrane solution (2.33); in particular this form confirms the interpretation of the solutions as describing fivebranes with world-volumes given by $\mathbb{R}^{1,2}\times\Sigma_3$. In addition the metric clearly displays the $SO(2)$ symmetry corresponding to rotations in the ρ^4,ρ^5 plane.

Of course to find explicit solutions we need to solve the BPS equations. When $l=-1$ it is easy to check that there is an exact solution given by

$$\begin{aligned} e^{10\lambda} &= 2 \\ e^{2g} &= \frac{e^{8\lambda}}{2m^2} \\ e^{2f} &= \frac{e^{8\lambda}}{m^2r^2}\ . \end{aligned} \tag{5.27}$$

[5] These types of coordinates were first noticed in the context of wrapped membranes in [**70**].

The D=7 metric is then the direct product $AdS_4 \times (H^3/\Gamma)$ (this D=7 solution was first found in [**124**]). The uplifted D=11 solution is a warped product of $AdS_4 \times (H^3/\Gamma)$ with a four-sphere which is non-trivially fibred over (H^3/Γ). The presence of the AdS_4 factor in the D=11 solution indicates that M-theory on this background is dual to a superconformal field theory in three spacetime dimensions. We will return to this point after analysing the general BPS solution. Note that when $l = 1$ there is no $AdS_4 \times (S^3/\Gamma)$ solution.

It seems plausible that the BPS equations could be solved exactly. However, much of the physical content can be deduced from a simple numerical investigation. If we introduce the new variables

$$\begin{aligned} a^2 &= e^{2g}e^{-8\lambda} \\ e^h &= e^{f-4\lambda} , \end{aligned} \tag{5.28}$$

then the BPS equations are given by

$$\begin{aligned} e^{-h}h' &= -\frac{m}{2}\left[2e^{10\lambda}-1\right] - \frac{l}{4ma^2} \\ e^{-h}\frac{a'}{a} &= -\frac{m}{2}\left[2e^{10\lambda}-1\right] - \frac{3l}{4ma^2} \\ e^{-h}\lambda' &= \frac{m}{5}\left[e^{10\lambda}-1\right] + \frac{l}{8ma^2} . \end{aligned} \tag{5.29}$$

We next define $x = a^2$ and $F = xe^{10\lambda}$ to obtain the ODE

$$\frac{dF}{dx} = \frac{F[m^2x - 5\alpha + 2\beta]}{x[m^2(2F-x)+2\beta]} . \tag{5.30}$$

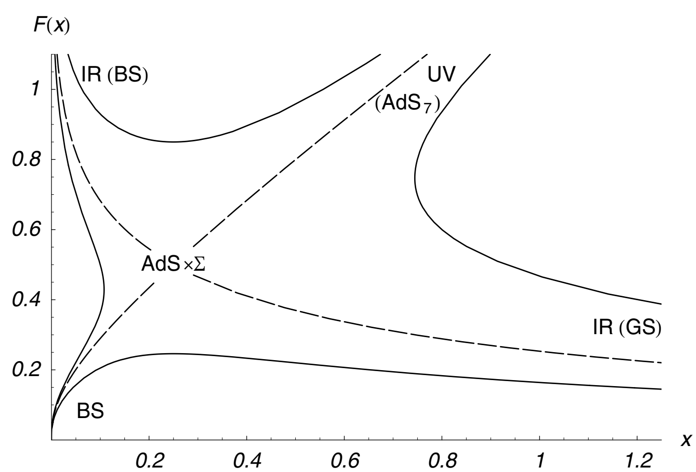


FIGURE 1. Behaviour of the orbits for five-branes wrapping SLAG 3-cycles with $l = -1$. Note the flow from the AdS_7-type region when F, x are large to the IR fixed point and the flows to the good and bad singularities in the IR, IR(GS) and IR(BS), respectively.

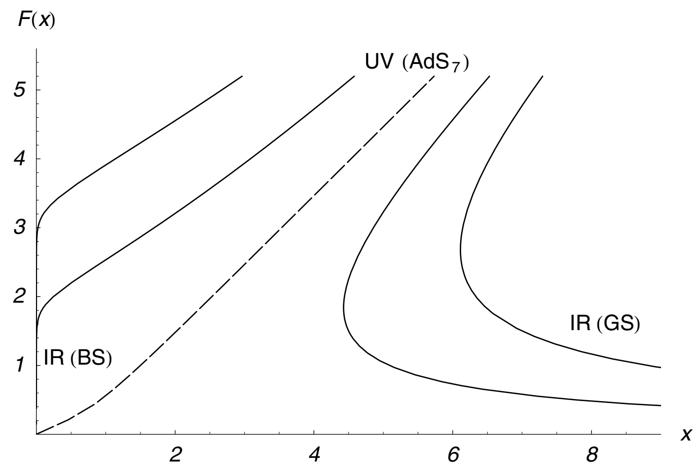


FIGURE 2. Behaviour of the orbits for fivebranes wrapping SLAG 3-cycles with $l = 1$.

The typical behaviour of $F(x)$ is illustrated in figure 1 for $l = -1$ and figure 2 for $l = 1$. The region where both x and F are large is interesting. There we have $F \approx x - l/m^2$ and using a as a radial variable we obtain the asymptotic behaviour of the metric:

$$ds^2 \approx \frac{4}{m^2 a^2} da^2 + a^2[ds^2(\mathbb{R}^{1,2}) + d\bar{s}^2(\Sigma_3)] \ .$$

This looks very similar to AdS_7 in Poincaré coordinates except that the sections with constant a are not $\mathbb{R}^{1,5}$ but $\mathbb{R}^{1,2} \times \Sigma_3$.

This clearly corresponds to the near horizon limit of the fivebrane wrapped on the SLAG 3-cycle. By the general discussion on the AdS/CFT correspondence earlier, this should be dual to the six-dimensional quantum field theory living on the wrapped fivebrane worldvolume $\mathbb{R}^{1,2} \times \Sigma_3$, after decoupling gravity. More precisely, the asymptotic behaviour of the solution (5.31), when lifted to D=11, is dual to the UV behaviour of the quantum field theory. Following the flow of the solution as in figures 1 and 2 correspond to flowing to the IR of the field theory. In the present context, the IR corresponds to length scales large compared to the size of the cycle Σ_3 on which the fivebrane is wrapped. In other words, going to the IR corresponds to taking Σ_3 to be very small (assuming it is compact) and the six-dimensional quantum field theory on $\mathbb{R}^{1,2} \times \Sigma_3$ behaves more and more like a three-dimensional quantum field theory on $\mathbb{R}^{1,2}$.

Perhaps the most interesting solutions occur for $l = -1$. There is a solution indicated by one of the dashed lines in figure 1 that flows from the UV AdS_7 type region to the $AdS_4 \times (H^3/\Gamma)$ fixed point that was given in (5.27). This is a supergravity solution that describes a kind of renormalisation group flow from a theory on $\mathbb{R}^{1,2} \times (H^3/\Gamma)$ to a superconformal field theory on $\mathbb{R}^{1,2}$. In particular, we see that the natural interpretation of the $AdS_4 \times (H^3/\Gamma)$ solution is that, when it is lifted to D=11, it is dual to a superconformal field theory on $\mathbb{R}^{1,2}$ that arises as the

IR limit of the fivebrane field theory on living on $\mathbb{R}^{1,2} \times (H^3/\Gamma)$. The interpretation for non-compact H^3/Γ is less clear.

The absence of an $AdS_4 \times (S^3/\Gamma)$ solution for $l = 1$ possibly indicates that the quantum field theories arising on fivebranes wrapping SLAG 3-cycles with positive curvature are not superconformal in the infrared. Alternatively, it could be that there are more elaborate solutions lying outside of our ansatz that have AdS_4 factors.

All other flows in figures 1 and 2 starting from the AdS_7 type region flow to singular solutions. Being singular does not exclude the possibility that they might be interesting physically. Indeed, a criteria for time-like singularities in static geometries to be "good singularities", i.e. dual to some quantum field theory behaviour, was presented in [**110**]. In particular, a good singularity is defined to be one in which the norm of the time-like Killing vector with respect to the D=11 supergravity metric does not increase as one goes to the singularity (one can also consider the weaker criterion that the norm is just bounded from above). It is not difficult to determine whether the singularities that arise in the different asymptotic limits are good or bad by this criterion and this has been presented in figures 1 and 2. It is likely that the good singularities describe some kind of Higgs branches of the quantum field theory corresponding to the possibility of moving the coincident wrapped fivebranes apart.

In summary, using D=7 gauged supergravity, we have been able to construct D=11 supergravity solutions that describe fivebranes wrapping SLAG 3-cycles. The cycle is Einstein and is either S^3/Γ or H^3/Γ where Γ is a discrete group of isometries. Probably the most important solutions that have been found are the $AdS_4 \times (H^3/\Gamma)$ solutions which are dual to new superconformal field theories, after being uplifted to D=11. More general flow solutions were also constructed numerically.

5.5. Fivebranes wrapping other cycles. The construction of supersymmetric solutions corresponding to fivebranes wrapping other supersymmetric cycles runs along similar lines. The ansatz for the D=7 metric is given by

$$ds^2 = e^{2f}[ds^2(\mathbb{R}^{1,5-d}) + dr^2] + e^{2g} d\bar{s}^2(\Sigma_d) \ , \tag{5.31}$$

where $d\bar{s}^2_d$ is the metric on the supersymmetric d-cycle, Σ_d, and the functions f and g depend on the radial coordinate r only.

The $SO(5)$-gauge fields are specified by the spin connection of the metric on Σ_d in a way determined by the structure of the normal bundle of the calibrated cycle being wrapped. In general, we decompose the $SO(5)$ symmetry into $SO(p) \times SO(q)$ with $p + q = 5$, and only excite the gauge fields in the $SO(p)$ subgroup. We will denote these directions by $a, b = 1, \ldots, p$. If we consider a probe fivebrane wrapping the cycle inside a manifold of special holonomy M, this decomposition corresponds to dividing the directions transverse to the brane into p directions within M and q directions perpendicular to M. The precise ansatz for the $SO(p)$ gauge fields is determined by some part of the spin connection on the cycle and will be discussed shortly.

In keeping with this decomposition, the solutions that we consider have a single scalar field excited. More precisely we take

$${\Pi_A}^i = diag(e^{q\lambda}, \ldots, e^{q\lambda}, e^{-p\lambda}, \ldots, e^{-p\lambda}) \ , \tag{5.32}$$

where we have p followed by q entries. Once again this implies that the composite gauge-field Q is then determined by the gauge fields via $Q^{ab} = 2mB^{ab}$.

It turns out that for the SLAG 5-cycle and most of the 4-cycle cases it is necessary to have non-vanishing three-forms S_A. The S-equation of motion is

$$m^2 \delta_{AC} \Pi^{-1}{}_i{}^C \Pi^{-1}{}_i{}^B S_B = -m * F_A + \frac{1}{4\sqrt{3}} \epsilon_{ABCDE} * (F^{BC} \wedge F^{DE}) \ . \tag{5.33}$$

The solutions have vanishing four-form field strength F_A and hence S_A are determined by the gauge fields.

Demanding that the configuration preserves supersymmetry, as for the SLAG 3-cycle case, we find that along the cycle directions:

$$(\partial + \frac{1}{4}\bar{\omega}^{bc}\gamma_{bc} + \frac{m}{2} B^{ab}\Gamma_{ab})\epsilon = 0 \ , \tag{5.34}$$

where $\bar{\omega}^{bc}$ is the spin connection one-form of the cycle. For each case the specific gauge fields, determined by the type of cycle being wrapped, go hand in hand with a set of projections which allow (5.34) to be satisfied for spinors independent of the coordinates along Σ_d. We can easily guess the appropriate ansatz for the gauge fields from our discussion of the structure of the normal bundles of calibrated cycles in section 4.1, and they are summarised below. The corresponding projections are then easily determined and are given explicitly in [**110, 3, 71, 68**]. In the uplifted D=11 solutions, these projections translate into a set of projections corresponding to those of the special holonomy manifold and an additional projection corresponding to the wrapped fivebrane.

By analysing the conditions for supersymmetry in more detail, one finds that the metric on the cycle is necessarily Einstein, and we again normalise so that

$$\bar{R}_{ab} = l\bar{g}_{ab} \ , \tag{5.35}$$

with $l = 0, \pm 1$. When $d = 2, 3$ the Einstein condition implies that the cycles have constant curvature and hence are either spheres for $l = 1$ or hyperbolic spaces for $l = -1$, or quotients of these spaces by a discrete group of isometries. When $d > 3$ the Einstein condition implies that the Riemann tensor can be written

$$\bar{R}_{abcd} = \bar{W}_{abcd} + \frac{2l}{d-1}\bar{g}_{a[c}\bar{g}_{d]b} \ , \tag{5.36}$$

where $\bar{W}$ is the Weyl tensor, and there are more possibilities. By analysing the D=7 Einstein equations one finds that for $d = 4, 5$ the part of the spin connection that is identified with the gauge fields must have constant curvature.

We now summarise the ansatz for the $SO(5)$ gauge fields for each case and discuss the types of cycle that arise for $d = 4, 5$. We can always take a quotient of the cycles listed by a discrete group of isometries.

SLAG n-cycles: Consider a probe fivebrane wrapping a SLAG n-cycle in a CY_n. The five directions transverse to the fivebrane consist of n directions tangent to the CY_n and $5-n$ normal to it. Thus, for the supergravity solution we decompose $SO(5) \to SO(n) \times SO(5-n)$, and let the only non-vanishing gauge fields lie in the $SO(n)$ factor. Up to a factor of $2m$ these $SO(n)$ gauge fields are identified with the $SO(n)$ spin connection on Σ_n, as in (5.16). This identification corresponds to the fact that $N(\Sigma_n) \cong T(\Sigma_n)$ for SLAG n-cycles. Since all of the spin connection is identified with the gauge fields, the metric on Σ_n must have constant curvature ($\bar{W} = 0$ for $d = 4, 5$) and hence Σ_n is S^n for $l = 1$ and H^n for $l = -1$.

Kähler 2-cycles: Kähler 2-cycles in CY_2 are SLAG 2-cycles and have just been discussed. For probe fivebranes wrapping Kähler 2-cycles in CY_3 the five directions transverse to the fivebrane consist of four directions tangent to the CY_3 and one flat direction normal to the CY_3. The normal bundle of the Kähler 2-cycle has structure group $U(2) \cong U(1) \times SU(2)$, and now recall (4.5). Thus, for the supergravity solution in D=7, we decompose $SO(5) \to SO(4) \to U(2) \cong U(1) \times SU(2)$ and identify the $U(1)$ spin connection of the cycle with the gauge fields in the $U(1)$ factor. We also set the $SU(2)$ gauge fields to zero, which corresponds to considering fivebranes wrapping Kähler 2-cycles with non-generic normal bundle. An example of such a cycle is the two-sphere in the resolved conifold. It would be interesting to find more general solutions with non-vanishing $SU(2)$ gauge fields, preserving the same amount of supersymmetry.

Kähler 4-cycles: We assume that Σ_4 in the D=7 supergravity solution has a Kähler metric with a $U(2) \cong U(1) \times SU(2)$ spin connection. When probe fivebranes wrap Kähler 4-cycles in CY_3 the five transverse directions consist of two directions tangent to the CY_3 and three flat directions normal to the CY_3. Thus we decompose $SO(5) \to SO(2) \times SO(3)$, set the $SO(3)$ gauge fields to zero and identify the $SO(2) \cong U(1)$ gauge fields with the $U(1)$ part of the $U(1) \times SU(2)$ spin connection on the 4-cycle. When probe fivebranes wrap Kähler 4-cycles in CY_4 the five transverse directions consist of four directions tangent to the CY_4 and one flat direction normal to the CY_4. Thus we now decompose $SO(5) \to SO(4) \to U(2) \cong U(1) \times SU(2)$ and we set the $SU(2)$ gauge fields to zero, which again corresponds to considering non-generic normal bundles. We identify the $U(1)$ gauge fields with the $U(1)$ part of the spin connection as dictated by (4.5). In both cases, the identification of the gauge fields with part of the spin connection doesn't place any further constraints on Σ_4 other than it is Kähler-Einstein. An example when $l = 1$ is $\mathbb{C}P^2$.

$\mathbb{C}$-Lag 4-cycles: We again assume that Σ_4 in the D=7 solution has a Kähler metric with a $U(2) \cong U(1) \times SU(2)$ spin connection. We again decompose $SO(5) \to SO(4) \to U(2) \cong U(1) \times SU(2)$ but now we do not set the $SU(2)$ gauge fields to zero. Indeed, since the cycle is both SLAG and Kähler, with respect to different complex structures, we must identify all of the $U(2)$ gauge fields with the $U(2)$ spin connection. Einstein's equations then imply that Σ_4 must have constant holomorphic sectional curvature. This means that for $l = 1$ it is $\mathbb{C}P^2$ while for $l = -1$ it is the open disc in $\mathbb{C}^2$ with the Bergman metric. Note that the solutions corresponding to fivebranes wrapping $\mathbb{C}$-Lag $\mathbb{C}P^2$ are different from the solutions corresponding to fivebranes wrapping Kähler $\mathbb{C}P^2$, since they have more gauge fields excited and preserve different amounts of supersymmetry.

Associative 3-cycles: When probe fivebranes wrap associative 3-cycles in G_2 manifolds, there are four transverse directions that are tangent to the G_2 manifold and one flat direction normal to the G_2 manifold. We thus decompose $SO(5) \to SO(4) \cong SU(2)^+ \times SU(2)^-$, where the superscripts indicate the self-dual and anti-self-dual parts. Recall that the normal bundle of associative 3-cycles is given by $S \otimes V$ where S was the $SU(2)$ spin bundle on Σ_3 and V is a rank $SU(2)$ bundle. In the non-generic case when V is trivial, for example for the G_2 manifold in [**28**], then the identification of the gauge fields is clear: we should identify the $SO(3) \cong SU(2)$ spin connection on Σ_3 with $SU(2)^+$ gauge fields and set the $SU(2)^-$ gauge fields to zero.

Co-associative 4-cycles: When probe fivebranes wrap co-associative 4-cycles in G_2 manifolds, there are three transverse directions that are tangent to the G_2 manifold and two flat directions normal to the G_2 manifold. We thus decompose $SO(5) \to SO(3) \times SO(2)$ and set the $SO(2)$ gauge fields to zero. Recall that the normal bundle of co-associative 4-cycles is isomorphic to the bundle of anti-self-dual two-forms on the 4-cycle. This indicates that we should identify the $SO(3) \cong SU(2)$ gauge fields with the anti-self-dual part, $SU(2)^-$, of the $SO(4) \cong SU(2)^+ \times SU(2)^-$ spin connection on Σ_4. For the co-associative 4-cycles, Einstein's equations imply that the anti-self-dual part of the spin connection has constant curvature, or in other words, the Weyl tensor is self-dual $\bar{W}^- = 0$. These manifolds are sometimes called conformally half-flat. If $l = 1$ the only compact examples are $\mathbb{C}P^2$ and S^4.

Cayley 4-cycles: When probe fivebranes wrap Cayley 4-cycles in $Spin(7)$ manifolds the five normal directions consist of four directions tangent to the $Spin(7)$ manifold and one flat direction normal to the $Spin(7)$ manifold. We thus again decompose $SO(5) \to SO(4) \cong SU(2)^+ \times SU(2)^-$, where the superscripts indicate the self-dual and anti-self-dual parts. Recall that the normal bundle of Cayley 4-cycles is given by $S_- \otimes V$ where S_- is the $SU(2)$ bundle of negative chirality spinors on Σ_3 and V is a rank $SU(2)$ bundle. In the non-generic case when V is trivial, for example for the $Spin(7)$ manifolds in [**28**], then the identification of the gauge fields is clear: if $SU(2)^{\pm}$ are the self-dual and anti-self-dual parts of the $SO(4)$ spin connection on Σ_4 then we should identify the $SU(2)^-$ part of the spin connection with $SU(2)^-$ gauge fields and set the $SU(2)^+$ gauge fields to zero. As for the co-associative 4-cycles, Einstein's equations imply that the Weyl tensor is self-dual, $\bar{W}^- = 0$.

With this data the BPS equations can easily be derived and we refer to [**110, 3, 71, 68**] for the explicit equations. They have a similar appearance to those of the SLAG 3-cycle case, with the addition of an extra term coming from the three-forms S for most of the 4-cycle cases and the SLAG 5-cycles. When the curvature of the cycle is negative, $l = -1$, in all cases except for Kähler 4-cycles in CY_3, we obtain an $AdS_{7-d} \times \Sigma_d$ fixed point. When $l = 1$, only for SLAG 5-cycles do we find such a fixed point. This is summarised in table 5.

We note that using exactly the same ansatz for the D=7 supergravity fields, some additional non-supersymmetric solutions of the form $AdS_{7-d} \times \Sigma_d$ were found [**62**]. In addition, by considering the possibility of extra scalar fields being excited, one more AdS_3 solution was found. Such solutions could be dual to non-supersymmetric conformal field theories, that are related to wrapped fivebranes with supersymmetry broken. To develop this interpretation it is necessary that the solutions be stable, which is difficult to determine. A preliminary perturbative investigation revealed that some of these solutions are unstable. We have also summarised these solutions in table 5.

5.6. Wrapped Membranes and D3-branes. D=11 supergravity solutions describing membranes wrapping Kähler 2-cycles can be found in an analogous manner [**70**]. The appropriate gauged supergravity for this case is maximal $SO(8)$ gauged supergravity in D=4 [**47, 48**] which can be obtained from a consistent truncation of the dimensional reduction of D=11 supergravity on a seven-sphere [**50, 49**]. The vacuum solution of this theory is AdS_4 and this uplifts to $AdS_4 \times S^7$, which is the near horizon limit of the planar membrane solution. More general

spacetime	embedding	cycle Σ_n	supersymmetry
$AdS_5 \times \Sigma_2$	Kähler 2-cycle in CY_2	H^2	yes
		S^2	no*
	Kähler 2-cycle in CY_3	H^2	yes
		H^2	no
$AdS_4 \times \Sigma_3$	SLAG 3-cycle in CY_3	H^3	yes
		H^3	no
	Associative 3-cycle	H^3	yes
		H^3	no
$AdS_3 \times \Sigma_4$	Co-associative 4-cycle	C^4_-	yes
	SLAG 4-cycle in CY_4	H^4	yes
		H^4	no
		S^4	no
	Kähler 4-cycle in CY_4	K^4_-	yes
		K^4_+	no
	Cayley 4-cycle	C^4_-	yes
		C^4_-	no
		CP^2, S^4	no
	CLAG 4-cycle in HK_8	B	yes
		B	no
		CP^2	no
	SLAG 4-cycle in $CY_2 \times CY_2$	$H^2 \times H^2$	yes
		$H^2 \times H^2$	no*
		$S^2 \times S^2$	no*
		$S^2 \times H^2$	no*
$AdS_2 \times \Sigma_5$	SLAG 5-cycle in CY_5	H^5	yes
		S^5	yes
	SLAG 5-cycle in $CY_2 \times CY_3$	$H^3 \times H^2$	yes
		$S^2 \times H^3$	no
		$S^2 \times H^3$	no

TABLE 5. AdS fixed point solutions for wrapped fivebranes: C_- and $K_\pm$ are conformally half-flat and Kähler–Einstein metrics with the subscript denoting positive or negative scalar curvature and B is the Bergman metric. Note that we can also take quotients of all cycles by discrete groups of isometries and this preserves supersymmetry. * denotes a solution shown to be unstable.

solutions can be found that uplift to solutions describing the near horizon limit of wrapped membranes.

Actually, the general formulae for obtaining $SO(8)$ gauged supergravity from the dimensional reduction of D=11 supergravity on the seven-sphere are rather implicit and not in a form that is useful for uplifting general solutions. Luckily, there is a further consistent truncation of the $SO(8)$ gauged supergravity theory to a $U(1)^4$ gauged supergravity where the formulae are known explicitly (in the special case that the axion fields are zero) [**39**] and this is sufficient for the construction of D=11 wrapped membrane solutions.

To see why, let us describe the ansatz for the gauge fields for the D=4 solutions. If we consider a probe membrane wrapping a Kähler 2-cycle in a CY_n then the eight directions transverse to the membrane consist of $2n-2$ directions that are tangent to the CY_n and $10-2n$ flat directions that are normal to the CY_n. In addition the normal bundle of the Kähler 2-cycle in CY_n has structure group $U(n-1) \cong U(1) \times SU(n-1)$, and recall (4.5). Thus, in the D=4 supergravity, we should first decompose $SO(8) \to SO(2n-2) \times SO(10-2n)$ and only have non-vanishing gauge fields in $U(n-1) \subset SO(2n-2)$. The gauge fields in the $U(1)$ factor of $U(n-1) \cong U(1) \times SU(n-1)$ are then identified with the $U(1)$ spin connection on the 2-cycle, corresponding to (4.5). In the solutions that have been constructed, the remaining $SU(n-1)$ gauge fields are set to zero, which corresponds to the normal bundle of the Kähler 2-cycle being non-generic when $n \geq 3$. Thus, the ansatz for the gauge fields is such that they always lie within the maximal Cartan subalgebra $U(1)^4$ of $SO(8)$ and hence the truncation formulae of [**39**] can be used.

Once again the metric on the 2-cycle is Einstein and hence is either S^2/Γ for $l = 1$ or H^2/Γ for $l = -1$. In particular, the cycle can be an arbitrary Riemann surface. General BPS equations have been found and analysed numerically. Interestingly, $AdS_2 \times \Sigma_2$ fixed points are found only for $l = -1$ and only for the cases of CY_4 and CY_5. When uplifted to D=11 these solutions become a warped product with a non-round seven-sphere that is non-trivially fibred over the cycle. The AdS_2 fixed point solutions should be dual to superconformal quantum mechanics living on the wrapped membranes.

The appropriate gauged supergravity theory for finding D=10 type IIB solutions describing D3-branes wrapping various calibrated cycles is the maximally supersymmetric $SO(6)$ gauged supergravity in D=5 [**85**]. This can be obtained from the consistent truncation of the dimensional reduction of the type IIB supergravity on a five-sphere. In particular, the vacuum solution is AdS_5 and this uplifts to $AdS_5 \times S^5$ which is the near horizon limit of the planar D3-brane. Actually the general formulae for this reduction are not yet known and one has to exploit further consistent truncations that are known [**129, 39, 107, 43**].

All cases in table 3 have been investigated, and BPS equations have been found and analysed. Once again, in the solutions, the 2- and 3-cycles that the D3-branes wrap have Einstein metrics and hence have constant curvature. D3-branes wrapping Kähler 2-cycles in CY_2 and CY_3 were studied in [**110**] while the CY_4 case was analysed in [**115**]. $AdS_3 \times H^2/\Gamma$ fixed points were found for the CY_3 and CY_4 cases. D3-branes wrapping associative 3-cycles were analysed in [**118**] and an $AdS_2 \times H^3/\Gamma$ fixed point was found. Finally, D3-branes wrapping SLAG 3-cycles were studied in [**115**] and no AdS_2 fixed point was found. Note that non-supersymmetric AdS solutions were sought in [**115**] for both wrapped membranes and D3-branes, generalising the fivebrane solutions in [**62**], but none were found.

5.7. Other wrapped brane solutions. Let us briefly mention some other supergravity solutions describing wrapped branes that have been constructed.

D6-branes of type IIA string theory carry charge under the $U(1)$ gauge field arising from the Kaluza-Klein reduction of D=11 supergravity on an S^1 (the field $C^{(1)}$ in (5.1)). The planar D6-brane uplifts to pure geometry: $\mathbb{R}^{1,6} \times M_4$, where M_4 is Taub-NUT space with $SU(2)$ holonomy. Similarly, when D6-branes wrap calibrated cycles they uplift to other special holonomy manifolds in D=11 and this has been studied in, e.g. [**1, 8, 80, 53, 91, 82**]. Other solutions related to wrapped D6-branes that are dual to non-commutative field theories have been studied in [**25, 26**].

There are solutions of massive type IIA supergravity with AdS_6 factors which are dual to the five-dimensional conformal field theory arising on the D4-D8-brane system [**56, 23**]. Supersymmetric and non-supersymmetric solutions describing the D4-D8 system wrapped on various calibrated cycles were found in [**119, 115**].

For some other supergravity solutions with possible applications to AdS/CFT, that are somewhat related to those described here, see [**7, 41, 54, 55, 24, 32, 105, 31, 106**].

6. Discussion

We have explained in some detail the construction of supergravity solutions describing branes wrapping calibrated cycles. There are a number of issues that are worth further investigation.

It seems plausible that the BPS equations can be solved exactly. To date this has only been achieved in a few cases. They were solved for the case of membranes wrapping Kähler 2-cycles in CY_5 [**70**], but this case is special in that all scalar fields in the gauged supergravity are set to zero. When there is a non-vanishing scalar field, the BPS equations were solved exactly for some cases in [**110**] and they have been partially integrated for other cases. Of particular interest are the exact solutions corresponding to the flows from an AdS_D region to an $AdS_{D-d}\times\Sigma_d$ fixed point (for example, one of the dashed lines in figure 1) as they are completely regular solutions.

For the case of membranes wrapping Kähler 2-cycles in CY_5, the general flow solution can be viewed as the "topological" AdS_4 black holes discussed in [**29**]. When $l=-1$, there is a supersymmetric rotating generalisation of this black hole [**29**]: when it is uplifted to D=11, it corresponds to waves on the wrapped membrane [**70**]. The rotating solution is completely regular provided that the angular momentum is bounded. It would be interesting to understand this bound from the point of view of the dual field theory. In addition, the existence of this rotating solution suggests, that for all of the regular flow solutions of wrapped branes starting from an AdS_D region and flowing to an $AdS_{D-d}\times\Sigma_d$ region, there should be rotating generalisations that are waiting to be found.

In all of the supergravity solutions describing wrapped branes that have been constructed, the cycle has an Einstein metric on it. It would be interesting if a more general ansatz could be found in which this condition is relaxed. While this seems possible, it may not be possible to find explicit solutions. In some cases, such as fivebranes wrapping Kähler 2-cycles in CY_3, we noted that the solutions constructed correspond to fivebranes wrapping cycles with non-generic normal bundles. This was because certain gauge fields were set to zero. We expect that more general solutions can be found corresponding to generic normal bundles. Note that such a solution, with an AdS factor, was found for D3-branes wrapping Kähler 2-cycles in CY_3 [**110**].

It would also be interesting to construct more general solutions that describe the wrapped branes beyond the near horizon limit. Such solutions would asymptote to a special holonomy manifold, which would necessarily be non-compact in order that the solution can carry non-zero flux (a no-go theorem for D=11 supergravity solutions with flux is presented in [**76**]). It seems likely that solutions can be found that asymptote to the known cohomogeneity-one special holonomy manifolds. For example, the deformed conifold is a cohomogeneity-one CY_3 that is a regular deformation of the conifold. It has a SLAG three-sphere and topologically the manifold is the cotangent bundle of the three-sphere, $T^*(S^3)$. It should be possible to generalise the solutions describing fivebranes wrapping SLAG three-spheres in the near horizon limit to solutions that include an asymptotic region far from the branes that approaches the conifold metric. Of course, these solutions will still be singular in the near horizon limit. It will be particularly interesting to construct similar solutions for the $l = -1$ case. $T^*(H^3)$ admits a CY_3 metric, with a SLAG H^3, but there is a singularity at some finite distance from the SLAG 3-cycle. There may be a solution with non-zero flux that interpolates from this singular behaviour down to the regular near horizon solutions that we constructed. Alternatively, it may be that the flux somehow "pushes off" the singularity to infinity and the entire solution is regular. The construction of these more general solutions, when the four-sphere transverse to the fivebrane is allowed to get large, will necessarily require new techniques, as they cannot be found in the gauged supergravity.

The construction of the wrapped brane solutions using gauged supergravity is rather indirect and it is desirable to characterise the D=11 geometries more directly. For example, this may lead to new methods to generalise the solutions along the lines mentioned above. One approach, is to guess general ansätze for D=11 supergravity configurations that might describe wrapped brane solutions and then impose the conditions to have supersymmetry. This approach has its origins in the construction of the intersecting brane solutions, reviewed in [**64, 133**], and was further extended in e.g. [**54, 35, 98, 99**]. Recently, it has been appreciated that it is possible to systematically characterise supersymmetric solutions of supergravity theories with non-zero fluxes using the notion of G-structures [**67, 60, 101, 74, 76, 75**] (see also [**66, 63**]). In particular, it was emphasised in some of these works that generalised calibrations play a central role and this is intimately connected with the fact that supergravity solutions with non-vanishing fluxes arise when branes wrap calibrated cycles. It should be noted that while these techniques provide powerful ways of characterising the D=11 geometries it is often difficult to obtain explicit examples: indeed, even recovering the known explicit solutions found via gauged supergravity can be non-trivial (see e.g. [**74**]).

The supergravity solutions can be used to learn a lot about the dual conformal field theory, assuming the AdS/CFT correspondence is valid. For example, the AdS fixed points can be used to determine the spectrum and correlation functions of the operators in the dual field theory. For the case of wrapped D3-branes, since the dual field theory is related to $\mathcal{N} = 4$ super Yang-Mills theory, some detailed comparisons can be made [**110**]. For the fivebrane case it will be more difficult to do this since the conformal field theory living on the fivebrane is still poorly understood. Perhaps some detailed comparisons can be made for the wrapped membranes.

Recently some new supersymmetric solutions with AdS factors were constructed in [**38, 37**]. It will be interesting to determine their dual CFT interpretation and to see if they are related to wrapped branes. The non-supersymmetric solutions containing AdS factors found in [**62**] might be dual to non-supersymmetric conformal field theories. A necessary requirement is that the solutions are stable: it would be useful to complete the preliminary analysis of the perturbative stability undertaken in [**62**].

Supergravity solutions that are dual to supersymmetric quantum field theories that are not conformally invariant can be constructed using wrapped NS 5-branes of type IIB string theory. The near horizon limit of the planar NS 5-brane is dual to what is known as "little string theory" in six dimensions (for a review see [**4**]). These still mysterious theories are not local quantum field theories but at low energies they give rise to supersymmetric Yang-Mills (SYM) theory in six dimensions. As a consequence, the geometries describing NS 5-branes wrapped on various calibrated cycles encode information about various SYM theories in lower-dimensions. The geometries describing NS 5-branes wrapped on Kähler 2-cycles in CY_3 were constructed in [**34, 111**] and are dual to $\mathcal{N} = 1$ SYM theory in four dimensions. If the 5-branes are wrapped on Kähler 2-cycles in CY_2 the geometries are dual to $\mathcal{N} = 2$ SYM theory in four dimensions [**69, 22**] (see also [**94**]). By wrapping on associative 3-cycles one finds geometries that encode information about $\mathcal{N} = 1$ SYM in D=3 [**3, 34, 130, 109, 81**], while wrapping on SLAG 3-cycles one finds $\mathcal{N} = 2$ SYM in D=3 [**67, 83**]. For the latter case, the solutions presented in [**67, 83**] are singular and correspond to vanishing Chern-Simons form in the dual SYM theory. There are strong physical arguments that suggest there are more general regular solutions that are dual to SYM with non-vanishing Chern-Simons form, and it would be very interesting to construct them. Note that supergravity solutions describing NS 5-branes wrapping various 4-cycles were found in [**115**]. The D=10 geometry for wrapped NS 5-branes has been analysed in some detail in [**61, 67, 60, 101, 74, 75**].

We hope to have given the impression that, while much is now known about supergravity solutions describing branes wrapped on calibrated cycles, there is still much to be understood.

References

1. B. S. Acharya, *On realising N = 1 super Yang-Mills in M theory*, hep-th/0011089.
2. B. S. Acharya, J. M. Figueroa-O'Farrill, C. M. Hull, and B. Spence, *Branes at conical singularities and holography*, Adv. Theor. Math. Phys. **2** (1999), 1249–1286.
3. B. S. Acharya, J. P. Gauntlett, and N. Kim, *Fivebranes wrapped on associative three-cycles*, Phys. Rev. **D63** (2001), 106003.
4. O. Aharony, *A brief review of 'little string theories'*, Class. Quant. Grav. **17** (2000), 929–938.
5. ———, *The non-AdS/non-CFT correspondence, or three different paths to QCD*, hep-th/0212193.
6. O. Aharony, S. S. Gubser, J. M. Maldacena, H. Ooguri, and Y. Oz, *Large N field theories, string theory and gravity*, Phys. Rept. **323** (2000), 183–386.
7. M. Alishahiha and Y. Oz, *AdS/CFT and BPS strings in four dimensions*, Phys. Lett. **B465** (1999), 136–141.
8. M. Atiyah, J. M. Maldacena, and C. Vafa, *An M-theory flop as a large N duality*, J. Math. Phys. **42** (2001), 3209–3220.
9. M. Atiyah and E. Witten, *M-theory dynamics on a manifold of G(2) holonomy*, Adv. Theor. Math. Phys. **6** (2003), 1–106.
10. I. Bandos, K. Lechner, A. Nurmagambetov, P. Pasti, D. P. Sorokin, and M. Tonin, *Covariant action for the super-five-brane of M-theory*, Phys. Rev. Lett. **78** (1997), 4332–4334.
11. O. Barwald, N. D. Lambert, and P. C. West, *A calibration bound for the M-theory fivebrane*, Phys. Lett. **B463** (1999), 33–40.

12. K. Becker, M. Becker, and A. Strominger, *Five-branes, membranes and nonperturbative string theory*, Nucl. Phys. **B456** (1995), 130–152.
13. Katrin Becker, M. Becker, D. R. Morrison, H. Ooguri, Y. Oz, and Z. Yin, *Supersymmetric cycles in exceptional holonomy manifolds and Calabi-Yau 4-folds*, Nucl. Phys. **B480** (1996), 225–238.
14. E. Bergshoeff, R. Kallosh, T. Ortin, and G. Papadopoulos, *kappa-symmetry, supersymmetry and intersecting branes*, Nucl. Phys. **B502** (1997), 149–169.
15. E. Bergshoeff, E. Sezgin, and P. K. Townsend, *Supermembranes and eleven-dimensional supergravity*, Phys. Lett. **B189** (1987), 75–78.
16. ———, *Properties of the eleven-dimensional super membrane theory*, Ann. Phys. **185** (1988), 330.
17. E. Bergshoeff, D. P. Sorokin, and P.K. Townsend, *The M5-brane hamiltonian*, Nucl. Phys. **B533** (1998), 303–316.
18. M. Bershadsky, C. Vafa, and V. Sadov, *D-branes and topological field theories*, Nucl. Phys. **B463** (1996), 420–434.
19. M. Bertolini, *Four lectures on the gauge-gravity correspondence*, Internat. J. Modern Phys. A. **18** (2003) 5647-5711.
20. M. Bertolini, P. Di Vecchia, M. Frau, A. Lerda, R. Marotta, and I. Pesando, *Fractional D-branes and their gauge duals*, JHEP **02** (2001), 014.
21. F. Bigazzi, A. L. Cotrone, M. Petrini, and A. Zaffaroni, *Supergravity duals of supersymmetric four-dimensional gauge theories*, hep-th/0303191.
22. F. Bigazzi, A. L. Cotrone, and A. Zaffaroni, *$N = 2$ gauge theories from wrapped five-branes*, Phys. Lett. **B519** (2001), 269–276.
23. A. Brandhuber and Y. Oz, *The D4-D8 brane system and five dimensional fixed points*, Phys. Lett. **B460** (1999), 307–312.
24. B. Brinne, A. Fayyazuddin, S. Mukhopadhyay, and D. J. Smith, *Supergravity M5-branes wrapped on Riemann surfaces and their QFT duals*, JHEP **12** (2000), 013.
25. J. Brugués, J. Gomis, T. Mateos, and T. Ramirez, *Commutative and noncommutative $N = 2$ SYM in 2+1 from wrapped D6-branes*, Classical Quantum Gravity **20** (2003) S441-S448.
26. ———, *Supergravity duals of noncommutative wrapped D6 branes and supersymmetry without supersymmetry*, JHEP **10** (2002), 016.
27. R. L. Bryant, *Pseudo-Riemannian metrics with parallel spinor fields and vanishing Ricci tensor*, Global analysis and harmonic analysis (Marseille-Luminy, 1999), Sémin. Congr., vol. 4, Soc. Math. France, Paris, 2000, pp. 53–94.
28. R. L. Bryant and S. M. Salamon, *On the construction of some complete metrics with exceptional holonomy*, Duke Math. J. **58** (1989), no. 3, 829–850.
29. M. M. Caldarelli and D. Klemm, *Supersymmetry of anti-de sitter black holes*, Nucl. Phys. **B545** (1999), 434–460.
30. I. C. G. Campbell and P. C. West, *N=2 $D = 10$ nonchiral supergravity and its spontaneous compactification*, Nucl. Phys. **B243** (1984), 112.
31. A. H. Chamseddine and W. A. Sabra, *Magnetic and dyonic black holes in $D = 4$ gauged supergravity*, Phys. Lett. **B485** (2000), 301–307.
32. ———, *Magnetic strings in five dimensional gauged supergravity theories*, Phys. Lett. **B477** (2000), 329–334.
33. A. Chamseddine and M. S. Volkov, *Non-Abelian solitons in $N = 4$ gauged supergravity and leading order string theory*, Phys. Rev. **D57** (1998), 6242–6254.
34. ———, *Non-Abelian vacua in $d = 5$, $N = 4$ gauged supergravity*, JHEP **04** (2001), 023.
35. H. Cho, M. Emam, D. Kastor and J. Traschen, *Calibrations and Fayyazuddin-Smith spacetimes*, Phys. Rev. **D63** (2001) 064003.
36. E. Cremmer, B. Julia, and J. Scherk, *Supergravity theory in 11 dimensions*, Phys. Lett. **B76** (1978), 409–412.
37. S. Cucu, H. Lü, and J. F. Vázquez-Poritz, *Interpolating from $AdS_{D-2} \times S^2$ to AdS_D*, Nucl. Phys. **B677** (2004) 181-222.
38. ———, *A supersymmetric and smooth compactification of M-theory to AdS_5*, Phys. Lett. **B568** (2003) 261-269.
39. M. Cvetic, M. J. Duff, P. Hoxha, J. T. Liu, H. Lu, J. X. Lu, R. Martinez-Acosta, C.N. Pope, H. Sati, and T.A. Tran, *Embedding AdS black holes in ten and eleven dimensions*, Nucl. Phys. **B558** (1999), 96–126.

40. M. Cvetic, G. W. Gibbons, H. Lü, and C. N. Pope, *Special holonomy spaces and M-theory*, Unity from Duality, Gravity, Gauge Theory and Strings (Les Houches 2001) 523-545. Nato Adv. Study Inst.
41. M. Cvetic, H. Lu, C. N. Pope, and J. F. Vazquez-Poritz, *Ad S in warped spacetimes*, Phys. Rev. **D62** (2000), 122003.
42. M. Cvetic, H. Lu, and C. N. Pope, *Brane resolution through transgression*, Nucl. Phys. **B600** (2001), 103–132.
43. M. Cvetic, H. Lu, C. N. Pope, A. Sadrzadeh, and T. A. Tran, *Consistent SO(6) reduction of type IIB supergravity on* S^5, Nucl. Phys. **B586** (2000), 275–286.
44. J. Dadok and R. Harvey, *Calibrations and spinors*, Acta Math. **170** (1993), no. 1, 83–120.
45. J. Dadok, R. Harvey, and F. Morgan, *Calibrations on* $\mathbf{R}^8$, Trans. Amer. Math. Soc. **307** (1988), no. 1, 1–40.
46. S. P. de Alwis, *A note on brane tension and M-theory*, Phys. Lett. **B388** (1996), 291–295.
47. B. de Wit and H. Nicolai, *N=8 supergravity*, Nucl. Phys. **B208** (1982), 323.
48. ______, *N=8 supergravity with local SO(8)×SU (8) invariance*, Phys. Lett. **B108** (1982), 285.
49. ______, *The consistency of the* S^7 *truncation in* $d = 11$ *supergravity*, Nucl. Phys. **B281** (1987), 211.
50. B. de Wit, H. Nicolai, and N. P. Warner, *The embedding of gauged N=8 supergravity into d = 11 supergravity*, Nucl. Phys. **B255** (1985), 29.
51. M. J. Duff and J. T. Liu, *Hidden spacetime symmetries and generalized holonomy in M-theory*, Nucl. Phys. **B674** (2003) 217-230.
52. M. J. Duff, J. T. Liu, and R. Minasian, *Eleven-dimensional origin of string / string duality: A one-loop test*, Nucl. Phys. **B452** (1995), 261–282.
53. J. D. Edelstein and C. Nunez, *D6 branes and M-theory geometrical transitions from gauged supergravity*, JHEP **04** (2001), 028.
54. A. Fayyazuddin and D. J. Smith, *Localized intersections of M5-branes and four-dimensional superconformal field theories*, JHEP **04** (1999), 030.
55. ______, *Warped AdS near-horizon geometry of completely localized intersections of M5-branes*, JHEP **10** (2000), 023.
56. S. Ferrara, A. Kehagias, H. Partouche, and A. Zaffaroni, *Ad* S_6 *interpretation of 5D superconformal field theories*, Phys. Lett. **B431** (1998), 57–62.
57. J. M. Figueroa-O'Farrill, *Breaking the M-waves*, Class. Quant. Grav. **17** (2000), 2925–2948.
58. J. M. Figueroa-O'Farrill and G. Papadopoulos, *Maximally supersymmetric solutions of ten-dimensional and eleven-dimensional supergravities*, JHEP **03** (2003) 048.
59. D. Z. Freedman, S. S. Gubser, K. Pilch, and N. P. Warner, *Renormalization group flows from holography supersymmetry and a c-theorem*, Adv. Theor. Math. Phys. **3** (1999), 363–417.
60. T. Friedrich and S. Ivanov, *Killing spinor equations in dimension 7 and geometry of integrable* G_2*-manifolds*, J. Geom. Phys. **48** (2003) 1-11.
61. ______, *Parallel spinors and connections with skew-symmetric torsion in string theory*, Asian J. Math. **6** (2001), no. 2, 303–335.
62. J. P. Gauntlett, N. Kim, S. Pakis, and D. Waldram, *M-theory solutions with AdS factors*, Class. Quant. Grav. **19** (2002), 3927–3946.
63. J.P. Gauntlett and J.B. Gutowski, *All supersymmetric solutions of minimal gauged supergravity in five dimensions*, Phys. Rev. **D68** (2003) 105009.
64. J. P. Gauntlett, *Intersecting branes*, hep-th/9705011.
65. ______, *Membranes on fivebranes*, Adv. Theor. Math. Phys. **3** (1999), 775–790.
66. J. P. Gauntlett, J. B. Gutowski, C. M. Hull, Stathis Pakis, and H. S. Reall, *All supersymmetric solutions of minimal supergravity in five dimensions*, Classical Quantum Gravity **20** (2003) 4587-4634.
67. J. P. Gauntlett, Kim, D. N., Martelli, and D. Waldram, *Fivebranes wrapped on SLAG three-cycles and related geometry*, JHEP **11** (2001), 018.
68. J. P. Gauntlett and N. Kim, *M-fivebranes wrapped on supersymmetric cycles II*, Phys. Rev. **D65** (2002), 086003.
69. J. P. Gauntlett, N. Kim, D. Martelli, and D. Waldram, *Wrapped fivebranes and* $N = 2$ *super Yang-Mills theory*, Phys. Rev. **D64** (2001), 106008.
70. J. P. Gauntlett, N. Kim, S. Pakis, and D. Waldram, *Membranes wrapped on holomorphic curves*, Phys. Rev. **D65** (2002), 026003.

71. J. P. Gauntlett, N. Kim, and D. Waldram, *M-fivebranes wrapped on supersymmetric cycles*, Phys. Rev. **D63** (2001), 126001.
72. J. P. Gauntlett, N. D. Lambert, and P. C. West, *Supersymmetric fivebrane solitons*, Adv. Theor. Math. Phys. **3** (1999), 91–118.
73. J. P. Gauntlett, N. D. Lambert, and P. C. West, *Branes and calibrated geometries*, Commun. Math. Phys. **202** (1999), 571–592.
74. J. P. Gauntlett, D. Martelli, S. Pakis, and D. Waldram, *G-structures and wrapped NS5-branes*, hep-th/0205050.
75. J. P. Gauntlett, D. Martelli, and D. Waldram, *Superstrings with intrinsic torsion*, hep-th/0302158.
76. J. P. Gauntlett and S. Pakis, *The geometry of D= 11 Killing spinors*, JHEP **0304** (2003) 039.
77. F. Giani and M. Pernici, *N=2 supergravity in ten-dimensions*, Phys. Rev. **D30** (1984), 325–333.
78. G. W. Gibbons, G. T. Horowitz, and P.K. Townsend, *Higher dimensional resolution of dilatonic black hole singularities*, Class. Quant. Grav. **12** (1995), 297–318.
79. G. W. Gibbons and G. Papadopoulos, *Calibrations and intersecting branes*, Commun. Math. Phys. **202** (1999), 593–619.
80. J. Gomis, *D-branes, holonomy and M-theory*, Nucl. Phys. **B606** (2001), 3–17.
81. ———, *On SUSY breaking and χSB from string duals*, Nucl. Phys. **B624** (2002), 181–199.
82. J. Gomis and T. Mateos, *D6 branes wrapping Kaehler four-cycles*, Phys. Lett. **B524** (2002), 170–176.
83. J. Gomis and J. G. Russo, *$D = 2+1$ $N = 2$ Yang-Mills theory from wrapped branes*, JHEP **10** (2001), 028.
84. S. S. Gubser, I. R. Klebanov and A. M. Polyakov, *Gauge theory correlators from non-critical string theory*, Phys. Lett. B **428** (1998), 105-114.
85. M. Gunaydin, L. J. Romans, and N. P. Warner, *Compact and noncompact gauged supergravity theories in five-dimensions*, Nucl. Phys. **B272** (1986), 598.
86. J. Gutowski, S. Ivanov, and G. Papadopoulos, *Deformations of generalized calibrations and compact non-Kähler manifolds with vanishing first Chern class*, Asian J. Math. **7** (2003) 39-79.
87. J. Gutowski and G. Papadopoulos, *AdS calibrations*, Phys. Lett. **B462** (1999), 81–88.
88. J. Gutowski, G. Papadopoulos, and P. K. Townsend, *Supersymmetry and generalized calibrations*, Phys. Rev. **D60** (1999), 106006.
89. R. Harvey, *Spinors and calibrations*, Perspectives in Mathematics, vol. 9, Academic Press Inc., Boston, MA, 1990.
90. R. Harvey and H. B. Lawson, Jr., *Calibrated geometries*, Acta Math. **148** (1982), 47–157.
91. R. Hernandez, *Branes wrapped on coassociative cycles*, Phys. Lett. **B521** (2001), 371–375.
92. P. Hořava and E. Witten, *Eleven-dimensional supergravity on a manifold with boundary*, Nucl. Phys. **B475** (1996), 94–114.
93. ———, *Heterotic and type I string dynamics from eleven dimensions*, Nucl. Phys. **B460** (1996), 506–524.
94. K. Hori and A. Kapustin, *Worldsheet descriptions of wrapped NS five-branes*, JHEP **11** (2002), 038.
95. P. S. Howe and E. Sezgin, *$D = 11$, $p = 5$*, Phys. Lett. **B394** (1997), 62–66.
96. P. S. Howe, E. Sezgin, and P. C. West, *Covariant field equations of the M-theory five-brane*, Phys. Lett. **B399** (1997), 49–59.
97. P. S. Howe and P. C. West, *The complete N=2, $d = 10$ supergravity*, Nucl. Phys. **B238** (1984), 181.
98. T. Z. Husain, *M2-branes wrapped on holomorphic curves*, JHEP **12** (2003) 037.
99. ———, *That's a wrap!*, JHEP **4** (2003) 053.
100. M. Huq and M. A. Namazie, *Kaluza-Klein supergravity in ten-dimensions*, Class. Quant. Grav. **2** (1985), 293.
101. S. Ivanov, *Connection with torsion, parallel spinors and geometry of Spin(7) manifolds*, math.dg/0111216.
102. A. Khavaev, K. Pilch, and N. P. Warner, *New vacua of gauged $N = 8$ supergravity in five dimensions*, Phys. Lett. **B487** (2000), 14–21.

103. I. R. Klebanov and E. Witten, *Superconformal field theory on threebranes at a calabi-yau singularity*, Nucl. Phys. **B536** (1998), 199–218.
104. I. R. Klebanov and M. J. Strassler, *Supergravity and a confining gauge theory: Duality cascades and χSB-resolution of naked singularities*, JHEP **08** (2000), 052.
105. D. Klemm and W. A. Sabra, *Supersymmetry of black strings in $D = 5$ gauged supergravities*, Phys. Rev. **D62** (2000), 024003.
106. ———, *Charged rotating black holes in 5d Einstein-Maxwell- (A)dS gravity*, Phys. Lett. **B503** (2001), 147–153.
107. H. Lu, C. N. Pope, and T. A. Tran, *Five-dimensional $N = 4$, $SU(2) \times U(1)$ gauged supergravity from type IIB*, Phys. Lett. **B475** (2000), 261–268.
108. J. M. Maldacena, *The large N limit of superconformal field theories and supergravity*, Adv. Theor. Math. Phys. **2** (1998), 231–252.
109. J. M. Maldacena and H. Nastase, *The supergravity dual of a theory with dynamical supersymmetry breaking*, JHEP **09** (2001), 024.
110. J. M. Maldacena and C. Nuñez, *Supergravity description of field theories on curved manifolds and a no go theorem*, Int. J. Mod. Phys. **A16** (2001), 822–855.
111. ———, *Towards the large N limit of pure $N = 1$ super Yang Mills*, Phys. Rev. Lett. **86** (2001), 588–591.
112. M. Marino, R. Minasian, G. W. Moore, and A. Strominger, *Nonlinear instantons from supersymmetric p-branes*, JHEP **01** (2000), 005.
113. R. C. McLean, *Deformations of calibrated submanifolds*, Comm. Anal. Geom. **6** (1998), no. 4, 705–747.
114. David R. Morrison and M. Ronen Plesser, *Non-spherical horizons. I*, Adv. Theor. Math. Phys. **3** (1999), 1–81.
115. M. Naka, *Various wrapped branes from gauged supergravities*, hep-th/0206141.
116. H. Nastase, D. Vaman, and P. van Nieuwenhuizen, *Consistent nonlinear KK reduction of 11d supergravity on $AdS_7 \times S^4$ and self-duality in odd dimensions*, Phys. Lett. **B469** (1999), 96–102.
117. ———, *Consistency of the $AdS_7 \times S^4$ reduction and the origin of self-duality in odd dimensions*, Nucl. Phys. **B581** (2000), 179–239.
118. H. Nieder and Y. Oz, *Supergravity and D-branes wrapping special lagrangian cycles*, JHEP **03** (2001), 008.
119. C. Nunez, I. Y. Park, M. Schvellinger, and T. A. Tran, *Supergravity duals of gauge theories from $F(4)$ gauged supergravity in six dimensions*, JHEP **04** (2001), 025.
120. G. Papadopoulos and A. Teschendorff, *Multi-angle five-brane intersections*, Phys. Lett. **B443** (1998), 159–166.
121. ———, *Grassmannians, calibrations and five-brane intersections*, Class. Quant. Grav. **17** (2000), 2641–2662.
122. P. Pasti, D. P. Sorokin, and M. Tonin, *Covariant action for a $D = 11$ five-brane with the chiral field*, Phys. Lett. **B398** (1997), 41–46.
123. M. Pernici, K. Pilch, and P. van Nieuwenhuizen, *Gauged maximally extended supergravity in seven-dimensions*, Phys. Lett. **B143** (1984), 103.
124. M. Pernici and E. Sezgin, *Spontaneous compactification of seven-dimensional supergravity theories*, Class. Quant. Grav. **2** (1985), 673.
125. K. Pilch and N. P. Warner, *$N = 2$ supersymmetric RG flows and the IIB dilaton*, Nucl. Phys. **B594** (2001), 209–228.
126. ———, *$N = 1$ supersymmetric renormalization group flows from IIB supergravity*, Adv. Theor. Math. Phys. **4** (2002), 627–677.
127. J. Polchinski, *$N = 2$ gauge-gravity duals*, Int. J. Mod. Phys. **A16** (2001), 707–718.
128. J. Polchinski and M. J. Strassler, *The string dual of a confining four-dimensional gauge theory*, hep-th/0003136.
129. L. J. Romans, *Gauged $N=4$ supergravities in five-dimensions and their magnetovac backgrounds*, Nucl. Phys. **B267** (1986), 433.
130. M. Schvellinger and T. A. Tran, *Supergravity duals of gauge field theories from $SU(2)\times U(1)$ gauged supergravity in five dimensions*, JHEP **06** (2001), 025.
131. J. H. Schwarz, *Covariant field equations of chiral $N=2$ $D = 10$ supergravity*, Nucl. Phys. **B226** (1983), 269.

132. S. Sethi, C. Vafa, and E. Witten, *Constraints on low-dimensional string compactifications*, Nucl. Phys. **B480** (1996), 213–224.
133. D. J. Smith, *Intersecting brane solutions in string and M-theory*, Class. Quant. Grav. **20** (2003), R233.
134. K. S. Stelle, *BPS branes in supergravity*, Quantum Field Theory: perspective and prospective (Les Houches 1998) 257-350 Kluwer.
135. A. Strominger, *Superstrings with torsion*, Nucl. Phys. **B274** (1986), 253.
136. L. Susskind and E. Witten, *The holographic bound in anti-de Sitter space*, hep-th/9805114.
137. C. Vafa and E. Witten, *A one loop test of string duality*, Nucl. Phys. **B447** (1995), 261–270.
138. E. Witten, *On flux quantization in M-theory and the effective action*, J. Geom. Phys. **22** (1997), 1–13.
139. ———, *Anti-de sitter space and holography*, Adv. Theor. Math. Phys. **2** (1998), 253–291.

Department of Physics, Queen Mary, University of London, Mile End Rd, London, E1 4NS, U.K.

E-mail address: J.P.Gauntlett@qmul.ac.uk

Clay Mathematics Proceedings
Volume 3, 2004

M-theory on Manifolds with Exceptional Holonomy

Sergei Gukov

Abstract. In these lectures I review recent progress on construction of manifolds with exceptional holonomy and their application in string theory and M-theory.

1. Introduction

Recently, M-theory compactifications on manifolds of exceptional holonomy have attracted considerable attention. These models allow one to geometrically engineer various minimally supersymmetric gauge theories, which typically have a rich dynamical structure. A particularly interesting aspect of such models is the behaviour near a classical singularity, where one might expect extra massless degrees of freedom, enhancement of gauge symmetry, or a phase transition to a different theory.

In these lectures I will try to explain two interesting problems in this subject — namely, construction of manifolds with exceptional holonomy and the analysis of the physics associated with singularities — and present general methods for their solution. I will also try to make these lectures self-contained and pedagogical, so that no special background is needed. In particular, below we start with an introduction to special holonomy, then explain its relation to minimal supersymmetry and proceed to the main questions.

2. Riemannian Manifolds of Special Holonomy

2.1. Holonomy Groups. Consider an oriented manifold X of real dimension n and a vector $\vec{v}$ at some point on this manifold. One can explore the geometry of X by doing a parallel transport of $\vec{v}$ along a closed contractible path in X; see Figure1. Under such an operation the vector $\vec{v}$ may not come back to itself. In fact, generically it will transform into a different vector that depends on the geometry of X, on the path, and on the connection which was used to transport $\vec{v}$. For

2000 *Mathematics Subject Classification.* 81T30, 53C29, 83E30 .

I would like to thank the organizers of the 2002 School on "Geometry and String Theory" for a wonderful meeting. Various parts of these lectures are based on the work done together with B. Acharya, A. Brandhuber, J. Gomis, S. Gubser, X. de la Ossa, D. Tong, J. Sparks, S.-T. Yau, and E. Zaslow, whom I wish to thank for enjoyable collaboration. This research was conducted during the period Sergei Gukov served as a Clay Mathematics Institute Research Fellow.

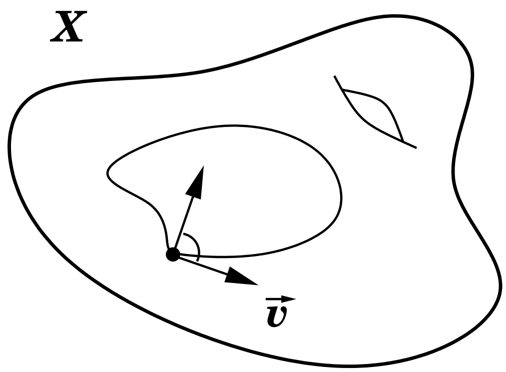


FIGURE 1. Parallel transport of a vector $\vec{v}$ along a closed path on the manifold X.

a Riemannian manifold X, the natural connection is the Levi-Civita connection. Furthermore, Riemannian geometry also tells us that the length of the vector covariantly transported along a closed path should be the same as the length of the original vector. But the orientation may be different, and this is precisely what we are going to discuss.

The relative orientation of the vector after parallel transport with respect to the orientation of the original vector $\vec{v}$ is described by *holonomy*. On an n-dimensional manifold holonomy is conveniently characterised by an element of the special orthogonal group, $SO(n)$. It is not hard to see that the set of all holonomies themselves form a group, called the *holonomy group*, where the group structure is induced by the composition of paths and an inverse corresponds to a path traversed in the opposite direction. From the way we introduced the holonomy group, $Hol(X)$, it seems to depend on the choice of the point where we start and finish parallel transport. However, for a generic choice of such point the holonomy group does not depend on it, and therefore $Hol(X)$ becomes a true geometric characteristic of the manifold X. By definition, we have

$$Hol(X) \subseteq SO(n) \tag{2.1}$$

where the equality holds for a generic Riemannian manifold X.

In some special instances, however, one finds that $Hol(X)$ is a proper subgroup of $SO(n)$. In such cases, we say that X is a special holonomy manifold or a manifold with restricted holonomy. These manifolds are in some sense distinguished, for they exhibit nice geometric properties. As we explain later in this section, these properties are typically associated with the existence of non-degenerate (in some suitable sense) p-forms which are covariantly constant. Such p-forms also serve as calibrations, and are related to the subject of minimal varieties.

The possible choices for $Hol(X) \subset SO(n)$ are limited, and were completely classified by M. Berger in 1955 [**12**]. Specifically, for X simply-connected and neither locally a product nor symmetric, the only possibilities for $Hol(X)$, other than the generic case of $SO(n)$, are $U\left(\frac{n}{2}\right)$, $SU\left(\frac{n}{2}\right)$, $Sp\left(\frac{n}{4}\right) \times Sp(1)$, $Sp\left(\frac{n}{4}\right)$, G_2[1], $Spin(7)$ or $Spin(9)$, see Table 1. The first four of these correspond, respectively, to a

[1]The fourteen-dimensional simple Lie group $G_2 \subset Spin(7)$ is precisely the automorphism group of the octonions, $\mathbb{O}$.

Metric	Holonomy	Dimension
Kähler	$U\left(\frac{n}{2}\right)$	$n =$ even
Calabi-Yau	$SU\left(\frac{n}{2}\right)$	$n =$ even
HyperKähler	$Sp\left(\frac{n}{4}\right)$	$n =$ multiple of 4
Quaternionic	$Sp\left(\frac{n}{4}\right) Sp(1)$	$n =$ multiple of 4
Exceptional	G_2	7
Exceptional	$Spin(7)$	8
Exceptional	$Spin(9)$	16

TABLE 1. Berger's list of holonomy groups.

Kähler, Calabi-Yau, quaternionic Kähler or hyper-Kähler manifold. The last three possibilities are the so-called exceptional cases, which occur only in dimensions 7, 8 and 16, respectively. The case of 16-manifold with $Spin(9)$ holonomy is in some sense trivial since the Riemannian metric on any such manifold is always symmetric [**3**].

2.2. Relation Between Holonomy and Supersymmetry. Roughly speaking, one can think of the holonomy group as a geometric characteristic of the manifold that tells us how much symmetry this manifold has. Namely, the smaller the holonomy group, the larger the symmetry of the manifold X. Conversely, for manifolds with larger holonomy groups the geometry is less restricted.

This philosophy becomes especially helpful in the physical context of superstring or M-theory compactifications on X. There, the holonomy of X becomes related to the degree of supersymmetry preserved in compactification: the manifolds with larger holonomy group typically preserve a smaller fraction of the supersymmetry. This provides a nice link between the 'geometric symmetry' (holonomy) and the 'physical symmetry' (supersymmetry). In Table 2 we illustrate this general pattern with a few important examples, which will be used later.

The first example in Table 2 is a torus, T^n, which we view as a quotient of an n-dimensional real vector space, $\mathbb{R}^n$, by a lattice. In this example, it is easy to deduce directly from our definition that $X = T^n$ has trivial holonomy group, inherited from the trivial holonomy of $\mathbb{R}^n$. Indeed, no matter which path we choose on T^n, a parallel transport of a vector $\vec{v}$ along this path always brings it back to itself. Hence, this example is the most symmetric one, in the sense of the previous paragraph, $Hol(X) = \mathbf{1}$. Correspondingly, in string theory toroidal compactifications preserve all of the original supersymmetries.

Our next example is $Hol(X) = SU(3)$, which corresponds to Calabi-Yau manifolds of complex dimension 3 (real dimension 6). These manifolds exhibit a number of remarkable properties, such as mirror symmetry, and are reasonably well studied both in the mathematical and in the physical literature. We just mention here that

Manifold X	T^n		CY_3		X_{G_2}		$X_{Spin(7)}$
$\dim_{\mathbb{R}}(X)$	n		6		7		8
$Hol(X)$	$\mathbf{1}$	$\subset$	$SU(3)$	$\subset$	G_2	$\subset$	$Spin(7)$
SUSY	1	$>$	1/4	$>$	1/8	$>$	1/16

TABLE 2. Relation between holonomy and supersymmetry for certain manifolds.

compactification on Calabi-Yau manifolds preserves 1/4 of the original supersymmetry. In particular, compactification of heterotic string theory on $X = CY_3$ yields $\mathcal{N} = 1$ effective field theory in $3+1$ dimensions.

The last two examples in Table 2 are G_2 and $Spin(7)$ manifolds; that is, manifolds with holonomy group G_2 and $Spin(7)$, respectively. They nicely fit into the general pattern, so that as we read Table 2 from left to right the holonomy increases, whereas the fraction of unbroken supersymmetry decreases. Specifically, compactification of M-theory on a manifold with G_2 holonomy leads to $\mathcal{N} = 1$ four-dimensional theory. This is similar to the compactification of heterotic string theory on Calabi-Yau three-folds. However, an advantage of M-theory on G_2 manifolds is that it is completely 'geometric'. Compactification on $Spin(7)$ manifolds breaks supersymmetry even further, to an amount which is too small to be realised in four-dimensional field theory.

Mathematically, the fact that all these manifolds preserve some supersymmetry is related to the existence of covariantly constant spinors:

$$\nabla \xi = 0 \tag{2.2}$$

For example, if $Hol(X) = G_2$ the covariantly constant spinor is the singlet in the decomposition of the spinor of $SO(7)$ into representations of G_2:

$$\mathbf{8} \rightarrow \mathbf{7} \oplus \mathbf{1}$$

Summarising, in Table 2 we listed some examples of special holonomy manifolds that will be discussed below. All of these manifolds preserve a certain fraction of supersymmetry, which depends on the holonomy group. Moreover, all of these manifolds are Ricci-flat,

$$R_{ij} = 0.$$

This useful property guarantees that all backgrounds of the form

$$\mathbb{R}^{11-n} \times X$$

automatically solve eleven-dimensional Einstein equations with vanishing source terms for matter fields.

Of particular interest are M-theory compactifications on manifolds with exceptional holonomy,

$$
\begin{array}{cc}
\text{M} - \text{theory on} & \text{M} - \text{theory on} \\
G_2 \text{ manifold} & Spin(7) \text{ manifold} \\
\Downarrow & \Downarrow \\
D = 3+1 \text{ QFT} & D = 2+1 \text{ QFT}
\end{array}
\tag{2.3}
$$

since they lead to effective theories with minimal supersymmetry in four and three dimensions, respectively. In such theories one can find many interesting phenomena, *e.g.* confinement, various dualities, rich phase structure, non-perturbative effects, *etc.* Moreover, minimal supersymmetry in three and four dimensions (partly) helps with the following important problems:

- The Hierarchy Problem
- The Cosmological Constant Problem
- The Dark Matter Problem

All this makes minimal supersymmetry very attractive and, in particular, motivates the study of M-theory on manifolds with exceptional holonomy. In this context, the spectrum of elementary particles in the effective low-energy theory and their interactions are encoded in the geometry of the space X. Therefore, understanding the latter may help us to learn more about dynamics of minimally supersymmetric field theories, or even about M-theory itself!

2.3. Invariant Forms and Minimal Submanifolds. For a manifold X, we have introduced the notion of special holonomy and related it to the existence of covariantly constant spinors on X, *cf.* (2.2). However, special holonomy manifolds can be also characerised by the existence of certain invariant forms and the corresponding minimal cycles.

Indeed, one can sandwich antisymmetric combinations of Γ-matrices with a covariantly constant spinor ξ on X to obtain antisymmetric tensor forms of various degree:

$$
\omega^{(p)} = \xi^{\dagger} \Gamma_{i_1 \ldots i_p} \xi
\tag{2.4}
$$

By construction, the p-form $\omega^{(p)}$ is covariantly constant and invariant under $Hol(X)$. In order to find all possible invariant forms on a special holonomy manifold X, we need to decompose the cohomology of X into representations[2] of $Hol(X)$ and identify all singlet components. For example, for a manifold with G_2 holonomy such a

[2]This decomposition makes sense because the Laplacian on X commutes with $Hol(X)$. In a sense, for exceptional holonomy manifolds, decomposition of cohomology groups into representations of $Hol(X)$ is a (poor) substitute for the Hodge decomposition in the realm of complex geometry.

decomposition reads [**37**]:

$$
\begin{array}{rcl}
H^0(X,\mathbb{R}) & = & \mathbb{R} \\
H^1(X,\mathbb{R}) & = & H^1_{\mathbf{7}}(X,\mathbb{R}) \\
H^2(X,\mathbb{R}) & = & H^2_{\mathbf{7}}(X,\mathbb{R}) \oplus H^2_{\mathbf{14}}(X,\mathbb{R}) \\
H^3(X,\mathbb{R}) & = & H^3_{\mathbf{1}}(X,\mathbb{R}) \oplus H^3_{\mathbf{7}}(X,\mathbb{R}) \oplus H^3_{\mathbf{27}}(X,\mathbb{R}) \\
H^4(X,\mathbb{R}) & = & H^4_{\mathbf{1}}(X,\mathbb{R}) \oplus H^4_{\mathbf{7}}(X,\mathbb{R}) \oplus H^4_{\mathbf{27}}(X,\mathbb{R}) \\
H^5(X,\mathbb{R}) & = & H^5_{\mathbf{7}}(X,\mathbb{R}) \oplus H^5_{\mathbf{14}}(X,\mathbb{R}) \\
H^6(X,\mathbb{R}) & = & H^6_{\mathbf{7}}(X,\mathbb{R}) \\
H^7(X,\mathbb{R}) & = & \mathbb{R}
\end{array} \tag{2.5}
$$

Here, we used the G_2 structure locally. The fact that the metric on X has irreducible G_2-holonomy implies global constraints on X and this forces some of the above groups to vanish when X is compact, *viz.*

$$H^k_{\mathbf{7}}(X,\mathbb{R}) = 0, \quad k = 1, \dots, 6$$

Let us now return to the construction (2.4) of the invariant forms on X. From the above decomposition we see that on a G_2 manifold such forms can appear only in degree $p = 3$ and $p = 4$. They are called *associative* and *coassociative* forms, respectively. In fact, the coassociative 4-form is the Hodge dual of the associative 3-form. These forms, which we denote Φ and $*\Phi$, enjoy a number of remarkable properties.

For example, the existence of a G_2 holonomy metric on X is equivalent to the closure and co-closure of the associative form[3],

$$
\begin{array}{rcl}
d\Phi & = & 0 \\
d * \Phi & = & 0.
\end{array} \tag{2.6}
$$

This may look a little surprising, especially since the number of metric components on a 7-manifold is different from the number of of components of a generic 3-form. However, given a G_2 holonomy metric,

$$ds^2 = \sum_{i=1}^{7} e^i \otimes e^i,$$

one can locally write the invariant 3-form Φ in terms of the vielbein e^i,

$$\Phi = \frac{1}{3!} \psi_{ijk} \ e^i e^j e^k \tag{2.7}$$

Here, ψ_{ijk} are totally antisymmetric structure constants of the imaginary octonions,

$$\sigma_i \sigma_j = -\delta_{ij} + \psi_{ijk} \ \sigma_k \ , \ i,j,k = 1, \dots, 7 \tag{2.8}$$

and G_2 is the automorphism group of the imaginary octonions. In a choice of basis the non-zero structure constants are given by

$$\psi_{ijk} = +1 \ , \ (abc) = \{(123), (147), (165), (246), (257), (354), (367)\} \ . \tag{2.9}$$

[3]Another, equivalent, condition is to say that the G_2-structure (g, Φ) is torsion-free: $\nabla \Phi = 0$.

$Hol(X)$	cycle S	$p = \dim(S)$	Deformations	dim(Def)
$SU(3)$	SLAG	3	unobstructed	$b_1(S)$
G_2	associative	3	obstructed	—
	coassociative	4	unobstructed	$b_2^+(S)$
$Spin(7)$	Cayley	4	obstructed	—

TABLE 3. Deformations of calibrated submanifolds.

It is, perhaps, less obvious that one can also locally reconstruct a G_2 metric from the associative 3-form:

$$\begin{aligned} g_{ij} &= \det(B)^{-1/9} B_{ij} \\ B_{jk} &= -\frac{1}{144} \Phi_{j i_1 i_2} \Phi_{k i_3 i_4} \Phi_{i_5 i_6 i_7} \epsilon^{i_1 \dots i_7} \end{aligned} \tag{2.10}$$

This will be useful to us in the following sections.

Similarly, on a $Spin(7)$ manifold X we find only one invariant form in degree $p = 4$, called the Cayley form, Ω. Indeed, the decomposition of the cohomology groups of X into $Spin(7)$ representations looks as follows [**37**]:

$$\begin{aligned} H^0(X,\mathbb{R}) &= \mathbb{R} \\ H^1(X,\mathbb{R}) &= H^1_{\mathbf{8}}(X,\mathbb{R}) \\ H^2(X,\mathbb{R}) &= H^2_{\mathbf{7}}(X,\mathbb{R}) \oplus H^2_{\mathbf{21}}(X,\mathbb{R}) \\ H^3(X,\mathbb{R}) &= H^3_{\mathbf{8}}(X,\mathbb{R}) \oplus H^3_{\mathbf{48}}(X,\mathbb{R}) \\ H^4(X,\mathbb{R}) &= H^4_{\mathbf{1}^+}(X,\mathbb{R}) \oplus H^4_{\mathbf{7}^+}(X,\mathbb{R}) \oplus H^4_{\mathbf{27}^+}(X,\mathbb{R}) \oplus H^4_{\mathbf{35}^-}(X,\mathbb{R}) \\ H^5(X,\mathbb{R}) &= H^5_{\mathbf{8}}(X,\mathbb{R}) \oplus H^5_{\mathbf{48}}(X,\mathbb{R}) \\ H^6(X,\mathbb{R}) &= H^6_{\mathbf{7}}(X,\mathbb{R}) \oplus H^6_{\mathbf{21}}(X,\mathbb{R}) \\ H^7(X,\mathbb{R}) &= H^7_{\mathbf{8}}(X,\mathbb{R}) \\ H^8(X,\mathbb{R}) &= \mathbb{R} \end{aligned} \tag{2.11}$$

The additional label "$\pm$" denotes self-dual/anti-self-dual four-forms, respectively. The cohomology class of the 4-form Ω generates $H^4_{\mathbf{1}^+}(X,\mathbb{R})$,

$$H^4_{\mathbf{1}^+}(X,\mathbb{R}) = \langle [\Omega] \rangle$$

Again, on a compact simply-connected $Spin(7)$ manifold we have extra constraints,

$$H^1_{\mathbf{8}} = H^3_{\mathbf{8}} = H^5_{\mathbf{8}} = H^7_{\mathbf{8}} = 0, \quad H^2_{\mathbf{7}} = H^4_{\mathbf{7}} = H^6_{\mathbf{7}} = 0$$

Another remarkable property of the invariant forms is that they represent the volume forms of the minimal submanifolds in X. The forms with these properties are called *calibrations*, and the corresponding submanifolds are called *calibrated submanifolds* [**36**]. More precisely, we say that Ψ is a calibration if it is less than or equal to the volume on each oriented p-dimensional submanifold $S \subset X$. Namely, combining the orientation of S with the restriction of the Riemann metric on X to the subspace S, we can define a natural volume form $\mathrm{vol}(T_x S)$ on the tangent space T_xS for each point $x \in S$. Then, $\Psi|_{T_xS} = \alpha \cdot \mathrm{vol}(T_xS)$ for some $\alpha \in \mathbb{R}$, and we write:

$$\Psi|_{T_xS} \leq \mathrm{vol}(T_xS)$$

if $\alpha \leq 1$. If equality holds for all points $x \in S$, then S is called a calibrated submanifold with respect to the calibration Ψ. According to this definition, the volume of a calibrated submanifold S can be expressed in terms of Ψ as:

$$\mathrm{Vol}(S) = \int_{x \in S} \Psi|_{T_xS} = \int_S \Psi \tag{2.12}$$

Since the right-hand side depends only on the cohomology class, we can write:

$$\mathrm{Vol}(S) = \int_S \Psi = \int_{S'} \Psi = \int_{x \in S'} \Psi|_{T_xS'} \leq \int_{x \in S'} \mathrm{vol}(T_xS') = \mathrm{Vol}(S')$$

for any other submanifold S' in the same homology class. Therefore, we have just demonstrated that calibrated manifolds have minimal area in their homology class. This important property of calibrated submanifolds allows us to identify them with supersymmetric cycles. In particular, branes in string theory and M-theory wrapped over calibrated submanifolds give rise to BPS states in the effective theory.

Deformations of calibrated submanifolds have been studied by McLean [**46**], and are briefly summarised in Table 3.

2.4. Why Exceptional Holonomy is Hard. Once we have introduced manifolds with restricted (or special) holonomy, let us try to explain why until recently so little was known about the exceptional cases, G_2 and $Spin(7)$. Indeed, on the physics side, these manifolds are very natural candidates for constructing minimally supersymmetric field theories from string/M-theory compactifications. Therefore, one might expect exceptional holonomy manifolds to be at least as popular and attractive as, say, Calabi-Yau manifolds. However, there are several reasons why exceptional holonomy appeared to be a difficult subject; here we will stress two of them:

- **Existence**
- **Singularities**

Let us now explain each of these problems in turn. The first problem refers to the existence of an exceptional holonomy metric on a given manifold X. Namely, it would be useful to have a general theorem which, under some favorable conditions, would guarantee the existence of such a metric. Indeed, the original Berger's classification, described earlier in this section, only tells us which holonomy groups can occur, but says nothing about examples of such manifolds or conditions under which they exist. To illustrate this further, let us recall that when we deal with Calabi-Yau manifolds we use such a theorem all the time — it is a theorem due to Yau (originally, Calabi's conjecture) which guarantees the existence of a Ricci-flat metric on a compact, complex, Kähler manifold X with $c_1(X) = 0$ [**58**]. Unfortunately, no analogue of this theorem is known in the case of G_2 and $Spin(7)$ holonomy (the local existence of such manifolds was first established in 1985 by Bryant [**14**]). Therefore, until such a general theorem is found we are limited to a case-by-case analysis of the specific examples. We will return to this problem in the next section.

The second reason why exceptional holonomy manifolds are hard is associated with singularities of these manifolds. As will be explained in the following sections, interesting physics occurs at the singularities. Moreover, the most interesting physics is associated with the types of singularities of maximal codimension, which exploit the geometry of the special holonomy manifold to the fullest. Indeed, singularities with smaller codimension can typically arise in higher-dimensional compactifications and, therefore, do not expose peculiar aspects of exceptional holonomy manifolds related to the minimal amount of supersymmetry. Until recently, little was known about these types of degenerations of manifolds with G_2 and $Spin(7)$ holonomy. Moreover, even for known examples of isolated singularities, the dynamics of the M-theory in these backgrounds was unclear. Finally, it is important to stress that the mathematical understanding of exceptional holonomy manifolds would be incomplete too without proper account being taken of singular limits.

3. Construction of Manifolds With Exceptional Holonomy

In this section we review various methods of constructing compact and non-compact manifolds with G_2 and $Spin(7)$ holonomy. In the absence of general existence theorems, akin to Yau's theorem [**58**], these methods become especially valuable. It is hard to give full justice to all the existing techniques in one section. So we will try to explain only a few basic methods, focusing mainly on those which have played an important role in recent developments in string theory. We also illustrate these general techniques with several concrete examples that will appear in the later sections.

3.1. Compact Manifolds. The first examples of metrics with G_2 and $Spin(7)$ holonomy on compact manifolds were constructed by D. Joyce [**37**]. The basic idea is to start with toroidal orbifolds of the form

$$T^7/\Gamma \qquad \text{or} \qquad T^8/\Gamma \tag{3.1}$$

where Γ is a finite group, *e.g.* a product of $\mathbb{Z}_2$ cyclic groups. Notice that T^7 and T^8 themselves can be regarded as G_2 and $Spin(7)$ manifolds, respectively. In fact, they possess infinitely many G_2 and $Spin(7)$ structures. Therefore, if Γ preserves one of these structures the quotient space automatically will be a manifold with exceptional holonomy.

Example 3.1 (Joyce [**37**]). Consider a torus T^7, parametrized by periodic variables $x_i \sim x_i + 1$, $i = 1, \ldots, 7$. As we pointed out, it admits many G_2 structures. Let us choose one of them:

$$\Phi = e^1\wedge e^2\wedge e^3 + e^1\wedge e^4\wedge e^5 + e^1\wedge e^6\wedge e^7 + e^2\wedge e^4\wedge e^6 - e^2\wedge e^5\wedge e^7 - e^3\wedge e^4\wedge e^7 - e^3\wedge e^5\wedge e^6$$

where $e^j = dx_j$. Furthermore, let us take

$$\Gamma = \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2$$

generated by three involutions

$$\begin{array}{lcl} \alpha & : & (x_1, \ldots, x_7) \mapsto (x_1, x_2, x_3, -x_4, -x_5, -x_6, -x_7) \\ \beta & : & (x_1, \ldots, x_7) \mapsto (x_1, -x_2, -x_3, x_4, x_5, \frac{1}{2} - x_6, -x_7) \\ \gamma & : & (x_1, \ldots, x_7) \mapsto (-x_1, x_2, -x_3, x_4, \frac{1}{2} - x_5, x_6, \frac{1}{2} - x_7) \end{array}$$

It is easy to check that these generators indeed satisfy $\alpha^2 = \beta^2 = \gamma^2 = 1$ and that the group $\Gamma = \langle \alpha, \beta, \gamma \rangle$ preserves the associative three-form Φ given above. It

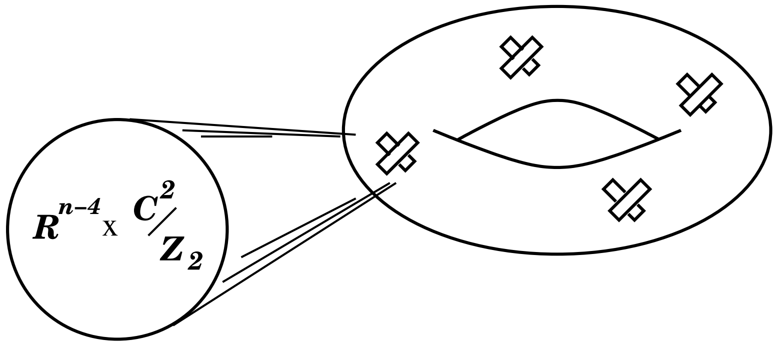


FIGURE 2. A cartoon representing a Joyce orbifold T^n/Γ with $\mathbb{C}^2/\mathbb{Z}_2$ orbifold points.

follows that the quotient space $X = T^7/\Gamma$ is a manifold with G_2 holonomy. More precisely, it is an orbifold since the group Γ has fixed points of the form $T^3 \times \mathbb{C}^2/\mathbb{Z}_2$. The existence of orbifold fixed points is a general feature of the Joyce construction.

The quotient space (3.1) typically has bad (singular) points, as shown in Fig. 2. In order to find a nice manifold X with G_2 or $Spin(7)$ holonomy one has to repair these singularities. In practice, this means removing the local neighbourhood of each singular point and replacing it with a smooth geometry, in a way which does not affect the holonomy group. This may be difficult (or even impossible) for generic orbifold singularities. However, if we have orbifold singularities that can also appear as degenerations of Calabi-Yau manifolds, then things simplify dramatically.

Suppose we have a $\mathbb{Z}_2$ orbifold, as in the previous example:

$$\mathbb{R}^{n-4} \times \mathbb{C}^2/\mathbb{Z}_2$$

where $\mathbb{Z}_2$ acts only on the $\mathbb{C}^2$ factor (by reflecting all the coordiantes). This type of orbifold singularity can be obtained as a singular limit of the A_1 ALE space:

$$\mathbb{R}^{n-4} \times \mathrm{ALE}_{A_1} \;\to\; \mathbb{R}^{n-4} \times \mathbb{C}^2/\mathbb{Z}_2$$

Since both the ALE space and its singular limit have $SU(2)$ holonomy group they represent the local geometry of the $K3$ surface. This is an important point; we used it implicitly to resolve the orbifold singularity with the usual tools from algebraic geometry. Moreover, Joyce proved that resolving orbifold singularities in this way does not change the holonomy group of the quotient space (3.1). Therefore, by the end of the day, when all singularities are removed, we obtain a smooth, compact manifold X with G_2 or $Spin(7)$ holonomy. In the above example, one finds a smooth manifold X with G_2 holonomy and Betti numbers:

$$b_2(X) = 12, \quad b_3(X) = 43$$

There are many other examples of this construction, which are modelled not only on $K3$ singularities, but also on orbifold singularities of Calabi-Yau three-folds [**37**]. More examples can be found by replacing tori in (3.1) by products of the $K3$ surface or Calabi-Yau three-folds with lower-dimensional tori. In such models, finite groups typically act as involutions on $K3$ or Calabi-Yau manifolds, to produce fixed points of a familiar kind. Again, upon resolving the singularities one finds compact, smooth manifolds with exceptional holonomy.

It may look a little disturbing that in Joyce's construction one always finds a compact manifold X with exceptional holonomy near a singular (orbifold) limit. However, from the physics point of view, this is not a problem at all since interesting phenomena usually occur when X develops a singularity. Indeed, compactification on a smooth manifold X whose dimensions are very large (compared to the Planck scale) leads to a very simple effective field theory; it is Abelian gauge theory with some number of scalar fields coupled to gravity. To find more interesting physics, such as non-Abelian gauge symmetry or chiral matter, one needs singularities.

Moreover, there is a close relation between various types of singularities and the effective physics they produce. A simple, but very important aspect of this relation is that a codimension d singularity of X can be associated with physics of $D \geq 11 - d$ dimensional field theory. For example, there is no way one can obtain four-dimensional chiral matter or parity symmetry breaking in $D = 2+1$ dimensions from a $\mathbb{C}/\mathbb{Z}_2$ singularity in X. Indeed, both of these phenomena are specific to their dimension and can not be lifted to a higher-dimensional theory. Therefore, in order to reproduce them from compactification on X one has to use the geometry of X 'to the fullest' and consider singularities of maximal codimension. This motivates us to study *isolated singular points* in G_2 and $Spin(7)$ manifolds.

Unfortunately, even though Joyce manifolds naturally admit orbifold singularities, none of them contains isolated G_2 or $Spin(7)$ singularities. Indeed, as we explained earlier, it is crucial that orbifold singularities are modelled on Calabi-Yau singularities, so that we can treat them using the familiar methods. Therefore, at best such singularities can give us the same physics as one finds in the corresponding Calabi-Yau manifolds.

Apart from a large class of Joyce manifolds, very few explicit constructions of compact manifolds with exceptional holonomy are known. One nice approach was recently suggested by A. Kovalev [**39**], where a smooth, compact 7-manifold X with G_2 holonomy is obtained by gluing 'back-to-back' two asymptotically cylindrical Calabi-Yau manifolds W_1 and W_2,

$$X \cong (W_1 \times S^1) \cup (W_2 \times S^1)$$

Although this construction is very elegant, so far it has been limited to very specific types of G_2 manifolds. In particular, it would be interesting to study deformations of these spaces and to see if they can develop isolated singularities interesting in physics. This leaves us with the following

Open Problem: Construct compact G_2 and $Spin(7)$ manifolds with various types of isolated singularities

3.2. Non-compact Manifolds. As we explained, interesting physics occurs at the singular points of the special holonomy manifold X. Depending on the singularity, one may find, for example, extra gauge symmetry, global symmetry, or massless states localized at the singularity. For each type of the singularity, the corresponding physics may be different. However, usually it depends only on the vicinity of the singularity, and not on the rest of the geometry of the space X. Therefore, in order to study the physics associated with a given singularity, one can imagine isolating the local neighbourhood of the singular point and studying it separately. This gives the so-called 'local model' of a singular point. This procedure is similar to considering one gauge factor in the standard model gauge group, rather than studying the whole theory at once. In this sense, non-compact manifolds

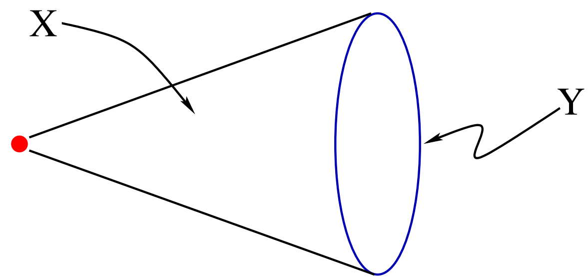


FIGURE 3. A cone over a compact space Y.

provide us with basic building blocks of low-energy energy physics that may appear in compactifications on compact manifolds.

Here, we discuss a particular class of isolated singularities, namely *conical singularities*. They correspond to degenerations of the metric on the space X of the form:

$$ds^2(X) = dt^2 + t^2 ds^2(Y), \tag{3.2}$$

where a compact space Y is the base of the cone; the dimension of Y is one less than the dimension of X. It is clear that X has an isolated singular point at the tip of the cone, except for the special case when Y is a sphere, $\mathbf{S}^{n-1}$, with a round metric.

The conical singularities of the form (3.2) are among the simplest isolated singularities one could study; see Figure3. In fact, the first examples of non-compact manifolds with G_2 and $Spin(7)$ holonomy, obtained by Bryant and Salamon [**15**] and, independently, by Gibbons, Page, and Pope [**28**], exhibit precisely this type of degeneration. Specifically, the complete metrics constructed in [**15, 28**] are smooth everywhere, and asymptotically look like (3.2), for various base manifolds Y. Therefore, they can be considered as resolutions of conical singularities. In Table 4 we list known asymptotically conical (AC) complete metrics with G_2 and $Spin(7)$ holonomy that were originally found in [**15, 28**] and in more recent literature [**17, 31**].

The method of constructing G_2 and $Spin(7)$ metrics originally used in [**15, 28**] was essentially based on the direct analysis of the Ricci-flatness equations,

$$R_{ij} = 0, \tag{3.3}$$

for a particular metric ansatz. We will not go into details of this approach here since it relies on finding the right form of the ansatz and, therefore, is not practical for generalizations. Instead, following [**32, 16**], we will describe a very powerful approach, recently developed by Hitchin [**34**], which allows us to construct all the G_2 and $Spin(7)$ manifolds listed in Table 4 (and many more!) in a systematic manner. Another advantage of this method is that it leads to first-order differential equations, which are much easier than the second-order Einstein equations (3.3).

Before we explain the basic idea of Hitchin's construction, notice that for all of the AC manifolds in Table 4 the base manifold Y is a homogeneous quotient space

$$Y = G/K, \tag{3.4}$$

Holonomy	Topology of X	Base Y
G_2	$\mathbf{S}^4 \times \mathbb{R}^3$	$\mathbb{C}\mathbf{P}^3$
	$\mathbb{C}\mathbf{P}^2 \times \mathbb{R}^3$	$SU(3)/U(1)^2$
	$\mathbf{S}^3 \times \mathbb{R}^4$	$SU(2) \times SU(2)$
	$T^{1,1} \times \mathbb{R}^2$	
$Spin(7)$	$\mathbf{S}^4 \times \mathbb{R}^4$	$SO(5)/SO(3)$
	$\mathbb{C}\mathbf{P}^2 \times \mathbb{R}^4$	$SU(3)/U(1)$
	$\mathbf{S}^5 \times \mathbb{R}^3$	

TABLE 4. Asymptotically conical manifolds with G_2 and $Spin(7)$ holonomy.

where G is some group and $K \subset G$ is a subgroup. Therefore, we can think of X as being foliated by *principal orbits* G/K over a positive real line, $\mathbb{R}_+$, as shown on Figure4. A real variable $t \in \mathbb{R}_+$ in this picture plays the role of the radial coordinate; the best way to see this is from the singular limit, in which the metric on X becomes exactly conical, *cf.* eq. (3.2).

As we move along $\mathbb{R}_+$, the size and the shape of the principal orbit changes, but only in a way consistent with the symmetries of the coset space G/K. In particular, at some point the principal orbit G/K may collapse into a degenerate orbit,

$$B = G/H \tag{3.5}$$

where symmetry requires

$$G \supset H \supset K \tag{3.6}$$

At this point (which we denote $t = t_0$) the "radial evolution" stops, resulting in a non-compact space X with a non-trivial topological cycle B, sometimes called a *bolt*. In other words, the space X is contractible to a compact set B, and from the relation (3.6) we can easily deduce that the normal space of B inside X is itself a cone on H/K. Therefore, in general, the space X obtained in this way is a singular space, with a conical singularity along the degenerate orbit $B = G/H$. However, if H/K is a round sphere, then the space X is smooth,

$$H/K = \mathbf{S}^k \quad \Longrightarrow \quad X \text{ smooth}$$

This simply follows from the fact that the normal space of B inside X in such a case is non-singular, $\mathbb{R}^{k+1}$ (= a cone over H/K). It is a good exercise to check that for all manifolds listed in Table 4, one indeed has $H/K = \mathbf{S}^k$, for some value of k. To show this, one should first write down the groups G, H, and K, and then find H/K.

The representation of the non-compact space X in terms of principal orbits which are homogeneous coset spaces is very useful. In fact, as we just explained, the

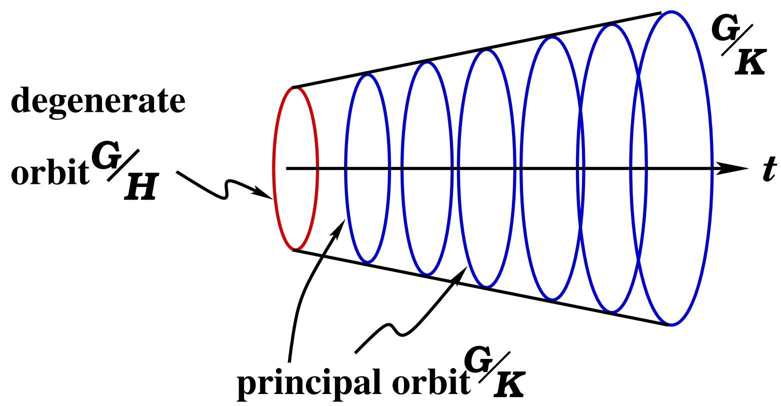


FIGURE 4. A non-compact space X can be viewed as a foliation by principal orbits $Y = G/K$. The non-trivial cycle in X correspond to the degenerate orbit G/H, where $G \supset H \supset K$.

topology of X simply follows from the group data (3.6). For example, if $H/K = \mathbf{S}^k$ so that X is smooth, we have

$$X \cong (G/H) \times \mathbb{R}^{k+1} \tag{3.7}$$

However, this structure can also be used to find a G-invariant metric on X. In order to do this, all we need to know are the groups G and K.

First, let us sketch the basic idea of Hitchin's construction [**34**], and then explain the details in some specific examples. For more details and further applications we refer the reader to [**32, 16**]. We start with a principal orbit $Y = G/K$ which can be, for instance, the base of the conical manifold that we want to construct. Let $\mathcal{P}$ be the space of (stable[4]) G-invariant differential forms on Y. This space is finite dimensional and, moreover, it turns out that there exists a symplectic structure on $\mathcal{P}$. This important result allows us to think of the space $\mathcal{P}$ as the phase space of some dynamical system:

$$\begin{aligned} \mathcal{P} = &\ \text{Phase Space} \\ \omega = &\sum dx_i \wedge dp_i \end{aligned} \tag{3.8}$$

where we parametrized $\mathcal{P}$ by some coordinate variables x_i and the conjugate momentum variables p_i.

Given a principal orbit G/K and a space of G-invariant forms on it, there is a canonical construction of a Hamiltonian $H(x_i, p_i)$ for our dynamical system, such that the Hamiltonian flow equations are equivalent to the special holonomy

[4]Stable forms are defined as follows [**35**]. Let X be a manifold of real dimension n, and $V = TX$. Then, the form $\rho \in \Lambda^p V^*$ is stable if it lies in an open orbit of the (natural) $GL(V)$ action on $\Lambda^p V^*$. In other words, this means that all forms in the neighborhood of ρ are $GL(V)$-equivalent to ρ. This definition is useful because it allows one to define a volume. For example, a symplectic form ω is stable if and only if $\omega^{n/2} \neq 0$.

condition [**34**]:

$$\left\{\begin{array}{lcl} \frac{dx_i}{dt} & = & \frac{\partial H}{\partial p_i} \\ \frac{dp_i}{dt} & = & -\frac{\partial H}{\partial x_i} \end{array}\right. \iff \begin{array}{c}\text{Special Holonomy Metric} \\ \text{on } (t_1,t_2)\times(G/K)\end{array} \tag{3.9}$$

where the ‘time’ in the Hamiltonian system is identified with the radial variable t. Thus, solving the Hamiltonian flow equations from $t = t_1$ to $t = t_2$ with a particular boundary condition leads to the special holonomy metric on $(t_1, t_2) \times (G/K)$. Typically, one can extend the boundaries of the interval (t_1, t_2) where the solution is defined to infinity on one side, and to a point $t = t_0$, where the principal orbit degenerates, on the other side. Then, this gives a complete metric with special holonomy on a non-compact manifold X of the form (3.7). Let us now illustrate these general ideas in more detail in a concrete example.

Example 3.2. Let us take $G = SU(2)^3$ and $K = SU(2)$. We can form the following natural sequence of subgroups:

$$\begin{array}{ccccc} G & & H & & K \\ \| & & \| & & \| \\ SU(2)^3 & \supset & SU(2)^2 & \supset & SU(2) \end{array} \tag{3.10}$$

From the general formula (3.4) it follows that in this example we deal with a space X, whose principal orbits are

$$Y = SU(2)\times SU(2) \cong \mathbf{S}^3\times\mathbf{S}^3$$

Furthermore, $G/H \cong H/K \cong \mathbf{S}^3$ implies that X is a smooth manifold with topology, *cf.* (3.7),

$$X \cong \mathbf{S}^3\times\mathbb{R}^4$$

In fact, X is one of the asymptotically conical manifolds listed in Table 4.

In order to find a G_2 metric on this manifold, we need to construct the “phase space”, $\mathcal{P}$, that is, the space of $SU(2)^3$–invariant 3-forms and 4-forms on $Y = G/K$:

$$\mathcal{P} = \Omega^3_G(G/K)\times\Omega^4_G(G/K)$$

In this example, it turns out that each of the factors is one-dimensional, generated by a 3-form ρ and by a 4-form σ, respectively,

$$\rho = \sigma_1\sigma_2\sigma_3 - \Sigma_1\Sigma_2\Sigma_3 + x\Big(d(\sigma_1\Sigma_1) + d(\sigma_2\Sigma_2) + d(\sigma_3\Sigma_3)\Big), \tag{3.11}$$

$$\sigma = p^{2/5}\Big(\sigma_2\Sigma_2\sigma_3\Sigma_3 + \sigma_3\Sigma_3\sigma_1\Sigma_1 + \sigma_1\Sigma_1\sigma_2\Sigma_2\Big). \tag{3.12}$$

where we introduced two sets of left invariant one-forms (σ_a, Σ_a) on Y

$$\begin{aligned} \sigma_1 &= \cos\psi d\theta + \sin\psi\sin\theta d\phi, & \Sigma_1 &= \cos\tilde\psi d\tilde\theta + \sin\tilde\psi\sin\tilde\theta d\tilde\phi, \\ \sigma_2 &= -\sin\psi d\theta + \cos\psi\sin\theta d\phi, & \Sigma_2 &= -\sin\tilde\psi d\tilde\theta + \cos\tilde\psi\sin\tilde\theta d\tilde\phi, \\ \sigma_3 &= d\psi + \cos\theta d\phi, & \Sigma_3 &= d\tilde\psi + \cos\tilde\theta d\tilde\phi \end{aligned} \tag{3.13}$$

which satisfy the usual $SU(2)$ algebra

$$d\sigma_a = -\frac{1}{2}\epsilon_{abc}\sigma_a\wedge\sigma_b,\ \ d\Sigma_a = -\frac{1}{2}\epsilon_{abc}\Sigma_a\wedge\Sigma_b. \tag{3.14}$$

Therefore, we have only one “coordinate” x and its conjugate “momentum” p, parametrizing the “phase space” $\mathcal{P}$ of our model. In order to see that there is a natural symplectic structure on $\mathcal{P}$, note that x and p multiply exact forms. For x

this obviously follows from (3.11) and for p this can be easily checked using (3.12) and (3.14). This observation can be used to define a non-degenerate symplectic structure on $\mathcal{P} = \Omega^3_{exact}(Y) \times \Omega^4_{exact}(Y)$. Explicitly, it can be written as

$$\omega\left((\rho_1,\sigma_1),(\rho_2,\sigma_2)\right) = \langle \rho_1, \sigma_2 \rangle - \langle \rho_2, \sigma_1 \rangle,$$

where, in general, for $\rho = d\beta \in \Omega^k_{exact}(Y)$ and $\sigma = d\gamma \in \Omega^{n-k}_{exact}(Y)$ one has a nondegenerate pairing

$$\langle \rho, \sigma \rangle = \int_Y d\beta \wedge \gamma = (-1)^k \int_Y \beta \wedge d\gamma. \tag{3.15}$$

Once we have the phase space $\mathcal{P}$, it remains to write down the Hamiltonian flow equations (3.9). In Hitchin's construction, the Hamiltonian $H(x,p)$ is defined as an invariant functional on the space of differential forms, $\mathcal{P}$. Specifically, in the context of G_2 manifolds it is given by

$$H = 2V(\sigma) - V(\rho), \tag{3.16}$$

where $V(\rho)$ and $V(\sigma)$ are suitably defined volume functionals[5] [**34**]. Computing (3.16) for the G-invariant forms (3.11) and (3.12) we obtain the Hamiltonian flow equations:

$$\begin{cases} \dot{p} &= x(x-1)^2 \\ \dot{x} &= p^2 \end{cases}$$

These first-order equations can be easily solved, and the solution for $x(t)$ and $p(t)$ determines the evolution of the forms ρ and σ, respectively. On the other hand, these forms define the associative three-form on the 7-manifold $Y \times (t_1, t_2)$,

$$\Phi = dt \wedge \omega + \rho, \tag{3.19}$$

where ω is a 2-form on Y, such that $\sigma = \omega^2/2$. For $x(t)$ and $p(t)$ satisfying the Hamiltonian flow equations the associative form Φ is automatically closed and co-closed, *cf.* (2.6). Therefore, as we explained in section 2, it defines a G_2 holonomy metric. Specifically, one can use (2.10) to find the explicit form of the metric, which after a simple change of variables becomes the G_2 metric on the spin bundle over $\mathbf{S}^3$, originally found in [**15, 28**]:

$$ds^2 = \frac{dr^2}{1 - r_0^2/r^2} + \frac{r^2}{12}\sum_{a=1}^{3}(\sigma_a - \Sigma_a)^2 + \frac{r^2}{36}\left(1 - \frac{r_0^2}{r^2}\right)\sum_{a=1}^{3}(\sigma_a + \Sigma_a)^2 \tag{3.20}$$

The above example can be easily generalised in a number of directions. For example, if instead of (3.10) we take $G = SU(2)^2$ and K to be its trivial subgroup, we end up with the same topology for X and Y, but with a larger space of G_2

---

[5]The volume $V(\sigma)$ for a 4-form $\sigma \in \Lambda^4 T^*Y \cong \Lambda^2 TY \otimes \Lambda^6 T^*Y$ is very easy to define. Indeed, we have $\sigma^3 \in \Lambda^6 TY \otimes (\Lambda^6 T^*Y)^3 \cong (\Lambda^6 T^*Y)^2$, and therefore we can take

$$V(\sigma) = \int_Y |\sigma^3|^{\frac{1}{2}}. \tag{3.17}$$

to be the volume for σ. In order to define the volume $V(\rho)$ for a 3-form $\rho \in \Lambda^3 T^*Y$, one first defines a map $K_\rho \colon TY \to TY \otimes \Lambda^6 T^*Y$, such that for a vector $v \in TY$ it gives $K(v) = \imath(v)\rho \wedge \rho \in \Lambda^5 T^*Y \cong TY \otimes \Lambda^6 T^*Y$. Hence, one can define $tr(K^2) \in (\Lambda^6 T^*Y)^2$. Since stable forms with stabilizer $SL(3,\mathbb{C})$ are characterised by $tr(K)^2 < 0$, following [**35**], we define

$$V(\rho) = \int_Y |\sqrt{-trK^2}|. \tag{3.18}$$

metrics on X. Indeed, for $G = SU(2)^2$ the space of G-invariant forms on Y is much larger. Therefore, the corresponding dynamical system is more complicated and has a richer structure. Some specific solutions of this more general system have been recently constructed in [**11, 18, 13, 19**], but the complete solution is still not known.

There is a similar systematic method, also developed by Hitchin [**34**], of constructing complete non-compact manifolds with $Spin(7)$ holonomy. Again, this method can be used to obtain the asymptotically conical metrics listed in Table 4, as well as other $Spin(7)$ metrics recently found in [**20, 29, 17, 42**].

4. M-theory, Circle Fibrations, and Calibrated Submanifolds

Since compactification on smooth manifolds does not produce interesting physics — in particular, does not lead to realistic quantum field theories — one has to study the dynamics of string theory and M-theory on *singular* G_2 and $Spin(7)$ manifolds. This is a very interesting problem which can provide us with many insights about infrared behaviour of minimally supersymmetric gauge theories and even about M-theory itself. From the experience with string theory and M-theory, one might expect to find some new physics at the singularities of G_2 and $Spin(7)$ manifolds, for example,

- New massless objects
- Extra gauge symmetry
- Restoration of continuous/discrete symmetry
- Topology changing transition

$\vdots$

Before one talks about physics associated with G_2 and $Spin(7)$ singularities, it would be nice to have a classification of all such degenerations. Unfortunately, this problem is not completely solved even for Calabi-Yau manifolds (apart from the ones in low dimension), and seems even less promising for real manifolds with exceptional holonomy. Therefore, one starts with some simple examples.

One simple kind of singularities — which we already encountered in section 3.1 in the Joyce construction of compact manifolds with exceptional holonomy — is a large class of orbifold singularities[6]. Locally, an orbifold singularity can be represented as a quotient of $\mathbb{R}^n$ by some discrete group Γ,

$$\mathbb{R}^n/\Gamma \tag{4.1}$$

Therefore, in string theory, the physics associated with such singularities can be systematically extracted from the orbifold conformal field theory [**21**]. For G_2 manifolds this was done in [**51, 23, 52, 9, 24, 49**], and for $Spin(7)$ manifolds in [**51, 10**]. Typically, one finds new massless degrees of freedom localized at the orbifold singularity and other phenomena listed above. However, the CFT technique is not applicable for studying M-theory on singular G_2 and $Spin(7)$ manifolds. Moreover, as we mentioned in section 3.1, many interesting phenomena occur at singularities which are not of the orbifold type, and to study the physics of those we need some new methods.

[6]A large list of G_2 orbifold singularities can be found in [**33**].

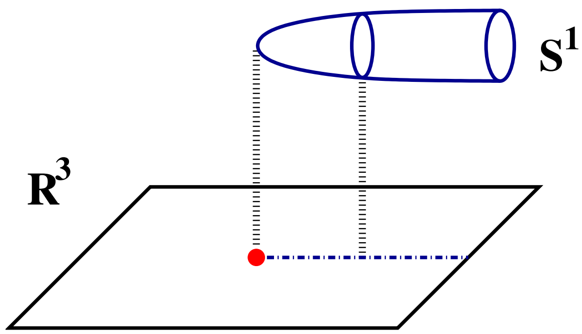


FIGURE 5. A cartoon representing Taub-NUT space as a circle fibration over a 3-plane.

4.1. Low Energy Dynamics via IIA Duals. One particularly useful method of analyzing M-theory on singular manifolds with special holonomy follows from the duality between type IIA string theory and M-theory compactified on a circle [**53, 55**]:

$$\boxed{\text{IIA Theory}} \quad \Longleftrightarrow \quad \boxed{\text{M-theory on } \mathbf{S}^1}$$

Among other things, this duality implies that any state in IIA theory can be identified with the corresponding state in M-theory. In this identification, some of the states acquire geometric origin when lifted to eleven dimensions. In order to see this explicitly, let us write the eleven-dimensional metric in the form

$$ds^2 = e^{-\frac{2}{3}\phi} g_{\mu\nu} dx^\mu dx^\nu + e^{\frac{4}{3}\phi}(dx_{11} + A_\mu dx^\mu)^2 \tag{4.2}$$

Upon reduction to ten-dimensionsional space-time (locally parametrized by x_μ), the field ϕ is identified with the dilaton, A_μ with the Ramond-Ramond 1-form, and $g_{\mu\nu}$ with the ten-dimensional metric. Therefore, any IIA background that involves excitations of these fields uplifts to a purely geometric background in eleven dimensions. Moreover, from the explicit form of the metric (4.2) it follows that the eleven-dimensional geometry is a circle fibration over the ten-dimensional space-time, such that the topology of this fibration is determined by the configuration of Ramond-Ramond 1-form field. This important observation will play a central role in this section.

To be specific, let us consider a D6-brane in type IIA string theory. Since a D6-brane is a source for the dilaton, for the Ramond-Ramond 1-form, and for the metric, it is precisely the kind of state that uplifts to pure geometry. For example, if both the D6-brane world-volume and the ambient space-time are flat, the dual background in M-theory is given by the Taub-NUT space:

$$\boxed{\begin{array}{l}\text{M-theory on}\\ \text{Taub-Nut space}\end{array}} \quad \Longleftrightarrow \quad \boxed{\begin{array}{l}\text{IIA theory in flat space-time}\\ \text{with a D6-brane}\end{array}}$$

The Taub-NUT space is a non-compact four-manifold with $SU(2)$ holonomy. It can be viewed as a circle fibration over a 3-plane; see Figure5. The $\mathbf{S}^1$ fiber degenerates at a single point — the origin of the $\mathbb{R}^3$ — which is identified with the location of the D6-brane in type IIA string theory. On the other hand, at large distance the size of the 'M-theory circle' stabilizes at some constant value (related to the value of the string coupling constant in IIA theory). Explicitly, the metric on the Taub-NUT space is given by [**22**]:

$$\begin{aligned} ds^2_{TN} &= H d\vec{r}^{\,2} + H^{-1}(dx_{11} + A_\mu dx^\mu)^2, \\ \vec{\nabla} \times \vec{A} &= -\vec{\nabla} H, \\ H &= 1 + \frac{1}{2r} \end{aligned} \tag{4.3}$$

This form of the metric makes especially clear the structure of the circle fibration. Indeed, if we fix a constant-r sphere inside the $\mathbb{R}^3$, then it is easy to see that the $\mathbf{S}^1$ fiber has 'winding number one' over this sphere. This indicates that there is a topological defect — namely, a D6-brane — located at $r = 0$, where the $\mathbf{S}^1$ fiber degenerates.

The relation between D6-branes and geometry can be extended to more general manifolds that admit a smooth $U(1)$ action. Indeed, if X is a space (not necessarily smooth and compact) with a $U(1)$ isometry, such that $X/U(1)$ is smooth, then the fixed point set, L, of the $U(1)$ action must be of codimension 4 inside X [**47, 43**]. This is just the right codimension to identify L with the D6-brane locus[7] in type IIA theory on $X/U(1)$:

$$\boxed{\text{M-theory on } X} \quad \Longleftrightarrow \quad \boxed{\begin{array}{l}\text{IIA theory on } X/U(1) \\ \text{with D6-branes on } L\end{array}}$$

For example, if X is the Taub-NUT space, then the $U(1)$ action is generated by the shift of the periodic variable x_{11} in (4.3). Dividing by this action one finds $X/U(1) \cong \mathbb{R}^3$ and that $L = \{\mathrm{pt}\} \in X$ indeed has codimension 4.

It may happen that a space X admits more than one $U(1)$ action. In that case, M-theory on X will have several IIA duals, which may look very different but, of course, should exhibit the same physics. This idea was used by Atiyah, Maldacena, and Vafa to realise geometric duality between certain IIA backgrounds as a topology changing transition in M-theory [**6**]. We will come back to this in section 5.

Now let us describe a particularly useful version of the duality between M-theory on a non-compact space X and IIA theory with D6-branes on $X/U(1)$ which occurs when X admits a $U(1)$ action, such that the quotient space is isomorphic to a flat Euclidean space. Suppose, for example, that X is a non-compact G_2 manifold, such that

$$X/U(1) \cong \mathbb{R}^6 \tag{4.4}$$

Then, the duality statement reads:

$$\boxed{\begin{array}{l}\text{M-theory on non-compact} \\ G_2 \text{ manifold } X\end{array}} \quad \Longleftrightarrow \quad \boxed{\begin{array}{l}\text{IIA theory in flat space-} \\ \text{time with D6-branes on} \\ \mathbb{R}^4 \times L\end{array}}$$

On the left-hand side of this equivalence the space-time is $\mathbb{R}^4 \times X$. On the other hand, the geometry of space-time in IIA theory is trivial (at least topologically) and all the interesting information about X is mapped into the geometry of the

[7]The part of the D-brane world-volume that is transverse to X is flat and does not play an important rôle in our discussion here.

D6-brane locus L. For example, the Betti numbers of L are determined by the corresponding Betti numbers of the space X [**29**]:

$$\begin{aligned} b_k(L) &= b_{k+2}(X), \quad k>0 \\ b_0(L) &= b_2(X)+1 \end{aligned} \tag{4.5}$$

Notice the shift in degree by 2, and also that the number of the D6-branes (= the number of connected components of L) is determined by the second Betti number of X. We will not present the derivation of this formula here. However, it is a useful exercise to check (*e.g.* using the Lefschetz fixed point theorem) that the Euler numbers of X and L must be equal, in agreement with (4.5).

The duality between M-theory on a non-compact manifold X and a configuration of D6-branes in a (topologically) flat space can be used to study singular limits of X. Indeed, when X develops a singularity, so does L. Moreover, L must be a supersymmetric (special Lagrangian) submanifold in $\mathbb{R}^6 \cong \mathbb{C}^3$ in order to preserve the same amount of supersymmetry[8] as the G_2 space X. Therefore, the problem of studying the dynamics of M-theory on G_2 singularities can be restated as a problem of studying D6-brane configurations on singular special Lagrangian submanifolds in flat space [**8**].

Following Atiyah and Witten [**8**], let us see how this duality can help us to analyze one of the conical G_2 singularities listed in Table 4.

EXAMPLE 4.1. Consider an asymptotically conical G_2 manifold X with $SU(3)$ / $U(1)^2$ principal orbits and topology

$$X \cong \mathbb{C}\mathbf{P}^2 \times \mathbb{R}^3 \tag{4.6}$$

Assuming that M-theory on this space X admits a circle reduction to IIA theory with D6-branes in flat space, we can apply the general formula (4.5) to find the topology of the D6-brane locus L. For the manifold (4.6) we find the following non-zero Betti numbers:

Betti numbers of X		Betti numbers of L
$b_4 = 1$	$\longrightarrow$	$b_2 = 1$
$b_2 = 1$	$\longrightarrow$	$b_0 = 2$

Therefore, we conclude that L should be a non-compact 3-manifold with two connected components (since $b_0 = 2$) and with a single topologically non-trivial 2-cycle (since $b_2 = 1$). A simple guess for a manifold that has these properties is

$$L \cong \big(\mathbf{S}^2 \times \mathbb{R}\big) \cup \mathbb{R}^3 \tag{4.7}$$

It turns out that there indeed exists a special Lagrangian submanifold in $\mathbb{C}^3$ with the right symmetries and topology (4.7); see [**8**] for an explicit construction of the circle action on X that has L as a fixed point set. If we choose to parametrize $\mathrm{Re}(\mathbb{C}^3) = \mathbb{R}^3$ and $\mathrm{Im}(\mathbb{C}^3) = \mathbb{R}^3$ by 3-vectors $\vec{\phi}_1$ and $\vec{\phi}_2$, respectively, then one can explicitly describe L as the zero locus of the polynomial relations [**38**]:

$$L = \Big\{ \vec{\phi}_1 \cdot \vec{\phi}_2 = -|\vec{\phi}_1||\vec{\phi}_2| \ ; \quad |\vec{\phi}_1|(3|\vec{\phi}_2|^2 - |\vec{\phi}_1|^2) = \rho \Big\} \cup \Big\{ |\vec{\phi}_1 - \sqrt{3}\vec{\phi}_2| = 0 \Big\} \tag{4.8}$$

[8] In general, in a reduction from M-theory down to Type IIA one does not obtain the standard flat metric on $X/U(1) \cong \mathbb{R}^6$ due to non-constant dilaton and other fields in the background. However, one would expect that near the singularities of the D-brane locus L these fields exhibit a regular behavior, and the metric on $X/U(1)$ is approximately flat, *cf.* [**8**]. In this case the condition for the Type IIA background to be supersymmetric can be expressed as a simple geometric criterion: it says that the D-brane locus L should be a calibrated submanifold in $X/U(1)$.

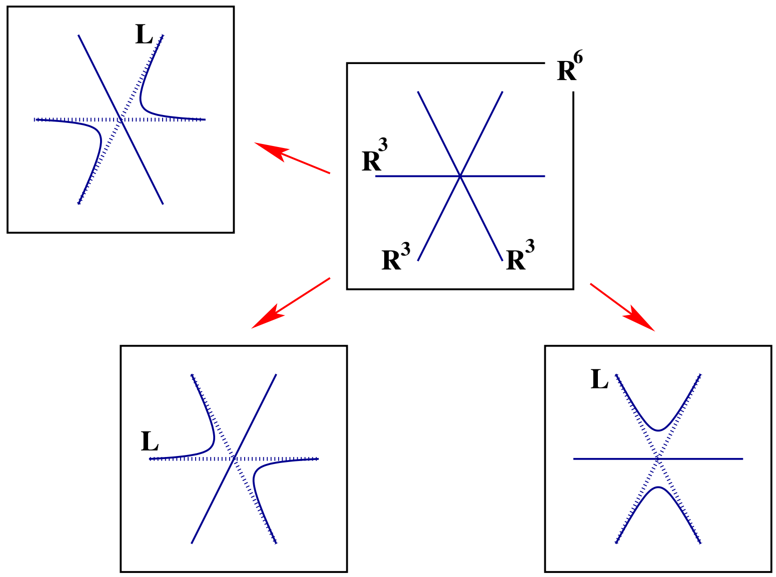


Figure 6. Intersection of special Lagrangian D6-branes dual to a G_2 cone over the flag manifold $SU(3)/U(1)^2$, and its three non-singular deformations.

It is easy to see that the first component of this manifold looks like a hyperboloid in $\mathbb{C}^3 = \mathbb{R}^3 \times \mathbb{R}^3$. It has a 'hole' in the middle, resulting in the $\mathbf{S}^2 \times \mathbb{R}$ topology, required by (4.7). The second component in (4.8) is clearly a 3-plane, which goes through this hole, as shown in Figure 6.

When X develops a conical singularity L degenerates into a collection of three planes,

$$L_{\mathrm{sing}} \cong \mathbb{R}^3 \cup \mathbb{R}^3 \cup \mathbb{R}^3 \tag{4.9}$$

intersecting at a single point, see Figure6. This can be seen explicitly by taking the $\rho \to 0$ limit in the geometry (4.8),

$$\begin{aligned} D6_1 : &\qquad \vec{\phi}_1 = 0 \\ D6_2 : &\qquad \frac{1}{2}\vec{\phi}_1 + \frac{\sqrt{3}}{2}\vec{\phi}_2 = 0 \\ D6_3 : &\qquad -\frac{1}{2}\vec{\phi}_1 + \frac{\sqrt{3}}{2}\vec{\phi}_2 = 0 \end{aligned}$$

which corresponds to the limit of a collapsing $\mathbf{S}^2$ cycle.

Therefore, in this example the physics of M-theory on a conical G_2 singularity can be reduced to the physics of three intersecting D6-branes in type IIA string theory. In particular, since D6-branes appear symmetrically in this dual picture, one can conclude that there must be three ways of resolving this singularity, depending on which pair of D6-branes is deformed into a smooth special Lagrangian submanifold (4.8). This is precisely what Atiyah and Witten found in a more careful analysis [**8**]. We will come back to this example later, in section 5.

There is a similar duality that relates M-theory on $Spin(7)$ manifolds to D6-brane configurations on coassociative cycles [**25**]. In particular, if X is a non-compact $Spin(7)$ manifold with a $U(1)$ action, such that

$$X/U(1) \cong \mathbb{R}^7 \tag{4.10}$$

then one finds the following useful duality:

M-theory on non-compact $Spin(7)$ manifold X	$\Longleftrightarrow$	IIA theory in flat space-time with D6-branes on $\mathbb{R}^3 \times L$

where L is a coassociative submanifold in $\mathbb{R}^7$. Again, on the left-hand side the geometry of space-time is $\mathbb{R}^3 \times X$, whereas on the right-hand side space-time is (topologically) flat and all the interesting information is encoded in the geometry of the D6-brane locus L. The topology of the latter can be determined from the general formula (4.5). Therefore, as in the G_2 case, M-theory dynamics on singular $Spin(7)$ manifolds can be obtained by investigating D6-brane configurations on singular coassociative submanifolds in flat space [**29**].

5. Topology Change in M-theory on Exceptional Holonomy Manifolds

5.1. Topology Change in M-theory. In this section we discuss topology changing transitions, by which we mean a particular behavior of the manifold X (and the associated physics) in the singular limit when one can go to a space with a different topology. In Calabi-Yau manifolds many examples of such transitions are known and can be understood using conformal field theory methods, see *e.g.* [**4**] and references therein. Some of these transitions give rise to analogous topology changing transitions in G_2 and $Spin(7)$ manifolds obtained from finite quotients of Calabi-Yau spaces that we discussed in section 3.1. In the context of compact manifolds with G_2 holonomy this was discussed in [**37, 48**]. One typically finds a transition between different topologies of X, such that the following sum of the Betti numbers remains invariant:

$$b_2 + b_3 = \text{const} \tag{5.1}$$

which is precisely what one would expect from the conformal field theory considerations [**51**]. One can interpret the condition (5.1) as a feature of the mirror phenomenon for G_2 manifolds [**2, 44, 32**].

Here, we shall discuss topology change in non-compact asymptotically conical G_2 and $Spin(7)$ manifolds of section 3.2, which are in a sense the most basic examples of singularities that reveal the peculiar aspects of exceptional holonomy. Notice that all of these manifolds have the form

$$X \cong B \times (\text{contractible})$$

where B is a non-trivial cycle (a bolt), *e.g.* $B = \mathbf{S}^3$, $\mathbf{S}^4$, $\mathbb{C}\mathbf{P}^2$, or something else. Therefore, it is natural to ask:

$$\text{“}\textit{What happens if } \mathrm{Vol}(B) \to 0 \text{ ?”}$$

In this limit the geometry becomes singular and, as we discussed earlier, there are many possibilities for M-theory dynamics associated with it. One possibility is a topology change, which is indeed what we shall find in some of the cases below.

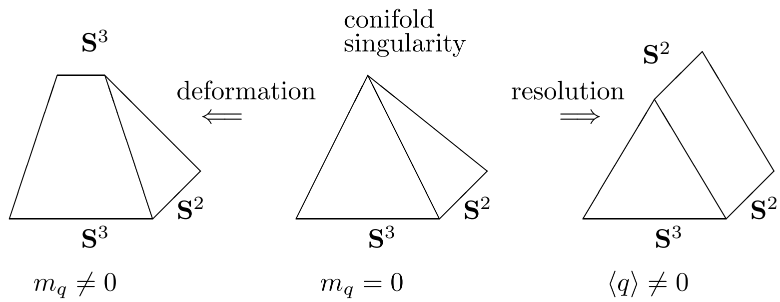


FIGURE 7. Conifold transition in type IIB string theory.

Although we are mostly interested in exceptional holonomy manifolds, it is instructive to start with topology changing transitions in Calabi-Yau manifolds, where one finds two prototypical examples:

The Flop is a transition between two geometries, where one two-cycle shrinks to a point and a (topologically) different two-cycle grows. This process can be schematically described by the diagram:

$$\mathbf{S}^2_{(1)} \longrightarrow \cdot \longrightarrow \mathbf{S}^2_{(2)}$$

This transition is smooth in string theory [**5, 56**].

The Conifold transition is another type of topology change, in which a three-cycle shrinks and is replaced by a two-cycle:

$$\mathbf{S}^3 \longrightarrow \cdot \longrightarrow \mathbf{S}^2$$

Unlike the flop, it is a real phase transition in the low-energy dynamics which can be understood as the condensation of massless black holes [**50, 26**]. Let us briefly recall the main arguments.

As the name indicates, the conifold is a cone over a five-dimensional base space which has topology $\mathbf{S}^2 \times \mathbf{S}^3$ (see Figure 7). Two different ways to desingularize this space — called the deformation and the resolution — correspond to replacing the singularity by a finite size $\mathbf{S}^3$ or $\mathbf{S}^2$, respectively. Thus, we have two different spaces, with topology $\mathbf{S}^3 \times \mathbb{R}^3$ and $\mathbf{S}^2 \times \mathbb{R}^4$, which asymptotically look the same.

In type IIB string theory, the two phases of the conifold geometry correspond to different branches in the four-dimensional $\mathcal{N} = 2$ low-energy effective field theory. In the deformed conifold phase, D3-branes wrapped around the 3-sphere give rise to a low-energy field q, with mass determined by the size of the $\mathbf{S}^3$. In the effective four-dimensional supergravity theory these states appear as heavy, point-like, extremal black holes. On the other hand, in the resolved conifold phase the field q acquires an expectation value reflecting the condensation of these black holes. Of course, in order to make the transition from one phase to the other, the field q must become massless somewhere and this happens at the conifold singularity, as illustrated in Figure7.

Now, let us proceed to topology change in G_2 manifolds. Here, again, one finds two kinds of topology changing transitions, which resemble the flop and the conifold transitions in Calabi-Yau manifolds:

G_2 **Flop** is a transition where a 3-cycle collapses and gets replaced by a (topologically) different 3-cycle:

$$\mathbf{S}^3_{(1)} \longrightarrow \cdot \longrightarrow \mathbf{S}^3_{(2)}$$

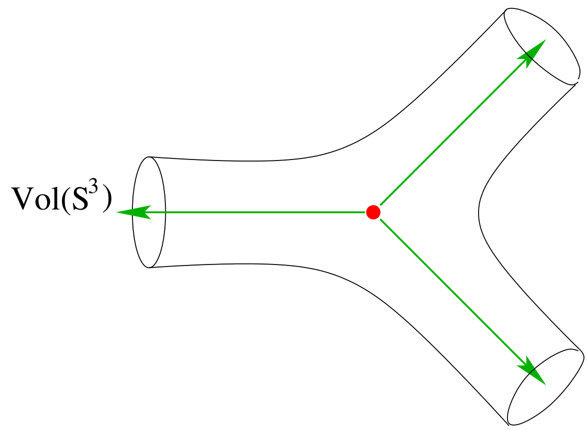


FIGURE 8. Quantum moduli space of M-theory on the G_2 manifold X with topology $\mathbf{S}^3 \times \mathbb{R}^4$. Green lines represent the 'geometric moduli space' parametrized by the volume of the $\mathbf{S}^3$ cycle, which is enlarged to a smooth complex curve by taking into account the C-field and quantum effects. The resulting moduli space has three classical limits, which can be connected without passing through the point where the geometry becomes singular (represented by the dot in this picture).

Note that this is indeed very similar to the flop transition in Calabi-Yau manifolds, where instead of a 2-cycle we have a 3-cycle shrinking. The physics is also similar, with membranes playing the role of string world-sheet instantons. Remember that the latter were crucial for a flop transition to be smooth in string theory. For a very similar reason, the G_2 flop transition is smooth in M-theory. This was first realised by Acharya [**1**], and by Atiyah, Maldacena, and Vafa [**6**], for a 7-manifold with topology

$$X \cong \mathbf{S}^3 \times \mathbb{R}^4$$

and studied further by Atiyah and Witten [**8**]. In particular, they found that M-theory on X has three classical branches, related by triality permutation symmetry, so that the quantum moduli space looks as shown on Figure8. Once again, the important point is that there is no singularity in quantum theory.

Let us proceed to another kind of topology changing transition in manifolds with G_2 holonomy.

A Phase Transition, somewhat similar to a hybrid of the conifold and the G_2 flop transition, can be found in M-theory on a G_2 manifold with topology

$$X \cong \mathbb{C}\mathbf{P}^2 \times \mathbb{R}^3$$

A singularity develops when the $\mathbb{C}\mathbf{P}^2$ cycle shrinks. As in the conifold transition, the physics of M-theory on this space also becomes singular at this point. Hence, this is a genuine phase transition [**8**]. Note, however, that unlike the conifold transition in type IIB string theory, this phase transition is not associated with

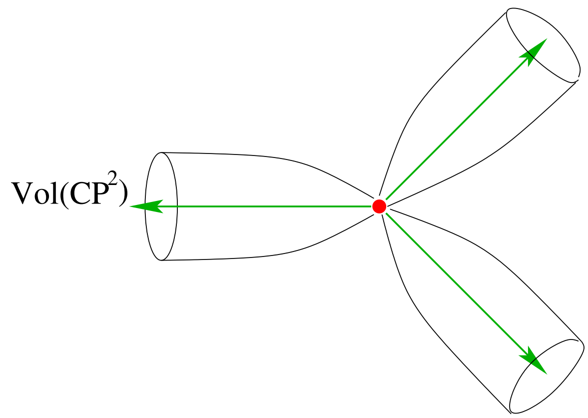


FIGURE 9. Quantum moduli space of M-theory on the G_2 manifold X with topology $\mathbb{C}\mathbf{P}^2 \times \mathbb{R}^3$. Green lines represent the 'geometric moduli space' parametrized by the volume of the $\mathbb{C}\mathbf{P}^2$ cycle, which is enlarged to a *singular* complex curve by taking into account the C-field and quantum effects. The resulting moduli space has three classical limits. In order to go from one branch to another one necessarily has to pass through the point where the geometry becomes singular (represented by the dot in this picture).

condensation of any particle-like states in M-theory[9] on X. Indeed, there are no 4-branes in M-theory, which could result in particle-like objects by wrapping around the collapsing $\mathbb{C}\mathbf{P}^2$ cycle.

Like the G_2 flop transition, this phase transition has three classical branches, which are related by triality symmetry, see Figure9. The important difference, of course, is that now one can go from one branch to another only through the singular point. In this transition one $\mathbb{C}\mathbf{P}^2$ cycle shrinks and another (topologically different) $\mathbb{C}\mathbf{P}^2$ cycle grows:

$$\mathbb{C}\mathbf{P}^2_{(1)} \longrightarrow \cdot \longrightarrow \mathbb{C}\mathbf{P}^2_{(2)}$$

One way to see that this is indeed the right physics of M-theory on X is to reduce it to type IIA theory with D6-branes in flat space-time [**8**].

Finally, we come to the last and the hardest case of holonomy groups, namely to $Spin(7)$ holonomy.

The $Spin(7)$ **Conifold** is the cone on $SU(3)/U(1)$. It was conjectured in [**31**] that the effective dynamics of M-theory on the $Spin(7)$ conifold is analogous to that of type IIB string theory on the usual conifold. Namely, the $Spin(7)$ cone on $SU(3)/U(1)$ has two different desingularizations, obtained by replacing the conical singularity either with a 5-sphere or with a $\mathbb{C}\mathbf{P}^2$; see Figure10. As a result, we obtain two different $Spin(7)$ manifolds, with topology

$$\mathbb{C}\mathbf{P}^2 \times \mathbb{R}^4 \quad \text{and} \quad \mathbf{S}^5 \times \mathbb{R}^3$$

which are connected via the topology changing transition

$$\mathbb{C}\mathbf{P}^2 \longrightarrow \cdot \longrightarrow \mathbf{S}^5$$

[9]However, such an interpretation can be given in type IIA string theory [**31**].

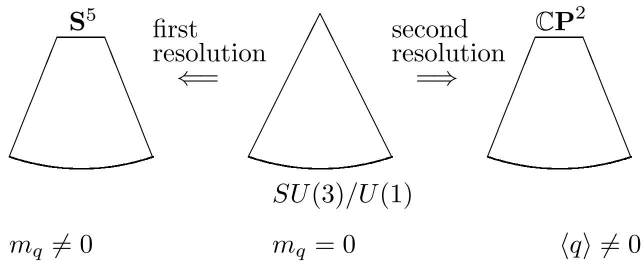


FIGURE 10. Conifold transition in M-theory on a manifold with $Spin(7)$ holonomy.

As with the conifold transition [**50, 26**], the $Spin(7)$ conifold has a nice interpretation in terms of the condensation of M5-branes. Namely, in the $\mathbf{S}^5$ phase we have extra massive states obtained upon quantization of the M5-brane wrapped around the five-sphere. The mass of these states is related to the volume of the $\mathbf{S}^5$. At the conifold point where the five-sphere shrinks, these M5-branes become massless as suggested by the classical geometry. At this point, the theory may pass through a phase transition into the Higgs phase, associated with the condensation of these five-brane states; see Figure10.

To continue the analogy with the Calabi-Yau conifold, recall that the moduli space of type II string theory on the Calabi-Yau conifold has three semi-classical regimes. The deformed conifold provides one of these, while there are two large-volume limits of the resolved conifold, related to each other by a flop transition. In fact, the same picture emerges for the $Spin(7)$ conifold. In this case, however, the two backgrounds differ not in geometry, but in the G-flux. It was shown in [**29**] that, due to the membrane anomaly of [**57**], M-theory on $X \cong \mathbb{C}\mathbf{P}^2 \times \mathbb{R}^4$ is consistent only for half-integral units of G_4 through the $\mathbb{C}\mathbf{P}^2$ bolt. Namely, after the transition from $X \cong \mathbf{S}^5 \times \mathbb{R}^3$, the G-flux may take the values $\pm 1/2$, with the two possibilities related by a parity transformation [**31**]. Thus, the moduli space of M-theory on the $Spin(7)$ cone over $SU(3)/U(1)$ also has three semi-classical limits: one with the parity invariant background geometry $\mathbf{S}^5 \times \mathbb{R}^3$, and two with the background geometry $X \cong \mathbb{C}\mathbf{P}^2 \times \mathbb{R}^4$ where parity is spontaneously broken; see Figure11. The last two limits are mapped into each other under parity transformation.

This picture is reproduced in the effective low-energy theory if we include in the spectrum light states corresponding to M5-branes wrapped over the five-sphere:

Effective Theory: $\mathcal{N} = 1$, $D = 3$ Maxwell-Chern-Simons theory with one charged complex scalar multiplet q

Here, it is the Higgs field q that appears due to quantizaion of the M5-branes. In this theory, different topological phases correspond to the Coulomb and Higgs branches:

$$\mathbf{S}^5 \times \mathbb{R}^3 \quad \Longleftrightarrow \quad \text{Coulomb branch} \tag{5.2}$$

$$\mathbb{C}\mathbf{P}^2 \times \mathbb{R}^4 \quad \Longleftrightarrow \quad \text{Higgs branch} \tag{5.3}$$

Further agreement in favor of this identification arises from examining the various extended objects that exist in M-theory on $\mathbb{C}\mathbf{P}^2 \times \mathbb{R}^4$, obtained from wrapped M5 or M2-branes. For example, we can consider an M2-brane over $\mathbb{C}\mathbf{P}^1$ inside $\mathbb{C}\mathbf{P}^2 \times \mathbb{R}^4$.

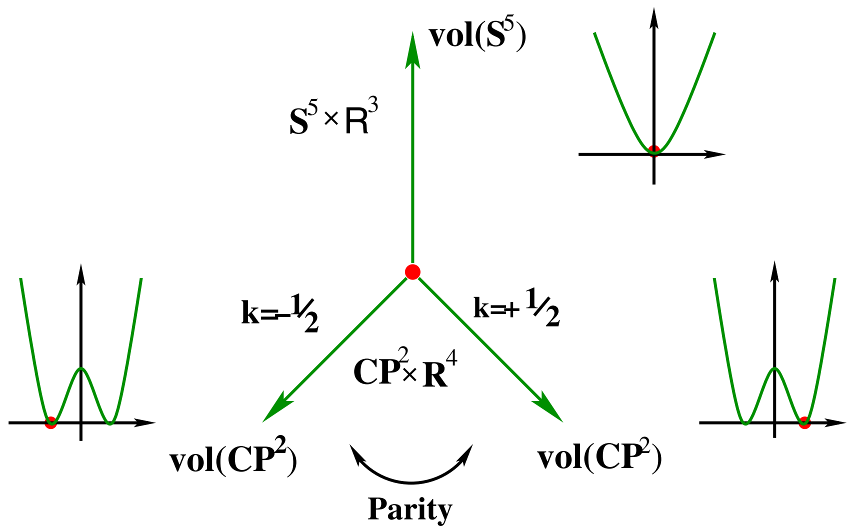


FIGURE 11. The moduli space of M-theory on the $Spin(7)$ cone over $SU(3)/U(1)$ can be compared to the vacuum structure of a system with spontaneous symmetry breaking. On this picture, the G-flux is measured by $k = \int_{\mathbb{C}\mathbf{P}^2} G/2\pi$.

This non-BPS state has a semi-classical mass proportional to the volume of $\mathbb{C}\mathbf{P}^1$, and is electrically charged under the global $U(1)_J$ symmetry of our gauge theory. Therefore, this state can be naturally identified with a vortex. Note that this state can be found only in the $\mathbb{C}\mathbf{P}^2$ phase (*i.e.* in the Higgs phase), in complete agreement with the low-energy physics.

In view of the interesting phenomena associated to branes in the conifold geometry, and their relationship to the conifold transition [**41, 40**], it would be interesting to learn more about the $Spin(7)$ transition using membrane probes in this background, and also to study the corresponding holographic renormalization group flows. For work in this area, see [**30, 45, 27**].

5.2. Relation to Geometric Transition. In the previous section we described the basic examples of topology changing transitions in exceptional holonomy manifolds, and commented on the important aspects of M-theory dynamics in these transitions. As we explain in this section, some of these transitions also have a nice interpretation in type IIA string theory, realizing dualities between backgrounds involving D6-branes and Ramond-Ramond fluxes in manifolds with more restricted holonomy. Specifically, we will consider two cases:

• $\mathbf{SU(3)} \to \mathbf{G_2}$: We start with a relation between the conifold transition in the presence of extra fluxes and branes and the G_2 flop transition in M-theory [**6**]. Note that these two transitions are associated with different holonomy groups[10] and, in particular, with different amounts of unbroken supersymmetry. The relation,

[10]That is why we refer to this case as $SU(3) \to G_2$.

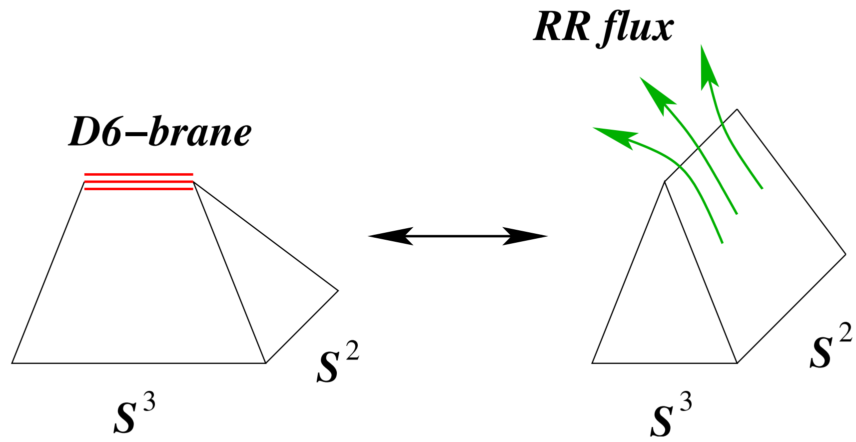


FIGURE 12. Geometric transition in IIA string theory connecting D6-branes wrapped around $\mathbf{S}^3$ in the *deformed* conifold geometry and *resolved* conifold with Ramond-Ramond 2-form flux through the $\mathbf{S}^2$.

however, appears when we introduce extra matter fields, represented either by D-branes or by fluxes. They break supersymmetry further, therefore, providing a relation between two different holonomy groups.

In order to explain how this works in the case of the conifold, let us consider type IIA theory on the deformed conifold geometry, $T^*\mathbf{S}^3 \cong \mathbf{S}^3 \times \mathbb{R}^3$. This already breaks supersymmetry down to $\mathcal{N} = 2$ in four dimensions. One can break supersymmetry further, to $\mathcal{N} = 1$, by wrapping a space-filling D6-brane around the supersymmetric (special Lagrangian) $\mathbf{S}^3$ cycle in this geometry. Then, a natural question to ask is: "What happens if one tries to go through the conifold transition with the extra D6-brane?". One possibility could be that the other branch is no longer connected and the transition is not possible. However, this is not what happens. Instead the physics is somewhat more interesting. According to [**54**], the transition proceeds, but now the two branches are smoothly connected, with the wrapped D6-brane replaced by Ramond-Ramond 2-form flux through the $\mathbf{S}^2$ cycle of the resolved conifold; see Figure12.

As we explained in section 4, both D6-branes and Ramond-Ramond 2-form tensor fields lift to purely geometric backgrounds in M-theory. Therefore, the geometric transition described above should lift to a transition between two purely geometric backgrounds in M-theory (hence, the name). Since these geometries must preserve the same amount of supersymmetry, namely $\mathcal{N} = 1$ in four dimensions, we conclude that we deal with a G_2 transition. In fact, it is the familiar flop transition in M-theory on a G_2 manifold [**1, 6**]:

$$X \cong \mathbf{S}^3 \times \mathbb{R}^4$$

Indeed, if we start on one of the three branches of this manifold and choose the 'M-theory circle' to be the fiber of the Hopf bundle in the 3-sphere

$$\mathbf{S}^1 \hookrightarrow \mathbf{S}^3 \to \mathbf{S}^2$$

we obtain the resolved conifold as a quotient space, $X/U(1) \cong \mathbf{S}^2 \times \mathbb{R}^4$. More precisely, we obtain a resolved conifold with Ramond-Ramond 2-form flux and no D6-branes because the circle action has no fixed points in this case. This gives us one side of the brane/flux duality, namely the right-hand side on the diagram below:

$$\begin{tikzcd}
\boxed{\begin{array}{l}\text{M-theory on } G_2 \\ \text{manifold } \mathbf{S}^3_{(1)} \times \mathbb{R}^4\end{array}} \arrow[r, leftrightarrow, "\text{flop}"] \arrow[d] & \boxed{\begin{array}{l}\text{M-theory on } G_2 \\ \text{manifold } \mathbf{S}^3_{(2)} \times \mathbb{R}^4\end{array}} \arrow[d] \\
\boxed{\begin{array}{l}\text{IIA on } \mathbf{S}^3 \times \mathbb{R}^3 \text{ with} \\ \text{D6-brane on } \mathbf{S}^3\end{array}} \arrow[r, leftrightarrow, "\text{geometric transition}"] & \boxed{\begin{array}{l}\text{IIA on } \mathbf{S}^2 \times \mathbb{R}^4 \text{ with} \\ \text{RR flux through } \mathbf{S}^2\end{array}}
\end{tikzcd}$$

Now, let us follow the G_2 flop transition in M-theory on the manifold $X \cong \mathbf{S}^3 \times \mathbb{R}^4$. As explained in the previous section, after the transition we obtain a G_2 manifold with similar topology, but the M-theory circle is now embedded in $\mathbb{R}^4$, rather than in $\mathbf{S}^3$. Acting on each $\mathbb{R}^4$ fiber, it yields $\mathbb{R}^3 = \mathbb{R}^4/U(1)$ as a quotient space with a single fixed point at the origin of the $\mathbb{R}^4$ (see the discussion of M-theory on the Taub-NUT space in section 4). Applying this to each fiber of the G_2 manifold X, we obtain the deformed conifold as the quotient space, $X/U(1) \cong \mathbf{S}^3 \times \mathbb{R}^3$, with the fixed point set $L = \mathbf{S}^3$. Since the latter is identified with the location of the space-filling D6-brane, we recover the other side of the brane/flux duality, illustrated in the above diagram.

Thus, we explained that the geometric transition — which is a highly non-trivial, non-perturbative phenomenon in string theory — can be understood as a G_2 flop transition in M-theory. Various aspects of this transition have been discussed in [**1, 6, 8, 11, 18, 13, 19, 7**]. As we show next, there is a similar relation between the phase transition in the G_2 holonomy manifold $X = \mathbb{C}\mathbf{P}^2 \times \mathbb{R}^3$ and the $Spin(7)$ conifold transition, discussed in the previous section.

• $\mathbf{G_2} \to \mathbf{Spin(7)}$: Consider type IIA string theory on the G_2 holonomy manifold

$$\mathbb{C}\mathbf{P}^2 \times \mathbb{R}^3 \tag{5.4}$$

which is obtained by resolving the cone over $SU(3)/U(1)^2$. As was discussed in the previous section, the corresponding moduli space has three classical branches connected by a singular phase transition. Motivated by the geometric transition in the conifold example, one could wrap an extra D6-brane over the $\mathbb{C}\mathbf{P}^2$ cycle and ask a similar question: "What happens if one tries to go through a phase transition?".

Using arguments similar to [**6**], one finds that the transition is again possible, via M-theory on a $Spin(7)$ manifold [**31**]. More precisely, after the geometric transition one finds type IIA string theory in a different phase of the G_2 manifold (5.4), where the D6-brane is replaced by RR flux through $\mathbb{C}\mathbf{P}^1 \subset \mathbb{C}\mathbf{P}^2$. This leads to a fibration:

$$\mathbf{S}^1 \hookrightarrow \mathbf{S}^5 \to \mathbb{C}\mathbf{P}^2$$

Hence the M-theory lift of this configuration gives a familiar $Spin(7)$ conifold,

$$X \cong \mathbf{S}^5 \times \mathbb{R}^3.$$

Similarly, one can identify the lift of $\mathbb{C}\mathbf{P}^2 \times \mathbb{R}^3$ with a D6-brane wrapped around $\mathbb{C}\mathbf{P}^2$ as the $Spin(7)$ manifold $\mathbb{C}\mathbf{P}^2 \times \mathbb{R}^4$, which is another topological phase of the

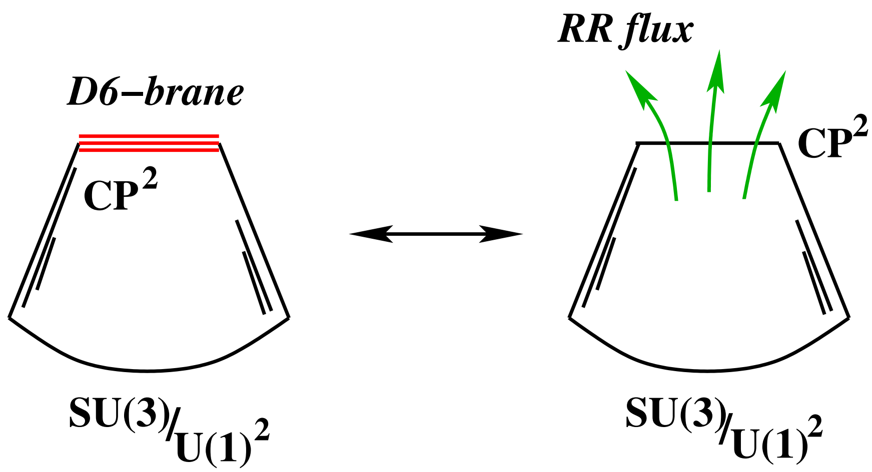


FIGURE 13. Geometric transition in IIA string theory connecting one branch of the G_2 manifold $\mathbb{C}\mathbf{P}^2 \times \mathbb{R}^3$, where D6-branes are warpped around the $\mathbb{C}\mathbf{P}^2$ cycle and another branch, where D6-branes are replaced by Ramond-Ramond 2-form flux through the $\mathbb{C}\mathbf{P}^1 \subset \mathbb{C}\mathbf{P}^2$.

Spin(7) conifold. Summarizing, we find that the conifold transition in M-theory on a *Spin*(7) manifold is related to a geometric transition in IIA string theory on the G_2 manifold (5.4) with branes/fluxes, as shown in the diagram below:

$$\begin{tikzcd}
\boxed{\begin{array}{l}\text{M-theory on } Spin(7) \\ \text{manifold } \mathbb{C}\mathbf{P}^2 \times \mathbb{R}^4\end{array}} \arrow[r, leftrightarrow, "\text{conifold transition}"] \arrow[d] & \boxed{\begin{array}{l}\text{M-theory on } Spin(7) \\ \text{manifold } \mathbf{S}^5 \times \mathbb{R}^3\end{array}} \arrow[d] \\
\boxed{\begin{array}{l}\text{IIA on } \mathbb{C}\mathbf{P}^2 \times \mathbb{R}^3 \text{ with} \\ \text{D6-brane on } \mathbb{C}\mathbf{P}^2\end{array}} \arrow[r, leftrightarrow, "\text{geometric transition}"] & \boxed{\begin{array}{l}\text{IIA on } \mathbb{C}\mathbf{P}^2 \times \mathbb{R}^3 \text{ with} \\ \text{RR flux through } \mathbb{C}\mathbf{P}^1\end{array}}
\end{tikzcd}$$

However, unlike its prototype with larger supersymmetry, this transition does not proceed smoothly.

References

1. B.S. Acharya, *On Realising N=1 Super Yang-Mills in M theory*, hep-th/0011089.
2. B.S. Acharya, *On Mirror Symmetry for Manifolds of Exceptional Holonomy*, Nucl. Phys. **B524** (1998) 269; "Exceptional Mirror Symmetry 1-14," in *Winter School on Mirror Symmetry, Vector Bundles and Lagrangian Submanifolds*, C. Vafa and S.-T. Yau, eds., AMS and International Press, Boston, 2001.
3. D.V. Alekseevsky, *Riemannian Spaces with Exceptional Holonomy*, Funct. Anal. Appl. **2** (1968) 97.
   R. Brown and A. Gray, *Riemannian Manifolds with Holonomy Group Spin*(9), *Differential Geometry (in honor of Kentaro Yano)*, Tokyo (1972) 41.
4. P. S. Aspinwall, B. R. Greene, D. R. Morrison, *Spacetime Topology Change: The Physics of Calabi-Yau Moduli Space*, hep-th/9311186.
5. P. Aspinwall, B. Greene and D. Morrison, *Multiple Mirror Manifolds and Topology Change in String Theory*, Phys. Lett. **B303** (1993) 249, hep-th/9301043.

6. M. Atiyah, J. Maldacena and C. Vafa, *An M-theory Flop as a Large N Duality*, J. Math. Phys. **42** (2001) 3209, hep-th/0011256.
7. M. Aganagic and C. Vafa, *G_2 manifolds, mirror symmetry and geometric engineering*, arXiv:hep-th/0110171.
8. M. Atiyah and E. Witten, *M-Theory Dynamics On A Manifold Of G_2 Holonomy*, Adv. Theor. Math. Phys. **6** (2002) 1.
9. R. Blumenhagen and V. Braun, *Superconformal field theories for compact G_2 manifolds*, JHEP **0112**, 006 (2001) [arXiv:hep-th/0110232].
10. R. Blumenhagen and V. Braun, *Superconformal field theories for compact manifolds with Spin(7) holonomy*, JHEP **0112**, 013 (2001) [arXiv:hep-th/0111048].
11. A. Brandhuber, J. Gomis, S. S. Gubser and S. Gukov, *Gauge theory at large N and new G_2 holonomy metrics*, Nucl. Phys. B. **611**, 179 (2001), hep-th/0106034.
12. M. Berger, *Sur les groupes d'holonomie des variétés á connexion affines et des variétés riemanniennes*, Bulletin de la Société Mathématique de France, **83** (1955) 279.
13. A. Brandhuber, *G_2 holonomy spaces from invariant three-forms*, Nucl. Phys. B. **629**, 393 (2002) [arXiv:hep-th/0112113].
14. R.L. Bryant, *Metrics with Exceptional Holonomy*, Ann. Math. **126** (1987) 525.
15. R. Bryant, S. Salamon, *On the Construction of some Complete Metrics with Exceptional Holonomy*, *Duke Math.* **J. 58** (1989) 829.
16. Z. W. Chong, M. Cvetic, G. W. Gibbons, H. Lu, C. N. Pope and P. Wagner, *General metrics of G_2 holonomy and contraction limits*, Nucl. Phys. B. **638** (2002) 459. arXiv:hep-th/0204064.
17. M. Cvetic, G.W. Gibbons, H. Lu, C.N. Pope, *Cohomogeneity One Manifolds of Spin(7) and G_2 Holonomy*, Phys. Rev. **D65** 106004 (2002).
18. M. Cvetic, G. W. Gibbons, H. Lu and C. N. Pope, *M-theory conifolds*, Phys. Rev. Lett. **88**, 121602 (2002) [arXiv:hep-th/0112098].
19. M. Cvetic, G. W. Gibbons, H. Lu and C. N. Pope, *A G_2 unification of the deformed and resolved conifolds*, Phys. Lett. B. **534**, 172 (2002) [arXiv:hep-th/0112138].
20. M. Cvetic, G. W. Gibbons, H. Lu, C. N. Pope, *New Complete Non-compact Spin(7) Manifolds*, Nucl. Phys. **B620** (2002) 29, hep-th/0103155.
21. L. J. Dixon, J. A. Harvey, C. Vafa and E. Witten, *Strings On Orbifolds.* 1 and 2, Nucl. Phys. B **261** (1985) 678; Nucl. Phys. B. **274** (1986) 285.
22. T. Eguchi and A. J. Hanson, *Asymptotically Flat Selfdual Solutions To Euclidean Gravity*, Phys. Lett. B **74** (1978) 249; T. Eguchi, P. B. Gilkey and A. J. Hanson, *Gravitation, Gauge Theories And Differential Geometry*, Phys. Rept. **66** (1980) 213.
23. T. Eguchi and Y. Sugawara, *CFT description of string theory compactified on non-compact manifolds with G_2 holonomy*, Phys. Lett. B **519**, 149 (2001) [arXiv:hep-th/0108091].
24. T. Eguchi and Y. Sugawara, *String theory on G_2 manifolds based on Gepner construction*, Nucl. Phys. B **630**, 132 (2002) [arXiv:hep-th/0111012].
25. J. Gomis, *D-Branes, Holonomy and M-Theory*, Nucl. Phys. **B606** (2001) 3, `hep-th/0103115`.
26. B. Greene, D. Morrison and A. Strominger, *Black Hole Condensation and the unification of String Vacua*, Nucl. Phys. **B451** (1995) 109, hep-th/9504145.
27. U. Gürsoy, C. Nuñez, M. Schvellinger, *RG flows from Spin(7), CY 4-fold and HK manifolds to AdS, Penrose limits and pp waves*, JHEP **6**, (2002) 015.
28. G. W. Gibbons, D. N. Page, C. N. Pope, *Einstein Metrics on S^3, $\mathbb{R}^3$ and $\mathbb{R}^4$ Bundles*, *Commun.Math.Phys* **127** (1990) 529-553.
29. S. Gukov and J. Sparks, *M theory on Spin(7) manifolds. I*, Nucl. Phys. **B625** (2002) 3, hep-th/0109025.
30. S. Gukov and D. Tong, *D-brane probes of G_2 holonomy manifolds*, Phys. Rev. **D66** (2002) 087901. hep-th/0202125.
S. Gukov and D. Tong, *D-brane probes of special holonomy manifolds, and dynamics of $N = 1$ three-dimensional gauge theories*, JHEP **0204**, 050 (2002), hep-th/0202126.
31. S. Gukov, J. Sparks and D. Tong, *Conifold transitions and five-brane condensation in M-theory on Spin(7) manifolds*, Classical Quantum Gravity **20** (2003) 665. hep-th/0207244.
32. S. Gukov, S. T. Yau and E. Zaslow, *Duality and fibrations on G_2 manifolds*, Turkish J. Math. **7** (2003) 61. arXiv:hep-th/0203217.
33. Y.-H. He, *G_2 Quivers*, JHEP **0302** (2003) 023.

34. N. Hitchin, *Stable forms and special metrics*, in *Global Differential Geometry: The mathematical legacy of Alfred Gray* (Bilbao, Spain 2000) 70 Contemp. Math. 288, AMS 2001. math.DG/0107101.
35. N. Hitchin, *The geometry of three-forms in six and seven dimensions*, math.DG/0010054.
36. R. Harvey and H.B. Lawson, Jr., *Calibrated geometries*, Acta Math. **148** (1982) 47.
37. D. Joyce, *Compact Manifolds with Special Holonomy*, Oxford University Press, 2000.
38. D. Joyce, *Special Lagrangian m-folds in C^m with symmetries*, Duke Math. J. 115 (2002) 6. math.DG/0008021.
39. A. Kovalev, *Twisted connected sums and special Riemannian holonomy*, J. Reine Ang. Math. **565** (2003) 125. math.dg/0012189.
40. I. R. Klebanov and M. J. Strassler, *Supergravity and a confining gauge theory: Duality cascades and chiSB-resolution of naked singularities*, JHEP **0008**, 052 (2000), hep-th/0007191.
41. I. Klebanov and E. Witten, *Superconformal field theory on threebranes at a Calabi-Yau singularity*, Nucl. Phys. **B536**, 199 (1998), hep-th/9807080.
42. H. Kanno, Y. Yasui, *On Spin(7) holonomy metric based on SU(3)/U(1) I* and *II*, J. Geom. Phys. **43** (2002) 293-209, 310-326. hep-th/0108226 and hep-th/0111198.
43. J. Levine, *Semi-Free Circle Actions on Spheres*, Invent. Math. **22** (1973) 161.
44. J.-H. Lee and N. C. Leung, *Geometric Structures on G_2 and Spin(7)-Manifolds*, math.DG/0202045.
45. A. Loewy, Y. Oz, *Branes in Special Holonomy Backgrounds*, Phys.Lett. **B537** (2002) 147. hep-th/0203092.
46. R. McLean, *Deformations of Calibrated Submanifolds*, Comm. Anal. Geom. **6** (1998) 705-747.
47. D. Montgomery, L. Zippin, *Topological transformation groups*, New York: Interscience, 1955.
48. H. Partouche and B. Pioline, *Rolling among G_2 vacua*, JHEP **0103** (2001) 005, [arXiv:hep-th/0011130].
49. R. Roiban, C. Romelsberger and J. Walcher, *Discrete torsion in singular G_2-manifolds and real LG*, Adv. Theor. Math. Phys. **6** (2003) 207. arXiv:hep-th/0203272.
50. A. Strominger, *Massless black holes and conifolds in string theory* Nucl. Phys. B. **451**, 96 (1995), hep-th/9504090.
51. S. Shatashvili, C. Vafa, *Superstrings and Manifolds of Exceptional Holonomy*, Selecta Math (N. S.) **1** (1995) 347. hep-th/9407025.
52. K. Sugiyama and S. Yamaguchi, *Cascade of special holonomy manifolds and heterotic string theory*, Nucl. Phys. **B622**, 3 (2002) [arXiv:hep-th/0108219].
53. P.K. Townsend, *The eleven-dimensional supermembrane revisited*, Phys. Lett. **B350** (1995) 184.
54. C. Vafa, *Superstrings and topological strings at large N*, J. Math. Phys. **42** (2001) 2798, hep-th/0008142.
55. E. Witten, *String theory dynamics in various dimensions*, Nucl. Phys. **B443** (1995) 85, [arXiv:hep-th/9503124].
56. E. Witten, "*Phases of $N = 2$ Models in Two Dimensions*", Nucl. Phys. **B403** (1993) 159, hep-th/9301042.
57. E. Witten, *On Flux Quantisation in M-Theory and the Effective Action*, J. Geom. Phys. **22**, 1 (1997), hep-th/9609122.
58. S.-T. Yau, *On the Ricci curvature of a compact Kähler manifold and the Monge-Ampére equations. I*, Commun. Pure Appl. Math. **31** (1978) 339.

Jefferson Physical Laboratory, Harvard University, Cambridge, MA 02138, U.S.A.
E-mail address: gukov@tomonaga.harvard.edu

Clay Mathematics Proceedings
Volume 3, 2004

Special holonomy and beyond

Nigel Hitchin

Abstract. Manifolds with special holonomy are traditionally associated with the existence of parallel spinors, but Calabi-Yau threefolds and G_2 manifolds also arise naturally in the context of a nonlinear form of Hodge theory: finding critical points of a natural functional defined on the differential p-forms in a fixed cohomology class. The advantage of this point of view is that it gives a natural approach to the moduli spaces of these structures, which appear as open sets in the appropriate cohomology group, and also leads to other, less familiar, geometric structures.

1. Invariant functionals and special holonomy

The list of special holonomy groups of Riemannian manifolds which emerged from Berger's work in the 1950's was recognized later to contain all those manifolds which possess covariant constant spinors, that is spinor fields ψ with $\nabla\psi = 0$. They are:

(1) $SU(n)$: Calabi-Yau manifolds
(2) $Sp(n)$: hyperkähler manifolds
(3) G_2: special manifolds in dimension 7
(4) $Spin(7)$: special manifolds in dimension 8.

All of these are Einstein metrics with vanishing Ricci tensor. A second list of special metrics consists of those with Killing spinors, spinor fields ψ satisfying $\nabla_X\psi = \lambda X\psi$. These are:

(1) spheres
(2) Einstein-Sasakian manifolds in dimension $2k+1$
(3) 3-Sasakian manifolds in dimension $4k+3$
(4) nearly Kähler manifolds in dimension 6
(5) weak holonomy G_2 manifolds in dimension 7.

These are Einstein metrics with $R_{ij} = \Lambda g_{ij}$ and $\Lambda > 0$. The link between them, established by Bär in 1993 [**1**], is that the Killing spinor on a manifold M^n extends to become a covariant constant spinor on the $(n+1)$-dimensional cone over M^n.

In fact all of these geometries are defined by reference to a closed form or collection of forms, and for some of them there is an alternative setting, independent

1991 *Mathematics Subject Classification.* Primary 53C29, 53C26, 32Q25,53C80, 58E11.
Key words and phrases. Special holonomy, Calabi-Yau.

of spinors or Riemannian metrics, which we shall present here. The approach is very natural, but leads on to some geometrical structures which appear not to have been considered before.

We know what a non-degenerate symmetric bilinear form is, and it gives us the notion of a metric, but what is a non-degenerate exterior form? One answer is to ask that it should be an element of the vector space $\Lambda^p V^*$ which lies in an *open orbit* of the action of the general linear group $GL(V)$. For example, $GL(2n, \mathbf{R})$ acts on $\Lambda^2\mathbf{R}^{2n}$ with an open orbit: the orbit consists of nondegenerate alternating bilinear forms.

Now $GL(V)$ acts on $\Lambda^p V^*$, but $\dim \Lambda^p V^* = n!/(n-p)!p!$ and $\dim GL(V) = n^2$ so in particular to get an open orbit we need

$$\frac{n!}{(n-p)!p!} \leq n^2$$

which puts serious constraints on when this occurs. We list the possibilities below, but first we note that the word "nondegenerate" has specific connotations for bilinear skew forms so instead of using this word we make the

Definition 1.1. If $\rho \in \Lambda^p V^*$ lies on an open orbit under $GL(V)$ it is called **stable**.

Note that an open orbit in $\Lambda^p V^*$ implies the existence of an open orbit in $\Lambda^{n-p}V^*$ by duality, so stability really occurs for pairs of forms in complementary dimensions.

There is a list, known in one form or another for many years, of when stable forms arise:

(1) $p = 1$ all n: stable = non-zero
(2) $p = 2$ all n: if $n = 2m+1$ or $2m$, then stable forms are of rank $2m$ when considered as bilinear forms. The stabilizer subgroup in dimension $2m$ is $Sp(2m, \mathbf{R})$.
(3) $p = 3$, $n = 6$: stabilizer $SL(3, \mathbf{C})$ or $SL(3, \mathbf{R}) \times SL(3, \mathbf{R})$
(4) $p = 3$, $n = 7$: stabilizer G_2 or its non-compact form
(5) $p = 3$, $n = 8$: stabilizer $SU(3), SU(2,1)$ or $SL(3, \mathbf{R})$ in the adjoint representation

and for the last three this means that we have normal forms at the linear algebra level (for stabilizers $SL(3, \mathbf{C}), G_2$ and $SU(3)$) given by:

- $dx_1dx_2dx_3 - dy_1dy_2dx_3 - dy_2dy_3dx_1 - dy_3dy_1dx_2$
- $(dx_1dx_2 - dx_3dx_4)dx_5 + (dx_1dx_3 - dx_4dx_2)dx_6 + (dx_1dx_4 - dx_2dx_3)dx_7 + dx_5dx_6dx_7$
- $dx_1dx_2dx_3 - dx_1(dx_7dx_4 + dx_5dx_6) - dx_2(dx_7dx_5 + dx_6dx_4) + dx_3(dx_7dx_6 + dx_4dx_5) + dx_8(dx_4dx_5 + dx_6dx_7)$

Now stability implies that there is an open set $U = GL(V)/H \subset \Lambda^p V^*$, and the stabilizer H of a point in U is $Sp(2m, \mathbf{R}), G_2 \ldots$ etc. Each one of these groups preserves also a volume form (metric or symplectic) so there is a $GL(V)$-invariant map

$$\phi : U \to \Lambda^n V^*.$$

Invariance under the scalar matrices implies that

$$\phi(\lambda^p \rho) = \lambda^n \phi(\rho)$$

so ϕ is homogeneous of degree n/p, and extends to a continuous map

$$\phi : \Lambda^p V^* \to \Lambda^n V^*$$

which is smooth on the open set $U \subset \Lambda^p V^*$. The derivative of ϕ at ρ is a linear map from $\Lambda^p V^*$ to $\Lambda^n V^*$ i.e.

$$D\phi \in (\Lambda^p V^*)^* \otimes \Lambda^n V^*$$

But $(\Lambda^p V^*)^* \cong \Lambda^{n-p} V^* \otimes \Lambda^n V$ so there is a unique

$$\hat{\rho} \in \Lambda^{n-p} V^*$$

such that the derivative can be expressed as

$$D\phi(\dot{\rho}) = \hat{\rho} \wedge \dot{\rho}.$$

Note that if we take $\dot{\rho} = \rho$, then Euler's formula for homogeneous functions implies that

$$\hat{\rho} \wedge \rho = \frac{n}{p} \phi(\rho).$$

This nonlinear algebraic operation $\rho \mapsto \hat{\rho}$ is familiar enough in the concrete cases:

(1) for $n = 6$, $p = 3$, $\hat{\rho}$ is determined by the property that $\Omega = \rho + i\hat{\rho}$ is a complex $(3, 0)$-form preserved by $SL(3, \mathbf{C})$,
(2) for $n = 7$, $p = 3$ or 4, $\hat{\rho} = *\rho$, where $*$ is the Hodge star operator for the inner product on V,
(3) for $n = 8$, $p = 3$ or $p = 5$, $\hat{\rho} = - * \rho$.

Example 1.2. Take the symplectic case of $\rho = \omega \in \Lambda^2 V^*$. Then ω is stable if it is nondegenerate, i.e. if ω lies in the open set U defined by $\det \omega_{ij} \neq 0$. The volume form is the usual Liouville volume $\phi(\omega) = \omega^m$ (up to a universal scale). Differentiating, we get

$$\hat{\omega} = m\omega^{m-1} \in \Lambda^{2m-2} V^*$$

and $\omega \wedge m\omega^{m-1} = m\omega^m$ verifying the homogeneity.

Suppose now that we have a compact manifold M^n, and a p-form $\rho \in \Omega^p(M)$ which is stable at each point. This means that topologically we must have a reduction of the structure group of the tangent bundle to the stabilizer – one of the groups in the list above. Then since $\rho \in C^\infty(M, \Lambda^p T^*)$ we obtain

$$\phi(\rho) \in C^\infty(M, \Lambda^n T^*).$$

Now define the diffeomorphism-invariant functional

$$\Phi(\rho) = \int_M \phi(\rho).$$

We now ask the question: *what is a critical point of the functional* $\Phi(\rho) = \int_M \phi(\rho)$ *for* ρ *a closed form in a fixed cohomology class in* $H^p(M, \mathbf{R})$*?*

This is a nonlinear version of Hodge theory. The answer is obtained by calculating the first variation:

$$\delta\Phi(\dot{\rho}) = \int_M D\phi(\dot{\rho}) = \int_M \hat{\rho}\wedge\dot{\rho}.$$

Because ρ is varying in a fixed cohomology class, $\dot{\rho} = d\alpha$ for some α and so

$$\delta\Phi(\dot{\rho}) = \int_M \hat{\rho}\wedge d\alpha = \pm\int_M d\hat{\rho}\wedge\alpha.$$

The critical points which are stable forms therefore occur when

$$d\rho = 0 = d\hat{\rho}.$$

EXAMPLE 1.3. Take the symplectic case. If $\omega \in \Omega^2(M)$ lies in a fixed cohomology class in $H^2(M,\mathbf{R})$ then

$$\Phi(\omega) = \int_M \omega^m = [\omega]^m[M] = const.$$

so any symplectic form is a critical point. On the other hand suppose we take the stable form $\rho = \omega^{m-1}$ as the basic object and only suppose that this one is closed and lies in a fixed cohomology class in $H^{2m-2}(M,\mathbf{R})$. This time $\Phi(\rho)$ is not constant, and the critical points are where

$$d\hat{\rho} = d\omega = 0$$

or in other words when ω is symplectic.

A critical point can never be an (isolated) Morse critical point since the functional is invariant by the full group of diffeomorphisms of M. We can ask nevertheless if it is nondegenerate transverse to the action of *Diff* (M) – an infinite dimensional version of a Morse-Bott critical point. In the symplectic case of $\rho = \omega^m$ above, we calculate the second variation to see this:

$$\delta^2\Phi(\dot{\rho}_1,\dot{\rho}_2) = m(m-1)\int_M \omega^{m-2}\dot{\omega}_1\dot{\omega}_2$$

The degeneracy subspace is defined by $\delta^2\Phi(\dot{\rho}_1,\dot{\rho}_2) = 0$ for all $\dot{\rho}_2 = d\sigma$, which gives

$$0 = (m-1)\int_M \omega^{m-2}\dot{\omega}_1\dot{\omega}_2 = \int_M \dot{\omega}_1\dot{\rho}_2 = \int_M \dot{\omega}_1\wedge d\sigma.$$

So $\dot{\rho}_1$ is in the degeneracy space iff $d\dot{\omega}_1 = 0$.

Now recall the *strong Lefschetz property*:

$$\cup[\omega]^{m-2} : H^2(M,\mathbf{R}) \cong H^{2m-2}(M,\mathbf{R})$$

for symplectic manifolds. We know that $d\dot{\omega}_1 = 0$ and since

$$\dot{\rho}_1 = (m-1)\omega^{m-2}\dot{\omega}_1$$

if strong Lefschetz holds then $\dot{\rho}_1$ exact implies $\dot{\omega}_1$ is exact, i.e. $\dot{\omega}_1 = d\alpha$. In this case we define the vector field X by $i_X\omega = \alpha$ and then

$$\dot{\omega}_1 = \mathcal{L}_X\omega$$

which means that the variation is tangential to the action of *Diff* (M). The features we encounter here are:

(1) Φ is nondegenerate transverse to *Diff* (M) orbits if strong Lefschetz holds,

(2) the local moduli space $\cong$ open set in $H^{2m-2}(M,\mathbf{R})$,
(3) the Hessian of Φ defines an indefinite metric on the moduli space.

Note that traditionally the local moduli space of symplectic structures is given by the de Rham cohomology class $[\omega] \in H^2(M,\mathbf{R})$ but $[\omega] \mapsto [\omega^{m-1}]$ is a local diffeomorphism if strong Lefschetz holds, which brings the two points of view together.

A more interesting example than the symplectic one is the case of 3-forms in six dimensions as in [**6**]. In this case $\rho \in \Lambda^3T^*$ is the real part of a complex locally decomposable form ν, and $\phi(\rho) = i\nu\wedge\bar{\nu}/2$. Here $\hat{\rho} = \operatorname{Im}\nu \in \Lambda^3T^*$ and so $\nu = \rho + i\hat{\rho}$ is a $(3,0)$ form for an almost complex structure on M. The critical points of Φ on a class in $H^3(M,\mathbf{R})$ give

$$d(\rho + i\hat{\rho}) = 0$$

and this implies that the complex structure is integrable and ν is a non-vanishing holomorphic three-form. Thus M^6 is given the structure of a complex threefold with trivial canonical bundle (for example a Calabi-Yau manifold). The essential features in this case are:

(1) Φ is nondegenerate transverse to *Diff* (M) orbits if the $\partial\bar{\partial}$-lemma holds,
(2) the local moduli space is isomorphic to an open set in $H^3(M,\mathbf{R})$, which gives another proof in this dimension of the unobstructedness of moduli,
(3) the Hessian of Φ defines an indefinite *special Kähler metric* on the moduli space.

In dimension 7, $\rho \in \Lambda^3T^*$ which is stable at each point defines an almost G_2-structure and $\phi(\rho)$ is the Riemannian volume of the G_2-metric. We find $\hat{\rho} = *\rho \in \Lambda^4T^*$, with $*$ the Hodge star operator of this metric and the critical points of Φ are where

$$d\rho = 0 = d*\rho.$$

This is known to be equivalent to the condition that ρ is covariant constant with respect to the metric and so defines a metric with holonomy G_2.

Using Hodge theory, the features here are that transverse nondegeneracy holds and, as a consequence, the moduli space is an open set in $H^3(M,\mathbf{R})$. The Hessian of Φ again defines an indefinite metric on this space.

All of these structures clearly have a distinguished class of flat coordinates on their moduli spaces. This would equally be true of the final structure thrown out by the stability condition in dimension 8 *if only we knew of a nontrivial example.* We study this geometry next nevertheless.

As noted above, $\rho \in \Lambda^3T^*$ defines the structure of the Lie algebra of $SU(3)$ on T. This means that we have a metric defined by the Killing form $\operatorname{tr}(XY)$ and the 3-form is then

$$\rho(X,Y,Z) = \operatorname{tr}(X[Y,Z]).$$

The integrand in the functional is then the Riemannian volume form $\phi(\rho)$ and

$$\hat{\rho} = *\rho \in \Lambda^5T^*.$$

Critical points of the functional on a cohomology class in $H^3(M,\mathbf{R})$ are therefore given by

$$d\rho = 0 = d*\rho.$$

This is directly analogous to the G_2 situation but there are essential differences, in particular we do not get a metric of special holonomy since $SU(3)$ does not appear in Berger's list.

One approach to this geometry is through *triality* in 8 dimensions. Recall that there are three real 8-dimensional representations of $Spin(8)$: the vector representation T and the two half-spin representations S^+, S^-. These are permuted by an outer automorphism $\Delta^+ : Spin(8) \to Spin(8)$ of order three. Now we can Clifford multiply a spinor $\psi \in S^+$ by a vector $x \in T$ to get $x \cdot \psi \in S^-$ and since x is skew adjoint in this action,

$$(x\cdot\psi, x\cdot\psi) = (-x^2\cdot\psi, \psi)$$

But in the Clifford algebra $x^2 = -(x,x)1$ so this expression is

$$\|x\|^2\|\psi\|^2.$$

If x, ψ and $x\cdot\psi$ all lived in the same space we would call this an orthogonal multiplication. In fact, a simple weight calculation shows that, restricted to $\mathrm{Ad} : SU(3) \to Spin(8)$,

$$T \cong S^+ \cong S^-$$

so we *do* get an orthogonal multiplication on the Lie algebra of $SU(3)$, in fact it is:

$$A\cdot B = \omega AB - \bar{\omega}BA - \frac{i}{\sqrt{3}}\,\mathrm{tr}(AB)1$$

where $\omega = (1+i\sqrt{3})/2$.

Thus if M^8 has an adjoint $SU(3)$ structure, as described above, the $SU(3)$-invariant isomorphism $S^+ \cong T$ defines a spin 3/2 field:

$$\psi \in C^\infty(M, S^+ \otimes \Lambda^1).$$

This ψ, it turns out, satisfies the *Rarita-Schwinger equation* (see [**7**])

$$D\psi = 0, \quad d^*\psi = 0$$

where for the first equation we think of ψ as a spinor with values in the bundle Λ^1 and apply the Dirac operator D and in the second we think of it as a one-form with values in S^+. Now take the second covariant derivative of ψ: this is a section $\nabla^2\psi \in C^\infty(S^+\otimes\Lambda^1\otimes\Lambda^1\otimes\Lambda^1)$ or in more conventional terms

$$\nabla^2\psi = \psi_{i;jk}$$

Covariantly differentiate the equation $D\psi = 0$ to obtain $\sum_j e_j\psi_{i;jk} = 0$ and then contract to get

$$\sum_{i,j} e_j\psi_{i;ji} = 0.$$

Now differentiate $d^*\psi = 0$ which gives $\sum_i \psi_{i;ij} = 0$ so that

$$\sum_{i,j} e_j\psi_{i;ij} = 0.$$

Subtract the two displayed formulae and re-express $\psi_{i;ji} - \psi_{i;ij}$ using the curvature tensor to get

$$\sum_{i,j} R_{ij} e_i \psi_j = 0$$

which implies that 8 of the 36 components of the Ricci tensor vanish. These manifolds are not quite Einstein manifolds, clearly, but yet have some common features. Unfortunately the only known compact example is the group manifold $SU(3)$, though one can find nontrivial local examples.

2. Cohomogeneity one metrics

Variational characterizations can be useful in a practical way in situations of symmetry but unfortunately compact manifolds with holonomy $SU(n), G_2$ or $Spin(7)$ have no continuous symmetries because the Ricci tensor vanishes and so any Killing vector is covariant constant. The formalism described above can still be used, however, and we shall see that, as in [**5**], [**3**], it provides a practical approach to deriving differential equations, if not solving them.

The first aspect of this involves duality of spaces of forms. If $d\beta \in \Omega^p$ is exact and $\gamma \in \Omega^{n-p}$ is closed then by Stokes' theorem

$$\int_M d\beta \wedge \gamma = 0.$$

This implies that the natural pairing of p-forms and $(n-p)$-forms gives a formal isomorphism

$$(\Omega^p_{exact})^* \cong \Omega^{n-p}/\Omega^{n-p}_{closed}$$

and since $d : \Omega^{n-p} \to \Omega^{n-p+1}_{exact}$ has kernel the closed $(n-p)$-forms we formally have (we are not concerned with distributions here)

$$(\Omega^p_{exact})^* \cong \Omega^{n-p+1}_{exact}.$$

In the previous lecture we were restricting a functional to a cohomology class. Now let's consider the internal geometry of such a class

$$\mathcal{A} = \{\alpha \in \Omega^p : d\alpha = 0, [\alpha] = a \in H^p(M^n, \mathbf{R})\}.$$

This is an affine space modelled on Ω^p_{exact}, and so its tangent bundle is trivial and

$$T\mathcal{A} \cong \mathcal{A} \times \Omega^p_{exact}.$$

Using the duality above, its cotangent bundle is

$$T^*\mathcal{A} \cong \mathcal{A} \times (\Omega^p_{exact})^* = \mathcal{A} \times \Omega^{n-p+1}_{exact}.$$

Bearing this in mind, let's return to the 7-dimensional case, but now considering 4-forms instead of 3-forms. There are stable 4-forms and we can do the variational approach of the last lecture and see G_2 manifolds as critical points of a functional on 4-forms in a given cohomology class, but the 4-form approach gives a bit more, because we now have

$$(\Omega^4_{exact})^* = \Omega^4_{exact}$$

so $\mathcal{A}$ has a (flat indefinite) metric

$$(d\eta, d\eta) = \int_M \eta \wedge d\eta.$$

To make use of this, first restrict to the trivial cohomology class. Then we have two functionals:

(1) $\Phi(d\eta)$ – the G_2 volume
(2) $Q(d\eta) = \int_M \eta \wedge d\eta$ – from the indefinite inner product.

We set up then a *constrained* variational problem – finding the critical points of Φ subject to the constraint $Q(d\eta) = 1$. The first variation gives:

$$\delta\Phi(d\dot{\eta}) = \int_M *d\eta \wedge d\dot{\eta}$$

$$\delta Q(d\dot{\eta}) = 2\int_M \dot{\eta} \wedge d\eta.$$

Using a Lagrange multiplier λ and Stokes' theorem the equations become

$$d(*\rho) = \lambda\rho$$

where $\rho = d\eta$.

A manifold with this structure is called a *weak holonomy G_2 manifold*. It is an Einstein manifold with positive scalar curvature and has a Killing spinor. As such it defines an incomplete $Spin(7)$ metric on the cone, but which also serves as a model for asymptotically conical $Spin(7)$ metrics. There are plenty of weak holonomy G_2 manifolds, since a 3-Sasakian manifold and its squashed deformation provides an example and thanks to the work of Boyer, Galicki et al (see [**2**]) there are infinitely many of these.

EXAMPLE 2.1. We can find homogeneous examples by applying the variational approach to invariant forms. For example, $S^7 \subset \mathbf{H}^2$ is acted on by $Sp(2) \times Sp(1)$ (acting by the quaternionic matrix on the left and the quaternionic scalar on the right). This makes $S^7 \to S^4$ into a principal $SU(2)$-bundle with $\alpha_1, \alpha_2, \alpha_3$ the components of a connection form (the 1-instanton bundle in fact) and $\omega_1, \omega_2, \omega_3$ the components of the curvature form. The closed invariant 4-forms are then spanned by $d\sigma, d\tau$ where

$$\sigma = \alpha_1 \wedge \alpha_2 \wedge \alpha_3, \qquad \tau = \alpha_1 \wedge \omega_1 + \alpha_2 \wedge \omega_2 + \alpha_3 \wedge \omega_3.$$

This provides a two-dimensional vector space of invariant exact 4-forms $xd\sigma + yd\tau$ on which we can define the two functionals and easily find the critical points, which give the squashed 7-sphere.

In the case where the cohomology class $\mathcal{A}$ is non-trivial, Q defines an indefinite metric and we take the *gradient vector field* $X = d\xi$ of $\Phi(\rho)$. Since

$$\int_M \xi \wedge d\varphi = \int_M *\rho \wedge d\varphi$$

we have $X = d\xi = d * \rho$ which yields the gradient flow equation:

$$\frac{\partial \rho}{\partial t} = d * \rho.$$

But $d\rho = 0$, so

$$d(*\rho \wedge dt + \rho) = 0$$

and then

$$*\rho \wedge dt + \rho$$

satisfies precisely the algebraic condition to be the 4-form stabilized by $Spin(7)$ (not a "stable form" in the sense above.) Being closed, it defines a Riemannian metric with holonomy $Spin(7)$ on $\mathbf{R} \times M^7$, or a subinterval thereof.

EXAMPLE 2.2. The formalism above can be used to give cohomogeneity one examples of non-compact $Spin(7)$ manifolds. Again using S^7 as the 7-manifold with symmetry group $Sp(2) \times Sp(1)$ one obtains the Bryant-Salamon $Spin(7)$ manifold, the first nontrivial complete example to be discovered. If we weaken the symmetry to $Sp(2) \times U(1)$ then the invariant forms are spanned by

$$d(\alpha_1 \wedge \alpha_2 \wedge \alpha_3), \qquad d(\alpha_1 \wedge \omega_1 + \alpha_2 \wedge \omega_2), \qquad (\alpha_3 \wedge \omega_3).$$

With this, one can easily write the gradient flow which gives the system of ODE's solved by Cvetič et al [**4**]. Their solutions give in particular on $M = \mathbf{R}^8$ an analogue of the Taub-NUT metric and one on the total space $S^+ \to S^4$ of the spin bundle over the sphere. They each have the form

$$\frac{(r+\ell)^2 dr^2}{(r+3\ell)(r-\ell)} + \frac{\ell^2(r+3\ell)(r-\ell)}{(r+\ell)^2}\sigma^2 + \frac{1}{4}(r+3\ell)(r-\ell)(D\mu_i)^2 + \frac{1}{2}(r^2-\ell^2)d\Omega_4^2$$

for a different range of values.

If we move now to six dimensions then the duality of forms tells us that

$$(\Omega^3_{exact})^* \cong \Omega^4_{exact}.$$

Here we work with two objects – $\rho \in \Omega^3$ a stable form with stabilizer $SL(3, \mathbf{C})$ and $\omega^2 \in \Omega^4$ with stabilizer $Sp(6, \mathbf{R})$. Compatibility of these two forms leads to a reduction of structure group to $SU(3)$ and the compatibility conditions are:

$$\omega \wedge \rho = 0, \quad \phi(\rho) = \lambda \omega^3.$$

The first says, from the complex point of view that ω is of type $(1,1)$, and from the symplectic point of view that ρ is primitive. The second says that the norm of ρ is constant.

We can consider here a constrained variational problem too, taking $\rho = d\alpha \in \Omega^3$ an exact 3-form and

$$\sigma = \omega^2 = d\beta \in \Omega^4$$

an exact 4-form. The duality defines the bilinear pairing

$$B(\rho, \omega^2) = \int_M d\alpha \wedge \beta$$

and we have two functionals

(1) $A(\rho, \sigma) = \Phi(\rho) + \Phi(\sigma)$ – the sum of volumes

(2) $B(\rho,\sigma)$.

The critical points of A subject to $B=1$ are given by the equations (after some scale changes)

$$d\hat{\rho}=\omega^2, \quad d\omega=\rho.$$

Interestingly, the compatibility conditions to reduce to $SU(3)$ follow from these equations, for firstly

$$\omega\wedge\rho=\omega\wedge d\omega=d(\omega^2)/2=0.$$

It follows from this that ω is type $(1,1)$ so that $\omega\wedge\hat{\rho}=0$ too. Then

$$\omega^3=\omega\wedge d\hat{\rho}=d(\omega\wedge\hat{\rho})-d\omega\wedge\hat{\rho}=\rho\wedge\hat{\rho}.$$

The structure on M^6 defined by this pair of forms is called a *weak holonomy* $SU(3)$ *structure* or *nearly Kähler structure*. Here there are few known examples: $S^6, \mathbf{C}P^3$, $U(3)/T$ and $S^3\times S^3$. The general class of manifolds of this type are Einstein with positive scalar curvature and have Killing spinors.

To pursue the analogue of the gradient flow here we take two cohomology classes $\mathcal{A}\in H^3(M,\mathbf{R})$ and $\mathcal{B}\in H^4(M,\mathbf{R})$. Then $T(\mathcal{A}\times\mathcal{B})=\mathcal{A}\times\mathcal{B}\times\Omega^3_{exact}\times\Omega^4_{exact}$. The pairing

$$B(\rho,\omega^2)=\int_M d\alpha\wedge\beta$$

defines an indefinite metric as before, but more usefully a symplectic structure

$$\omega((\rho_1,\sigma_1),(\rho_2,\sigma_2))=\langle\rho_1,\sigma_2\rangle-\langle\rho_2,\sigma_1\rangle$$

on $\mathcal{A}\times\mathcal{B}$. We now take the function

$$H=\Phi(\rho)-\Phi(\sigma)$$

and derive the *Hamiltonian flow* equations which are:

$$\begin{aligned}\frac{\partial\rho}{\partial t} &= d\omega\\ \frac{\partial\sigma}{\partial t} &= \omega\wedge\frac{\partial\omega}{\partial t}=-d\hat{\rho}.\end{aligned}$$

It happens that the compatibility conditions for $SU(3)$ are preserved by the flow and then

$$dt\wedge\omega+\rho$$

defines a G_2 metric on an interval of $\mathbf{R}\times M^6$.

EXAMPLE 2.3. In [**3**] the authors take $M^6=S^3\times S^3$ and cohomology classes in

$$H^3(M)\cong\mathbf{Z}\oplus\mathbf{Z}, \quad H^4(M)=0.$$

Left-invariance of forms yields

$$\rho=n\,\Sigma_1\,\Sigma_2\,\Sigma_3-m\,\sigma_1\,\sigma_2\,\sigma_3+x_1\,(\sigma_1\,\Sigma_2\,\Sigma_3-\sigma_2\,\sigma_3\,\Sigma_1)$$
$$+\ 2\text{ cyclic terms}$$
$$\sigma=y_1\,\sigma_2\,\Sigma_2\,\sigma_3\,\Sigma_3+y_2\,\sigma_3\,\Sigma_3\,\sigma_1\,\Sigma_1+y_3\,\sigma_1\,\Sigma_1\,\sigma_2\,\Sigma_2$$

The equations are then:

$$\dot{x}_1=\sqrt{\frac{y_2\,y_3}{y_1}}$$

$$\dot{y}_1 = \frac{m\,n\,x_1 + (m+n)\,x_2\,x_3 + x_1\,(x_2^2 + x_3^2 - x_1^2)}{\sqrt{y_1\,y_2\,y_3}}$$

EXAMPLE 2.4. In [**5**] the authors take $M^6 = S^3 \times T^3$ for which

$$H^3(M) \cong \mathbf{Z} \oplus \mathbf{Z}, \quad H^4(M) \cong \mathbf{Z}^3.$$

Left-invariance of forms gives:

$$\begin{aligned} \rho = n\,\Sigma_1\,\Sigma_2\,\Sigma_3 - m\,\sigma_1\,\sigma_2\,\sigma_3 + [&x_1\,\sigma_1\,\Sigma_2\,\Sigma_3 \\ &+ 2 \text{ cyclic terms}] \\ \sigma = y_1\,\sigma_2\,\Sigma_2\,\sigma_3\,\Sigma_3 + y_2\,&\sigma_3\,\Sigma_3\,\sigma_1\,\Sigma_1 + y_3\,\sigma_1\,\Sigma_1\,\sigma_2\,\Sigma_2 \end{aligned}$$

The equations are then:

$$\dot{x}_1 = \sqrt{\frac{y_2\,y_3}{y_1}}$$
$$\dot{y}_1 = \frac{m\,x_2\,x_3}{\sqrt{y_1\,y_2\,y_3}}$$

3. Generalized Calabi-Yau manifolds

We saw in the first Lecture that the Calabi-Yau condition could be derived by a variational approach using 3-forms in six dimensions. In this lecture we shall see how a variational approach using forms of arbitrary degree in six dimensions produces a new differential geometric structure. It is perhaps easier to introduce the structure first. This involves the *Courant bracket* on sections of $T \oplus \Lambda^p T^*$. If $X + \xi, Y + \eta \in C^\infty(T \oplus \Lambda^p T^*)$ one defines

$$[X + \xi, Y + \eta] = [X, Y] + \mathcal{L}_X \eta - \mathcal{L}_Y \xi - \frac{1}{2} d(i_X \eta - i_Y \xi)$$

Unlike the Lie bracket on vector fields this has nontrivial automorphisms defined by forms: if $\alpha \in \Omega^{p+1}$ and $d\alpha = 0$ then defining

$$A(X + \xi) = X + \xi + i_X \alpha$$

one obtains

$$A([X + \xi, Y + \eta]) = [A(X + \xi), A(Y + \eta)].$$

EXAMPLE 3.1. In the case $p = 0$ we have $X + f, Y + g \in C^\infty(T) \oplus C^\infty$ and one finds

$$[X + f, Y + g] = [X, Y] + Xg - Yf.$$

The automorphisms here are the closed 1-forms $\alpha \in \Omega^1$ where

$$A(X + f) = X + f + i_X \alpha$$

In fact this example has another interpretation as S^1-invariant vector fields on $M \times S^1$, for if

$$X + f\frac{d}{d\theta}$$

then the Lie bracket is equal to the Courant bracket. Moreover a gauge transformation $g : M \to S^1$ defined by

$$(x, e^{i\theta}) \mapsto (x, g(x)e^{i\theta})$$

gives the automorphism of invariant vector fields

$$X + f\frac{d}{d\theta} \mapsto X + (f + \frac{1}{i}i_X g^{-1}dg)\frac{d}{d\theta}$$

so that $\alpha = g^{-1}dg/2\pi i$ is the closed 1-form defining it. Note however that $\alpha \in \Omega^1$ in this case is closed with *integral periods* so that a general automorphism is a flat connection and not just one gauge-equivalent to the trivial connection.

There is in fact a more general setting for this if we replace the product $M \times S^1$ by a principal S^1-bundle P over M. In that case we have an exact sequence of vector bundles over M (the Atiyah sequence)

$$0 \to 1 \to TP/S^1 \to TM \to 0.$$

The sections of TP/S^1 are the invariant vector fields on P and so have their own Lie bracket. A connection on P is a splitting of the exact sequence: $TP/G \cong T \oplus 1$, which therefore defines a bracket on sections of $T \oplus 1$. which is

$$[X + f, Y + g] - 2F(X, Y)$$

where F is the curvature of the connection. Now we see that a non-flat connection defines a deformation of the bracket by a closed 2-form with integral periods.

It is the case $p = 1$ which will interest us:

$$X + \xi, Y + \eta \in C^\infty(T \oplus T^*).$$

In this case the automorphisms are closed 2-forms

$$B \in \Omega^2, \quad dB = 0$$

which play a role very close to that of the B-field in string theory. Concretely, the automorphism is

$$A(X + \xi) = X + \xi + i_X B.$$

We can also twist the bracket by a closed *3-form* H:

$$[X + \xi, Y + \eta] + 6i_X i_Y H.$$

When H has integral periods it brings us close to the differential geometry of gerbes, but this aspect will not be pursued here.

The vector bundle $T \oplus T^*$ has another structure defined on it apart from the Courant bracket. The natural pairing between T and T^* defines an indefinite metric on the bundle, with signature (n, n): for $X + \xi \in T \oplus T^*$ the inner product is

$$(X + \xi, X + \xi) = -\langle X, \xi\rangle.$$

By naturality, any endomorphism of T preserves the inner product, but the full orthogonal Lie algebra is

$$\operatorname{End} T \oplus \Lambda^2 T^* \oplus \Lambda^2 T.$$

In particular a 2-form B, thought of as a map from T to T^*, lies in the Lie algebra. The exponentiation of this action is

$$\exp B(X + \xi) = X + \xi + i_X B$$

which is just what we have seen in the context of the Courant bracket.

More interesting than the vector representation is the spin representation, which has a concrete form in our situation, because of the decomposition $T \oplus T^*$ into maximal isotropic subspaces. We define the two spaces

$$\begin{aligned} S^+ &= \Lambda^{ev}T^* \otimes (\Lambda^n T)^{1/2} \\ S^- &= \Lambda^{od}T^* \otimes (\Lambda^n T)^{1/2} \end{aligned}$$

and the action of $X + \xi$ by Clifford multiplication as:

$$(X + \xi) \cdot \varphi = i_X\varphi + \xi \wedge \varphi.$$

In this representation, the action of $B \in \Lambda^2 T^*$ is

$$\exp B(\varphi) = (1 + B + \frac{1}{2}B \wedge B + \dots) \wedge \varphi.$$

We can also consider the exterior forms alone as spinors, removing the line bundle twist by $(\Lambda^n T)^{1/2}$ (like $Spin^c$ structures). The spin representations S^+ and S^- have an invariant symmetric or skew-symmetric form $\langle \varphi, \psi \rangle$ defined on them. On exterior forms this takes values in the one-dimensional space $\Lambda^n T^*$. Concretely, this is just the exterior product pairing, but with an alternating sign, sometimes called in algebraic geometry the *Mukai pairing*.

Now we introduce a geometric structure compatible with this natural background. To motivate this recall one approach to the definition of a complex structure on a manifold of even dimension. An almost complex structure $J : T \to T$, with $J^2 = -1$ has a $+i$ eigenspace in $T \otimes \mathbf{C}$ which is a subbundle E satisfying the following properties:

(1) $T \otimes \mathbf{C} = E \oplus \bar{E}$
(2) the sections of E are closed under the Lie bracket.

The Newlander-Nirenberg Theorem tells us that any such subbundle actually defines an integrable complex structure. We now make a more general definition:

Definition 3.2. A **generalized complex structure** on an even-dimensional manifold is a subbundle $E \subset (T \oplus T^*) \otimes \mathbf{C}$ such that

(1) $(T \oplus T^*) \otimes \mathbf{C} = E \oplus \bar{E}$
(2) sections of E are closed under the Courant bracket
(3) E is isotropic with respect to the natural inner product on $T \oplus T^*$.

The final isotropy condition has two connotations. On the one hand it says that the almost complex structure $J : T \oplus T^* \to T \oplus T^*$ is compatible with the indefinite metric in the sense that the metric is pseudo-Hermitian. On the other, although it is easy to let the words "closed under the Courant bracket" slip off the tongue, there is a complication – the obstruction is only tensorial if E is isotropic. For example, suppose that $X + \xi, Y + \eta$ are sections of E and consider for some function f

$$[f(X + \xi), Y + \eta] = f[X + \xi, Y + \eta] + \frac{1}{2}df \wedge (i_X\eta + i_Y\xi) - (Y \cdot f)(X + \xi).$$

Since

$$\frac{1}{2}(i_X\eta + i_Y\xi) = (X + \xi, Y + \eta)$$

if E is isotropic then $i_X\eta + i_Y\xi = 0$ and the obstruction to being closed under the bracket lies in

$$\Lambda^2 E^* \otimes E^*.$$

Without the isotropy condition, the embarrassing extra term df is involved.

A generalized complex structure is a reduction of the structure group of $T \oplus T^*$ from $SO(2n, 2n)$ to $U(n, n)$ together with an integrability condition. An ordinary complex manifold is an example, where $J = I$ on T and $-I$ on T^*, but there are more as we shall see.

Isotropic subspaces have a close connection with spinors. It is easy to see that for any spinor φ the set of vectors $X + \xi$ which annihilate φ is an isotropic subspace, since Clifford multiplication has the characteristic property $(X + \xi)^2 = -(X + \xi, X + \xi)1$. A *pure* spinor $\varphi \in S^{\pm}$ is defined by the condition

$$U = \{X + \xi \in (T \oplus T^*) \otimes \mathbf{C} : (X + \xi) \cdot \varphi = 0\}$$

is maximally isotropic.

Example 3.3. We can build up examples from the simplest situation:

(1) take the pure spinor $1 \in \Lambda^0 T^*$; its maximal isotropic subspace is $U = T \subset T \oplus T^*$
(2) transform by B: one gets the pure spinor $\exp B(1) = 1 + B + \frac{1}{2}B^2 + \dots$; its maximal isotropic subspace is $U = \exp B(T)$

A generalized complex structure is thus defined in a neighbourhood of each point by a complex pure spinor field, well defined up to scalar multiplication. In our set-up such a spinor for $T \oplus T^*$ is a form of mixed degree on the manifold. In general there is a complex line bundle, analogous to the canonical bundle, for which these locally defined forms are local trivializations, but there is a situation where the pure spinor is global:

Definition 3.4. A **generalized Calabi-Yau manifold** is an even-dimensional manifold M with a *closed* form $\rho \in \Omega^{ev/od} \otimes \mathbf{C}$ which is a pure spinor and satisfies $\langle \rho, \bar{\rho} \rangle \neq 0$ at each point.

In the definition, the pure spinor ρ defines a maximal isotropic subbundle $E \subset (T \oplus T^*) \otimes \mathbf{C}$ and the condition $\langle \rho, \bar{\rho} \rangle \neq 0$ says that $E \oplus \bar{E} = (T \oplus T^*) \otimes \mathbf{C}$. One can also prove the following:

Lemma 3.5. *The condition $d\rho = 0$ for a pure spinor ρ implies that sections of the isotropic subbundle $E \subset (T \oplus T^*) \otimes \mathbf{C}$ are closed under Courant bracket.*

We see then that a generalized Calabi-Yau manifold is a particular case of a generalized complex manifold. In practical terms it is often easier to check that a form is closed rather than that a subbundle is closed under Courant bracket.

Surprisingly both complex and symplectic structures appear here, as in the following examples:

(1) a Calabi-Yau manifold with non-vanishing holomorphic n-form ρ
(2) a symplectic manifold (M, ω) with $\rho = \exp i\omega$
(3) the B-field transform of a symplectic manifold

$$\rho = \exp B \exp i\omega = \exp(B + i\omega)$$

Example 3.6. Here is the situation in dimension 2, in the compact case:

(1) *Odd type:* $\rho \in \Omega^1(M)$, $d\rho = 0$, $\rho \wedge \bar{\rho} \neq 0$. There is no purity condition here: the multiples of ρ define the $(1,0)$-forms of a complex structure, and then ρ is a non-vanishing holomorphic 1-form, so M^2 is an *elliptic curve*.

(2) *Even type:* here $\rho = c + \beta$ (c constant, β a 2-form) such that

$$\langle c+\beta, \bar{c}+\bar{\beta}\rangle = c\bar{\beta} - \bar{c}\beta \neq 0$$

This gives $c \neq 0$ and $\mathrm{Im}\,(\beta/c) \neq 0$ implies that $\beta/c = B + i\omega$ where ω is a symplectic form, so

$$\rho = c\exp(B + i\omega)$$

and M^2 is the B-field transform of a symplectic manifold.

In dimension 4, the purity condition is easy to state because of triality for $SO(4,4)$. As we noted in Lecture 1, the vector representation and the two spin representations in eight dimensions are related by an outer automorphism. In the present context this means that the internal structure of S^+ and S^- is simply that of an 8-dimensional vector space with an inner product – the spin quadratic form $\langle\varphi,\varphi\rangle$ – and a pure spinor is simply one that is null with respect to this.

EXAMPLE 3.7. Here then is the situation in dimension 4 in the compact case:

(1) *Odd type:* $\rho = \beta + \gamma \in \Omega^1 \oplus \Omega^3$, where β, γ are closed and

$$\beta \wedge \gamma = 0, \quad \beta \wedge \bar{\gamma} + \bar{\beta} \wedge \gamma \neq 0.$$

This implies that $\gamma = \beta \wedge \nu$ and β defines a foliation by surfaces with a transverse holomorphic 1-form, while $\mathrm{Im}\,(\nu)$ defines a symplectic form on the leaves. An example is $M^4 = M_1^2 \times M_2^2$, where M_1 is an elliptic curve, and M_2 a symplectic surface.

(2) *Even type:* $\rho = c + \beta + \gamma \in \Omega^0 \oplus \Omega^2 \oplus \Omega^4$ where the null (purity) condition is

$$\beta^2 = 2c\gamma$$

If $c \neq 0$, then

$$\rho = c\exp\beta = c + \beta + \frac{1}{2c}\beta^2$$

and as before, this is the B-field transform of a symplectic structure, but if $c = 0$, $\beta^2 = 0$ and is a closed, locally decomposable complex 2-form. It defines the structure of a complex surface with nonvanishing holomorphic 2-form on M, which thus must be a torus or K3 surface.

Complex structures are not the only ones one can generalize in this way. Here are some more obvious ones in 4 dimensions, determined by the number of spinors fixed by a subgroup of $Spin(4,4)$:

(1) $SU(2,2) \to SO(4,2)$: generalized Calabi-Yau structure
(2) $Sp(1,1) \to SO(4,1)$: generalized hyperkähler structure
(3) $Sp(1) \times Sp(1) \to SO(4)$: generalized hyperkähler metric

The last example applied to a K3 surface gives a geometrical structure whose moduli space is a space of $N = (4,4)$ conformal field theories [**9**].

The case of dimension 6 brings us back to the idea of open orbits of forms and invariant functionals, which is where we started. In fact, as described in [**8**], this is the context in which this author found these objects.

Note that in general the dimension of the space of complex pure spinors for $SO(2m, 2m)$ is the dimension of $\mathbf{C}\exp B$ which is $1 + m(2m-1)$. When $m = 3$, the real dimension is therefore

$$2 \times 16 = 32 = \dim S^+ = \dim S^-.$$

This numerical coincidence manifests itself in the more geometrical statement:

LEMMA 3.8. *For the spin representations of $Spin(6,6)$ in $S^\pm$ there is an invariant open set of real spinors φ such that $\varphi = \rho + \bar\rho$ where ρ is a pure spinor with $\langle \rho, \bar\rho\rangle \neq 0$.*

As a consequence of this, $\langle \rho, \bar\rho\rangle$ defines $SO(6,6)$-invariant maps

$$\phi : \Lambda^{ev/od} V^* \to \Lambda^6 V^*$$

and on a compact manifold M^6 we get a functional

$$\Phi(\varphi) = \int_M \phi(\varphi)$$

defined on even forms or odd forms, invariant under diffeomorphisms and, this time, the action of the B-field $\varphi \mapsto \exp B\varphi$.

EXAMPLE 3.9. If $\rho = \rho_0 + \rho_2 + \rho_4 + \rho_6$, then

$$\langle \rho, \bar\rho\rangle = \rho_0\bar\rho_6 - \bar\rho_0\rho_6 - \rho_2\bar\rho_4 + \bar\rho_2\rho_4.$$

Take $\varphi = 1 - \omega^2/2$, where ω is a symplectic form, then

$$\rho = \exp(i\omega) = 1 + i\omega - \frac{1}{2}\omega^2 - i\frac{1}{6}\omega^3$$

and

$$\phi(\varphi) = \frac{4}{3} i\omega^3.$$

So the functional Φ is essentially the Liouville volume of the symplectic form.

We now address our version of nonlinear Hodge theory and ask: *what is a critical point of the functional*

$$\Phi(\varphi) = \int_M \phi(\varphi)$$

for φ a closed form in a fixed cohomology class in $H^{ev/od}(M, \mathbf{R})$? Just as in Lecture 1, the critical points are forms where $\varphi = \rho + \bar\rho$ and

$$d\rho = 0$$

which means that the critical points are *generalized Calabi-Yau manifolds.*

In this situation, there is a non-degeneracy condition for the critical points which is both a generalization of the $\partial\bar\partial$ lemma and of strong Lefschetz and implies the existence of a unique nearby critical point in each nearby cohomology class, so we deduce that the *moduli space is an open set in $H^{ev/od}(M, \mathbf{R})$*. The set-up here however, is different from what we encountered before because we consider structures modulo diffeomorphism and B-fields. In fact we have to take diffeomorphisms homotopic to the identity (which is normal as in Teichmüller theory) and also cohomologically trivial B-fields – exact 2-forms.

To summarize, all the special geometries encountered above are in some way associated to open orbits of groups. Such actions have in fact been classified and in truth there are not many more, but to finish let me just point out that a real form of the exceptional group E_7 in its 56-dimensional representation shares many of the properties of the groups considered above, but I have no idea what sort of special geometry this generates or even in what dimension.

References

1. C. Bär, *Real Killing spinors and holonomy*, Comm. Math. Phys. **154** (1993), 509–521.
2. C. Boyer and K. Galicki, *3-Sasakian manifolds*, Surveys in differential geometry: essays on Einstein manifolds, Surv. Differ. Geom., VI, Int. Press, Boston, MA, 1999 pp. 123–184.
3. Z.W. Chong, M. Cvetič, G.W. Gibbons, H. Lü, C.N. Pope, P. Wagner, *General Metrics of G_2 Holonomy and Contraction Limits*, Nuclear Phys. B **620** (2002), 3–28.
4. M. Cvetič, G, Gibbons, H. Lü, C.N. Pope, *New complete noncompact Spin(7) manifolds*, Nuclear Phys. B **620** (2002), 29–54
5. S. Gukov, S.-T. Yau, E. Zaslow, *Duality and fibrations on G_2 manifolds*, Turkish J. Math. **27** (2003) 61-97
6. N. J. Hitchin, *The geometry of three-forms in six dimensions.*, J. Differential Geom. **55** (2000), 547–576.
7. N. J. Hitchin, *Stable forms and special metrics.*, Global differential geometry: the mathematical legacy of Alfred Gray (Bilbao, 2000), Contemp. Math., 288, Amer. Math. Soc., Providence, RI, 2001, pp. 70–89.
8. N. J. Hitchin, *Generalized Calabi-Yau manifolds*, Quarterly J. Math. **54** (2003) 281-308.
9. W. Nahm and K. Wendland *A hiker's guide to K3: aspects of $N = (4,4)$ superconformal field theory with central charge $c = 6$*, Comm. Math. Phys. **216** (2001) 85–138.

Mathematical Institute, 24-29 St Giles, Oxford, OX1 3LB, UK
E-mail address: hitchin@maths.ox.ac.uk

Clay Mathematics Proceedings
Volume 3, 2004

Constructing compact manifolds with exceptional holonomy

Dominic Joyce

Abstract. The exceptional cases in Berger's classification of Riemannian holonomy groups are G_2 in 7 dimensions, and Spin(7) in 8 dimensions. Metrics with these holonomy groups are Ricci-flat, and have many interesting geometrical properties. We survey the known constructions of examples of *compact* 7- and 8-manifolds with holonomy G_2 and Spin(7).

We first construct a compact, singular Riemannian orbifold X with holonomy a proper subgroup of G_2 or Spin(7), such as $X = T^7/\Gamma$ for certain finite groups Γ. Then we resolve the singularities of X using ideas from Calabi–Yau geometry and analytic techniques, to get a compact, nonsingular 7- or 8-manifold M with holonomy G_2 or Spin(7).

1. Introduction

In the theory of Riemannian holonomy groups, perhaps the most mysterious are the two exceptional cases, the holonomy group G_2 in 7 dimensions and the holonomy group Spin(7) in 8 dimensions. This is a survey paper on the construction of examples of *compact* 7- and 8-manifolds with holonomy G_2 and Spin(7).

All of the material described can be found in the author's book [**9**]. Some, but not all, is also in the papers [**5, 6, 7, 8, 10, 11**]. In particular, the most complicated and powerful form of the construction of compact manifolds with exceptional holonomy by resolving orbifolds T^n/Γ, and many of the examples, are given only in [**9**] and not in any published paper.

The rest of this section introduces the holonomy groups G_2, Spin(7) and SU(m), and the relations between them. Section 2 discusses constructions for compact 7-manifolds with holonomy G_2. Most of the section explains how to do this by resolving the singularities of orbifolds T^7/Γ, but in § 2.5 we briefly discuss two other methods starting from Calabi–Yau 3-folds.

Section 3 explains constructions of compact 8-manifolds with holonomy Spin(7). One way to do this is to resolve orbifolds T^8/Γ, but as this is very similar in outline to the G_2 material of § 2 we say little about it. Instead we describe a second construction which begins with a Calabi–Yau 4-orbifold.

2000 *Mathematics Subject Classification.* 53C29.

1.1. Riemannian holonomy groups.

Let M be a connected n-dimensional manifold, g a Riemannian metric on M, and ∇ the Levi-Civita connection of g. Let x, y be points in M joined by a smooth path γ. Then *parallel transport* along γ using ∇ defines an isometry between the tangent spaces T_xM, T_yM at x and y.

Definition 1.1. The *holonomy group* $\mathrm{Hol}(g)$ of g is the group of isometries of T_xM generated by parallel transport around piecewise-smooth closed loops based at x in M. We consider $\mathrm{Hol}(g)$ to be a subgroup of $\mathrm{O}(n)$, defined up to conjugation by elements of $\mathrm{O}(n)$. Then $\mathrm{Hol}(g)$ is independent of the base point x in M.

The classification of holonomy groups was achieved by Berger [**1**] in 1955.

Theorem 1.2. *Let M be a simply-connected, n-dimensional manifold, and g an irreducible, nonsymmetric Riemannian metric on M. Then either*

(i) $\mathrm{Hol}(g) = \mathrm{SO}(n)$,
(ii) $n = 2m$ *and* $\mathrm{Hol}(g) = \mathrm{SU}(m)$ *or* $\mathrm{U}(m)$,
(iii) $n = 4m$ *and* $\mathrm{Hol}(g) = \mathrm{Sp}(m)$ *or* $\mathrm{Sp}(m)\,\mathrm{Sp}(1)$,
(iv) $n = 7$ *and* $\mathrm{Hol}(g) = G_2$, *or*
(v) $n = 8$ *and* $\mathrm{Hol}(g) = \mathrm{Spin}(7)$.

Now G_2 and $\mathrm{Spin}(7)$ are the exceptional cases in this classification, so they are called the *exceptional holonomy groups*. For some time after Berger's classification, the exceptional holonomy groups remained a mystery. In 1987, Bryant [**2**] used the theory of exterior differential systems to show that locally there exist many metrics with these holonomy groups, and gave some explicit, incomplete examples. Then in 1989, Bryant and Salamon [**3**] found explicit, *complete* metrics with holonomy G_2 and $\mathrm{Spin}(7)$ on noncompact manifolds.

In 1994-5 the author constructed the first examples of metrics with holonomy G_2 and $\mathrm{Spin}(7)$ on *compact* manifolds [**5, 6, 7**]. These, and the more complicated constructions developed later by the author [**8, 9**] and by Kovalev [**12**], are the subject of this article.

1.2. The holonomy group G_2.

Let $(x_1, \dots, x_7)$ be coordinates on $\mathbb{R}^7$. Write $\mathrm{d}\mathbf{x}_{ij\dots l}$ for the exterior form $\mathrm{d}x_i \wedge \mathrm{d}x_j \wedge \cdots \wedge \mathrm{d}x_l$ on $\mathbb{R}^7$. Define a metric g_0, a 3-form φ_0 and a 4-form $*\varphi_0$ on $\mathbb{R}^7$ by $g_0 = \mathrm{d}x_1^2 + \cdots + \mathrm{d}x_7^2$,

$$\begin{aligned} \varphi_0 &= \mathrm{d}\mathbf{x}_{123} + \mathrm{d}\mathbf{x}_{145} + \mathrm{d}\mathbf{x}_{167} + \mathrm{d}\mathbf{x}_{246} - \mathrm{d}\mathbf{x}_{257} - \mathrm{d}\mathbf{x}_{347} - \mathrm{d}\mathbf{x}_{356} \ \text{ and} \\ *\varphi_0 &= \mathrm{d}\mathbf{x}_{4567} + \mathrm{d}\mathbf{x}_{2367} + \mathrm{d}\mathbf{x}_{2345} + \mathrm{d}\mathbf{x}_{1357} - \mathrm{d}\mathbf{x}_{1346} - \mathrm{d}\mathbf{x}_{1256} - \mathrm{d}\mathbf{x}_{1247}. \end{aligned} \tag{1}$$

The subgroup of $GL(7, \mathbb{R})$ preserving φ_0 is the *exceptional Lie group* G_2. It also preserves $g_0, *\varphi_0$ and the orientation on $\mathbb{R}^7$. It is a compact, semisimple, 14-dimensional Lie group, a subgroup of $\mathrm{SO}(7)$.

A *G_2-structure* on a 7-manifold M is a principal subbundle of the frame bundle of M, with structure group G_2. Each G_2-structure gives rise to a 3-form φ and a metric g on M, such that every tangent space of M admits an isomorphism with $\mathbb{R}^7$ identifying φ and g with φ_0 and g_0 respectively. By an abuse of notation, we will refer to (φ, g) as a G_2-structure.

Proposition 1.3. *Let M be a 7-manifold and (φ, g) a G_2-structure on M. Then the following are equivalent:*

(i) $\mathrm{Hol}(g) \subseteq G_2$, *and* φ *is the induced* 3*-form,*
(ii) $\nabla\varphi = 0$ *on* M*, where* ∇ *is the Levi-Civita connection of* g*, and*
(iii) $\mathrm{d}\varphi = \mathrm{d}^*\varphi = 0$ *on* M.

We call $\nabla\varphi$ the *torsion* of the G_2-structure (φ, g), and when $\nabla\varphi = 0$ the G_2-structure is *torsion-free*. A triple (M, φ, g) is called a G_2*-manifold* if M is a 7-manifold and (φ, g) a torsion-free G_2-structure on M. If g has holonomy $\mathrm{Hol}(g) \subseteq G_2$, then g is Ricci-flat.

THEOREM 1.4. *Let* M *be a compact* 7*-manifold, and suppose that* (φ, g) *is a torsion-free* G_2*-structure on* M*. Then* $\mathrm{Hol}(g) = G_2$ *if and only if* $\pi_1(M)$ *is finite. In this case the moduli space of metrics with holonomy* G_2 *on* M*, up to diffeomorphisms isotopic to the identity, is a smooth manifold of dimension* $b^3(M)$.

1.3. The holonomy group Spin(7)**.**

Let $\mathbb{R}^8$ have coordinates $(x_1, \ldots, x_8)$. Define a 4-form Ω_0 on $\mathbb{R}^8$ by

$$\begin{aligned}\Omega_0 = {} & \mathrm{d}\mathbf{x}_{1234} + \mathrm{d}\mathbf{x}_{1256} + \mathrm{d}\mathbf{x}_{1278} + \mathrm{d}\mathbf{x}_{1357} - \mathrm{d}\mathbf{x}_{1368} - \mathrm{d}\mathbf{x}_{1458} - \mathrm{d}\mathbf{x}_{1467} \\ & - \mathrm{d}\mathbf{x}_{2358} - \mathrm{d}\mathbf{x}_{2367} - \mathrm{d}\mathbf{x}_{2457} + \mathrm{d}\mathbf{x}_{2468} + \mathrm{d}\mathbf{x}_{3456} + \mathrm{d}\mathbf{x}_{3478} + \mathrm{d}\mathbf{x}_{5678}.\end{aligned} \tag{2}$$

The subgroup of $\mathrm{GL}(8, \mathbb{R})$ preserving Ω_0 is the holonomy group Spin(7). It also preserves the orientation on $\mathbb{R}^8$ and the Euclidean metric $g_0 = \mathrm{d}x_1^2 + \cdots + \mathrm{d}x_8^2$. It is a compact, semisimple, 21-dimensional Lie group, a subgroup of SO(8).

A Spin(7)-structure on an 8-manifold M gives rise to a 4-form Ω and a metric g on M, such that each tangent space of M admits an isomorphism with $\mathbb{R}^8$ identifying Ω and g with Ω_0 and g_0 respectively. By an abuse of notation we will refer to the pair (Ω, g) as a Spin(7)-structure.

PROPOSITION 1.5. *Let* M *be an* 8*-manifold and* (Ω, g) *a* Spin(7)*-structure on* M*. Then the following are equivalent:*

(i) $\mathrm{Hol}(g) \subseteq \mathrm{Spin}(7)$, *and* Ω *is the induced* 4*-form,*
(ii) $\nabla\Omega = 0$ *on* M*, where* ∇ *is the Levi-Civita connection of* g*, and*
(iii) $\mathrm{d}\Omega = 0$ *on* M.

We call $\nabla\Omega$ the *torsion* of the Spin(7)-structure (Ω, g), and (Ω, g) *torsion-free* if $\nabla\Omega = 0$. A triple (M, Ω, g) is called a Spin(7)*-manifold* if M is an 8-manifold and (Ω, g) a torsion-free Spin(7)-structure on M. If g has holonomy $\mathrm{Hol}(g) \subseteq \mathrm{Spin}(7)$, then g is Ricci-flat.

Here is a result on *compact* 8-manifolds with holonomy Spin(7).

THEOREM 1.6. *Let* (M, Ω, g) *be a compact* Spin(7)*-manifold. Then* $\mathrm{Hol}(g) = \mathrm{Spin}(7)$ *if and only if* M *is simply-connected, and* $b^3(M) + b^4_+(M) = b^2(M) + 2b^4_-(M) + 25$*. In this case the moduli space of metrics with holonomy* Spin(7) *on* M*, up to diffeomorphisms isotopic to the identity, is a smooth manifold of dimension* $1 + b^4_-(M)$.

1.4. The holonomy groups $\mathrm{SU}(m)$**.**

Let $\mathbb{C}^m \cong \mathbb{R}^{2m}$ have complex coordinates $(z_1, \ldots, z_m)$, and define the metric g_0, Kähler form ω_0 and complex volume form θ_0 on $\mathbb{C}^m$ by

$$\begin{gathered} g_0 = |\mathrm{d}z_1|^2 + \cdots + |\mathrm{d}z_m|^2, \quad \omega_0 = \frac{i}{2}(\mathrm{d}z_1 \wedge \mathrm{d}\bar{z}_1 + \cdots + \mathrm{d}z_m \wedge \mathrm{d}\bar{z}_m), \\ \text{and} \quad \theta_0 = \mathrm{d}z_1 \wedge \cdots \wedge \mathrm{d}z_m. \end{gathered} \tag{3}$$

The subgroup of $\mathrm{GL}(2m,\mathbb{R})$ preserving g_0, ω_0 and θ_0 is the special unitary group $\mathrm{SU}(m)$. Manifolds with holonomy $\mathrm{SU}(m)$ are called *Calabi–Yau manifolds*.

Calabi–Yau manifolds are automatically Ricci-flat and Kähler, with trivial canonical bundle. Conversely, any Ricci-flat Kähler manifold (M, J, g) with trivial canonical bundle has $\mathrm{Hol}(g) \subseteq \mathrm{SU}(m)$. By Yau's proof of the Calabi conjecture [**16**], we have:

THEOREM 1.7. *Let (M, J) be a compact complex m-manifold admitting Kähler metrics, with trivial canonical bundle. Then there is a unique Ricci-flat Kähler metric g in each Kähler class on M, and $\mathrm{Hol}(g) \subseteq \mathrm{SU}(m)$.*

Using this and complex algebraic geometry one can construct many examples of compact Calabi–Yau manifolds. The theorem also applies in the orbifold category, yielding examples of *Calabi–Yau orbifolds*.

1.5. Relations between G_2, $\mathrm{Spin}(7)$ and $\mathrm{SU}(m)$.

Here are the inclusions between the holonomy groups $\mathrm{SU}(m), G_2$ and $\mathrm{Spin}(7)$:

$$\begin{array}{ccccc} \mathrm{SU}(2) & \longrightarrow & \mathrm{SU}(3) & \longrightarrow & G_2 \\ \downarrow & & \downarrow & & \downarrow \\ \mathrm{SU}(2)\times\mathrm{SU}(2) & \longrightarrow & \mathrm{SU}(4) & \longrightarrow & \mathrm{Spin}(7). \end{array}$$

We shall illustrate what we mean by this using the inclusion $\mathrm{SU}(3) \hookrightarrow G_2$. As $\mathrm{SU}(3)$ acts on $\mathbb{C}^3$, it also acts on $\mathbb{R} \oplus \mathbb{C}^3 \cong \mathbb{R}^7$, taking the $\mathrm{SU}(3)$-action on $\mathbb{R}$ to be trivial. Thus we embed $\mathrm{SU}(3)$ as a subgroup of $\mathrm{GL}(7,\mathbb{R})$. It turns out that $\mathrm{SU}(3)$ is a subgroup of the subgroup G_2 of $\mathrm{GL}(7,\mathbb{R})$ defined in § 1.2.

Here is a way to see this in terms of differential forms. Identify $\mathbb{R} \oplus \mathbb{C}^3$ with $\mathbb{R}^7$ in the obvious way in coordinates, so that $\bigl(x_1, (x_2 + ix_3, x_4 + ix_5, x_6 + ix_7)\bigr)$ in $\mathbb{R} \oplus \mathbb{C}^3$ is identified with $(x_1, \ldots, x_7)$ in $\mathbb{R}^7$. Then $\varphi_0 = \mathrm{d}x_1 \wedge \omega_0 + \mathrm{Re}\,\theta_0$, where φ_0 is defined in (1) and ω_0, θ_0 in (3). Since $\mathrm{SU}(3)$ preserves ω_0 and θ_0, the action of $\mathrm{SU}(3)$ on $\mathbb{R}^7$ preserves φ_0, and so $\mathrm{SU}(3) \subset G_2$.

It follows that if (M, J, h) is a Calabi–Yau 3-fold, then $\mathbb{R} \times M$ and $\mathcal{S}^1 \times M$ have torsion-free G_2-structures, that is, are G_2-manifolds.

PROPOSITION 1.8. *Let (M, J, h) be a Calabi–Yau 3-fold, with Kähler form ω and complex volume form θ. Let x be a coordinate on $\mathbb{R}$ or $\mathcal{S}^1$. Define a metric $g = \mathrm{d}x^2 + h$ and a 3-form $\varphi = \mathrm{d}x \wedge \omega + \mathrm{Re}\,\theta$ on $\mathbb{R} \times M$ or $\mathcal{S}^1 \times M$. Then (g, φ) is a torsion-free G_2-structure on $\mathbb{R} \times M$ or $\mathcal{S}^1 \times M$, and $*\varphi = \frac{1}{2}\omega \wedge \omega - \mathrm{d}x \wedge \mathrm{Im}\,\theta$.*

Similarly, the inclusions $\mathrm{SU}(2) \hookrightarrow G_2$ and $\mathrm{SU}(4) \hookrightarrow \mathrm{Spin}(7)$ give:

PROPOSITION 1.9. *Let (M, J, h) be a Calabi–Yau 2-fold, with Kähler form ω and complex volume form θ. Let (x_1, x_2, x_3) be coordinates on $\mathbb{R}^3$ or T^3. Define a metric $g = \mathrm{d}x_1^2 + \mathrm{d}x_2^2 + \mathrm{d}x_3^2 + h$ and a 3-form*

$$\varphi = \mathrm{d}x_1 \wedge \mathrm{d}x_2 \wedge \mathrm{d}x_3 + \mathrm{d}x_1 \wedge \omega + \mathrm{d}x_2 \wedge \mathrm{Re}\,\theta - \mathrm{d}x_3 \wedge \mathrm{Im}\,\theta \tag{4}$$

on $\mathbb{R}^3 \times M$ or $T^3 \times M$. Then (φ, g) is a torsion-free G_2-structure on $\mathbb{R}^3 \times M$ or $T^3 \times M$, and

$$*\varphi = \tfrac{1}{2}\omega \wedge \omega + \mathrm{d}x_2 \wedge \mathrm{d}x_3 \wedge \omega - \mathrm{d}x_1 \wedge \mathrm{d}x_3 \wedge \mathrm{Re}\,\theta - \mathrm{d}x_1 \wedge \mathrm{d}x_2 \wedge \mathrm{Im}\,\theta. \tag{5}$$

PROPOSITION 1.10. *Let (M, J, g) be a Calabi–Yau 4-fold, with Kähler form ω and complex volume form θ. Define a 4-form Ω on M by $\Omega = \frac{1}{2}\omega \wedge \omega + \operatorname{Re}\theta$. Then (Ω, g) is a torsion-free* $\mathrm{Spin}(7)$*-structure on M.*

2. Constructing G_2-manifolds from orbifolds T^7/Γ

We now explain the method used in [**5, 6**] and [**9**, §11–§12] to construct examples of compact 7-manifolds with holonomy G_2. It is based on the *Kummer construction* for Calabi–Yau metrics on the $K3$ surface, and may be divided into four steps.

Step 1. Let T^7 be the 7-torus and (φ_0, g_0) a flat G_2-structure on T^7. Choose a finite group Γ of isometries of T^7 preserving (φ_0, g_0). Then the quotient T^7/Γ is a singular, compact 7-manifold, an *orbifold*.

Step 2. For certain special groups Γ there is a method to resolve the singularities of T^7/Γ in a natural way, using complex geometry. We get a nonsingular, compact 7-manifold M, together with a map $\pi : M \to T^7/\Gamma$, the resolving map.

Step 3. On M, we explicitly write down a 1-parameter family of G_2-structures (φ_t, g_t) depending on $t \in (0, \epsilon)$. They are not torsion-free, but have small torsion when t is small. As $t \to 0$, the G_2-structure (φ_t, g_t) converges to the singular G_2-structure $\pi^*(\varphi_0, g_0)$.

Step 4. We prove using analysis that for sufficiently small t, the G_2-structure (φ_t, g_t) on M, with small torsion, can be deformed to a G_2-structure $(\tilde\varphi_t, \tilde g_t)$, with zero torsion. Finally, we show that $\tilde g_t$ is a metric with holonomy G_2 on the compact 7-manifold M.

We will now explain each step in greater detail.

2.1. Step 1: Choosing an orbifold.

Let (φ_0, g_0) be the Euclidean G_2-structure on $\mathbb{R}^7$ defined in § 1.2. Suppose Λ is a *lattice* in $\mathbb{R}^7$, that is, a discrete additive subgroup isomorphic to $\mathbb{Z}^7$. Then $\mathbb{R}^7/\Lambda$ is the torus T^7, and (φ_0, g_0) pushes down to a torsion-free G_2-structure on T^7. We must choose a finite group Γ acting on T^7 preserving (φ_0, g_0). That is, the elements of Γ are the push-forwards to T^7/Λ of affine transformations of $\mathbb{R}^7$ which fix (φ_0, g_0), and take Λ to itself under conjugation.

Here is an example of a suitable group Γ, taken from [**9**, §12.2].

EXAMPLE 2.1. Let $(x_1, \ldots, x_7)$ be coordinates on $T^7 = \mathbb{R}^7/\mathbb{Z}^7$, where $x_i \in \mathbb{R}/\mathbb{Z}$. Let (φ_0, g_0) be the flat G_2-structure on T^7 defined by (1). Let α, β and γ be the involutions of T^7 defined by

$$\alpha : (x_1, \ldots, x_7) \longmapsto (x_1, x_2, x_3, -x_4, -x_5, -x_6, -x_7), \tag{6}$$

$$\beta : (x_1, \ldots, x_7) \longmapsto (x_1, -x_2, -x_3, x_4, x_5, \tfrac{1}{2} - x_6, -x_7), \tag{7}$$

$$\gamma : (x_1, \ldots, x_7) \longmapsto \bigl(-x_1, x_2, -x_3, x_4, \tfrac{1}{2} - x_5, x_6, \tfrac{1}{2} - x_7). \tag{8}$$

By inspection, α, β and γ preserve (φ_0, g_0), because of the careful choice of exactly which signs to change. Also, $\alpha^2 = \beta^2 = \gamma^2 = 1$, and α, β and γ commute. Thus they generate a group $\Gamma = \langle \alpha, \beta, \gamma \rangle \cong \mathbb{Z}_2^3$ of isometries of T^7 preserving the flat G_2-structure (φ_0, g_0).

Having chosen a lattice Λ and finite group Γ, the quotient T^7/Γ is an *orbifold*, a singular manifold with only quotient singularities. The singularities of T^7/Γ come from the fixed points of non-identity elements of Γ. We now describe the singularities in our example.

LEMMA 2.2. *In Example* 2.1, *$\beta\gamma, \gamma\alpha, \alpha\beta$ and $\alpha\beta\gamma$ have no fixed points on T^7. The fixed points of α, β, γ are each* 16 *copies of T^3. The singular set S of T^7/Γ is a disjoint union of* 12 *copies of T^3,* 4 *copies from each of α, β, γ. Each component of S is a singularity modelled on that of $T^3 \times \mathbb{C}^2/\{\pm 1\}$.*

The most important consideration in choosing Γ is that we should be able to resolve the singularities of T^7/Γ within holonomy G_2. We will explain how to do this next.

2.2. Step 2: Resolving the singularities.

Our goal is to resolve the singular set S of T^7/Γ to get a compact 7-manifold M with holonomy G_2. How can we do this? In general we cannot, because we have no idea of how to resolve general orbifold singularities with holonomy G_2. However, suppose we can arrange that every connected component of S is locally isomorphic to either

(a) $T^3 \times \mathbb{C}^2/G$, for G a finite subgroup of SU(2), or
(b) $\mathcal{S}^1 \times \mathbb{C}^3/G$, for G a finite subgroup of SU(3) acting freely on $\mathbb{C}^3 \setminus \{0\}$.

One can use complex algebraic geometry to find a *crepant resolution* X of $\mathbb{C}^2/G$ or Y of $\mathbb{C}^3/G$. Then $T^3 \times X$ or $\mathcal{S}^1 \times Y$ gives a local model for how to resolve the corresponding component of S in T^7/Γ. Thus we construct a nonsingular, compact 7-manifold M by using the patches $T^3 \times X$ or $\mathcal{S}^1 \times Y$ to repair the singularities of T^7/Γ. In the case of Example 2.1, this means gluing 12 copies of $T^3 \times X$ into T^7/Γ, where X is the blow-up of $\mathbb{C}^2/\{\pm 1\}$ at its singular point.

Now the point of using crepant resolutions is this. In both case (a) and (b), there exists a Calabi–Yau metric on X or Y which is asymptotic to the flat Euclidean metric on $\mathbb{C}^2/G$ or $\mathbb{C}^3/G$. Such metrics are called *Asymptotically Locally Euclidean (ALE)*. In case (a), the ALE Calabi–Yau metrics were classified by Kronheimer [**13, 14**], and exist for all finite $G \subset$ SU(2). In case (b), crepant resolutions of $\mathbb{C}^3/G$ exist for all finite $G \subset$ SU(3) by Roan [**15**], and the author [**10**], [**9**, §8] proved that they carry ALE Calabi–Yau metrics, using a noncompact version of the Calabi Conjecture.

By Propositions 1.8 and 1.9, we can use the Calabi–Yau metrics on X or Y to construct a torsion-free G_2-structure on $T^3 \times X$ or $\mathcal{S}^1 \times Y$. This gives a local model for how to resolve the singularity $T^3 \times \mathbb{C}^2/G$ or $\mathcal{S}^1 \times \mathbb{C}^3/G$ with holonomy G_2. So, this method gives not only a way to smooth out the singularities of T^7/Γ as a manifold, but also a family of torsion-free G_2-structures on the resolution which shows how to smooth out the singularities of the G_2-structure.

The requirement above that S be divided into connected components of the form (a) and (b) is in fact unnecessarily restrictive. There is a more complicated and powerful method, described in [**9**, §11–§12], for resolving singularities of a more general kind. We require only that the singularities should *locally* be of the form $\mathbb{R}^3 \times \mathbb{C}^2/G$ or $\mathbb{R} \times \mathbb{C}^3/G$, for G a finite subgroup of SU(2) or SU(3), and when $G \subset$ SU(3) we do *not* require that G act freely on $\mathbb{C}^3 \setminus \{0\}$.

If X is a crepant resolution of $\mathbb{C}^3/G$, where G does not act freely on $\mathbb{C}^3 \setminus \{0\}$, then the author shows [**9**, §9], [**11**] that X carries a family of Calabi–Yau metrics satisfying a complicated asymptotic condition at infinity, called *Quasi-ALE* metrics. These yield the local models necessary to resolve singularities locally of the form $\mathbb{R}\times\mathbb{C}^3/G$ with holonomy G_2. Using this method we can resolve many orbifolds T^7/Γ, and prove the existence of large numbers of compact 7-manifolds with holonomy G_2.

2.3. Step 3: Finding G_2-structures with small torsion.

For each resolution X of $\mathbb{C}^2/G$ in case (a), and Y of $\mathbb{C}^3/G$ in case (b) above, we can find a 1-parameter family $\{h_t : t > 0\}$ of metrics with the properties

(a) h_t is a Kähler metric on X with $\mathrm{Hol}(h_t) = \mathrm{SU}(2)$. Its injectivity radius satisfies $\delta(h_t) = O(t)$, its Riemann curvature satisfies $\|R(h_t)\|_{C^0} = O(t^{-2})$, and $h_t = h + O(t^4 r^{-4})$ for large r, where h is the Euclidean metric on $\mathbb{C}^2/G$, and r the distance from the origin.
(b) h_t is Kähler on Y with $\mathrm{Hol}(h_t) = \mathrm{SU}(3)$, where $\delta(h_t) = O(t)$, $\|R(h_t)\|_{C^0} = O(t^{-2})$, and $h_t = h + O(t^6 r^{-6})$ for large r.

In fact we can choose h_t to be isometric to $t^2 h_1$, and the properties above are easy to prove.

Suppose one of the components of the singular set S of T^7/Γ is locally modelled on $T^3 \times \mathbb{C}^2/G$. Then T^3 has a natural flat metric h_{T^3}. Let X be the crepant resolution of $\mathbb{C}^2/G$ and let $\{h_t : t > 0\}$ satisfy property (a). Then Proposition 1.9 gives a 1-parameter family of torsion-free G_2-structures $(\hat\varphi_t, \hat g_t)$ on $T^3 \times X$ with $\hat g_t = h_{T^3} + h_t$. Similarly, if a component of S is modelled on $\mathcal{S}^1 \times \mathbb{C}^3/G$, using Proposition 1.8 we get a family of torsion-free G_2-structures $(\hat\varphi_t, \hat g_t)$ on $\mathcal{S}^1 \times Y$.

The idea is to make a G_2-structure (φ_t, g_t) on M by gluing together the torsion-free G_2-structures $(\hat\varphi_t, \hat g_t)$ on the patches $T^3\times X$ and $\mathcal{S}^1\times Y$, and (φ_0, g_0) on T^7/Γ. The gluing is done using a partition of unity. Naturally, the first derivative of the partition of unity introduces 'errors', so that (φ_t, g_t) is not torsion-free. The size of the torsion $\nabla\varphi_t$ depends on the difference $\hat\varphi_t - \varphi_0$ in the region where the partition of unity changes. On the patches $T^3\times X$, since $h_t - h = O(t^4 r^{-4})$ and the partition of unity has nonzero derivative when $r = O(1)$, we find that $\nabla\varphi_t = O(t^4)$. Similarly $\nabla\varphi_t = O(t^6)$ on the patches $\mathcal{S}^1 \times Y$, and so $\nabla\varphi_t = O(t^4)$ on M.

For small t, the dominant contributions to the injectivity radius $\delta(g_t)$ and Riemann curvature $R(g_t)$ are made by those of the metrics h_t on X and Y, so we expect $\delta(g_t) = O(t)$ and $\|R(g_t)\|_{C^0} = O(t^{-2})$ by properties (a) and (b) above. In this way we prove the following result [**9**, Th. 11.5.7], which gives the estimates on (φ_t, g_t) that we need.

THEOREM 2.3. *On the compact 7-manifold M described above, and on many other 7-manifolds constructed in a similar fashion, one can write down the following data explicitly in coordinates:*

- *Positive constants A_1, A_2, A_3 and ϵ,*
- *A G_2-structure (φ_t, g_t) on M with $\mathrm{d}\varphi_t = 0$ for each $t \in (0,\epsilon)$, and*
- *A 3-form ψ_t on M with $\mathrm{d}^*\psi_t = \mathrm{d}^*\varphi_t$ for each $t \in (0,\epsilon)$.*

These satisfy three conditions:

(i) $\|\psi_t\|_{L^2} \leqslant A_1 t^4$, $\|\psi_t\|_{C^0} \leqslant A_1 t^3$ *and* $\|\mathrm{d}^*\psi_t\|_{L^{14}} \leqslant A_1 t^{16/7}$,
(ii) *the injectivity radius $\delta(g_t)$ satisfies $\delta(g_t) \geqslant A_2 t$,*
(iii) *the Riemann curvature $R(g_t)$ of g_t satisfies $\|R(g_t)\|_{C^0} \leqslant A_3 t^{-2}$.*

Here the operator d^* and the norms $\|\,.\,\|_{L^2}$, $\|\,.\,\|_{L^{14}}$ and $\|\,.\,\|_{C^0}$ depend on g_t.

One should regard ψ_t as a *first integral* of the torsion $\nabla\varphi_t$ of (φ_t, g_t). Thus the norms $\|\psi_t\|_{L^2}$, $\|\psi_t\|_{C^0}$ and $\|\mathrm{d}^*\psi_t\|_{L^{14}}$ are measures of $\nabla\varphi_t$. So parts (i)–(iii) say that the torsion $\nabla\varphi_t$ must be small compared to the injectivity radius and Riemann curvature of (M, g_t).

2.4. Step 4: Deforming to a torsion-free G_2-structure.

We prove the following analysis result.

THEOREM 2.4. *Let A_1, A_2, A_3 be positive constants. Then there exist positive constants κ, K such that whenever $0 < t \leqslant \kappa$, the following is true.*

Let M be a compact 7-manifold, and (φ, g) a G_2-structure on M with $\mathrm{d}\varphi = 0$. Suppose ψ is a smooth 3-form on M with $\mathrm{d}^*\psi = \mathrm{d}^*\varphi$, and

(i) $\|\psi\|_{L^2} \leqslant A_1 t^4$, $\|\psi\|_{C^0} \leqslant A_1 t^{1/2}$ *and* $\|\mathrm{d}^*\psi\|_{L^{14}} \leqslant A_1$,
(ii) *the injectivity radius $\delta(g)$ satisfies $\delta(g) \geqslant A_2 t$, and*
(iii) *the Riemann curvature $R(g)$ satisfies $\big\|R(g)\big\|_{C^0} \leqslant A_3 t^{-2}$.*

Then there exists a smooth, torsion-free G_2-structure $(\tilde\varphi, \tilde g)$ on M with $\|\tilde\varphi - \varphi\|_{C^0} \leqslant K t^{1/2}$.

Basically, this result says that if (φ, g) is a G_2-structure on M, and the torsion $\nabla\varphi$ is sufficiently small, then we can deform to a nearby G_2-structure $(\tilde\varphi, \tilde g)$ that is torsion-free. Here is a sketch of the proof of Theorem 2.4, ignoring several technical points. The proof is that given in [**9**, §11.6–§11.8], which is an improved version of the proof in [**5**].

We have a 3-form φ with $\mathrm{d}\varphi = 0$ and $\mathrm{d}^*\varphi = \mathrm{d}^*\psi$ for small ψ, and we wish to construct a nearby 3-form $\tilde\varphi$ with $\mathrm{d}\tilde\varphi = 0$ and $\mathrm{d}^*\tilde\varphi = 0$. Set $\tilde\varphi = \varphi + \mathrm{d}\eta$, where η is a small 2-form. Then η must satisfy a nonlinear p.d.e., which we write as

$$\mathrm{d}^*\mathrm{d}\eta = -\mathrm{d}^*\psi + \mathrm{d}^*F(\mathrm{d}\eta), \tag{9}$$

where F is nonlinear, satisfying $F(\mathrm{d}\eta) = O\big(|\mathrm{d}\eta|^2\big)$.

We solve (9) by iteration, introducing a sequence $(\eta_j)_{j=0}^{\infty}$ with $\eta_0 = 0$, satisfying the inductive equations

$$\mathrm{d}^*\mathrm{d}\eta_{j+1} = -\mathrm{d}^*\psi + \mathrm{d}^*F(\mathrm{d}\eta_j), \qquad \mathrm{d}^*\eta_{j+1} = 0. \tag{10}$$

If such a sequence exists and converges to η, then taking the limit in (10) shows that η satisfies (9), giving us the solution we want.

The key to proving this is an *inductive estimate* on the sequence $(\eta_j)_{j=0}^{\infty}$. The inductive estimate we use has three ingredients, the equations

$$\|\mathrm{d}\eta_{j+1}\|_{L^2} \leqslant \|\psi\|_{L^2} + C_1\|\mathrm{d}\eta_j\|_{L^2}\|\mathrm{d}\eta_j\|_{C^0}, \tag{11}$$

$$\|\nabla\mathrm{d}\eta_{j+1}\|_{L^{14}} \leqslant C_2\big(\|\mathrm{d}^*\psi\|_{L^{14}} + \|\nabla\mathrm{d}\eta_j\|_{L^{14}}\|\mathrm{d}\eta_j\|_{C^0} + t^{-4}\|\mathrm{d}\eta_{j+1}\|_{L^2}\big), \tag{12}$$

$$\|\mathrm{d}\eta_j\|_{C^0} \leqslant C_3\big(t^{1/2}\|\nabla\mathrm{d}\eta_j\|_{L^{14}} + t^{-7/2}\|\mathrm{d}\eta_j\|_{L^2}\big). \tag{13}$$

Here C_1, C_2, C_3 are positive constants independent of t. Equation (11) is obtained from (10) by taking the L^2-inner product with η_{j+1} and integrating by parts. Using the fact that $\mathrm{d}^*\varphi = \mathrm{d}^*\psi$ and $\|\psi\|_{L^2} = O(t^4)$, $|\psi| = O(t^{1/2})$ we get a powerful estimate of the L^2-norm of $\mathrm{d}\eta_{j+1}$.

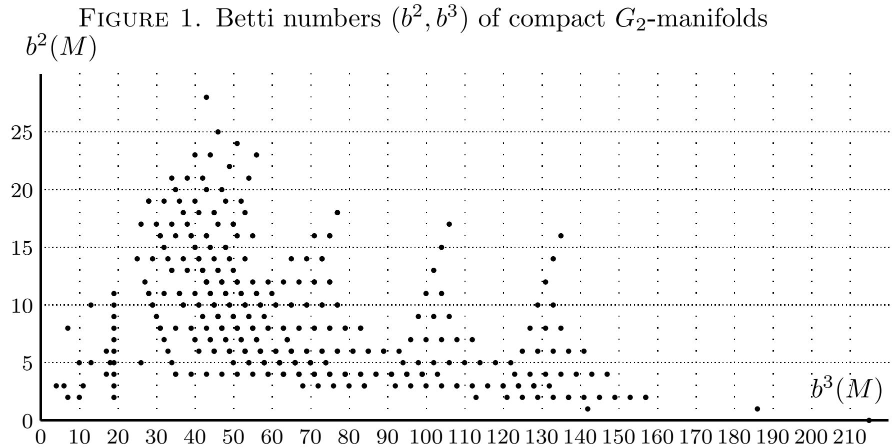
Equation (12) is derived from an *elliptic regularity estimate* for the operator $\mathrm{d} + \mathrm{d}^*$ acting on 3-forms on M. Equation (13) follows from the *Sobolev embedding theorem*, since $L^{14}_1(M)$ embeds in $C^0(M)$. Both (12) and (13) are proved on small balls of radius $O(t)$ in M, using parts (ii) and (iii) of Theorem 2.3, and this is where the powers of t come from.

Using (11)-(13) and part (i) of Theorem 2.3 we show that if

$$\|\mathrm{d}\eta_j\|_{L^2} \leqslant C_4 t^4, \quad \|\nabla \mathrm{d}\eta_j\|_{L^{14}} \leqslant C_5, \quad \text{and} \quad \|\mathrm{d}\eta_j\|_{C^0} \leqslant K t^{1/2}, \tag{14}$$

where C_4, C_5 and K are positive constants depending on C_1, C_2, C_3 and A_1, and if t is sufficiently small, then the same inequalities (14) apply to $\mathrm{d}\eta_{j+1}$. Since $\eta_0 = 0$, by induction (14) applies for all j and the sequence $(\mathrm{d}\eta_j)_{j=0}^{\infty}$ is bounded in the Banach space $L^{14}_1(\Lambda^3 T^* M)$. One can then use standard techniques in analysis to prove that this sequence converges to a smooth limit $\mathrm{d}\eta$. This concludes the proof of Theorem 2.4.

FIGURE 1. Betti numbers (b^2, b^3) of compact G_2-manifolds



From Theorems 2.3 and 2.4 we see that the compact 7-manifold M constructed in Step 2 admits torsion-free G_2-structures $(\tilde{\varphi}, \tilde{g})$. Theorem 1.4 then shows that $\mathrm{Hol}(\tilde{g}) = G_2$ if and only if $\pi_1(M)$ is finite. In the example above M is simply-connected, and so $\pi_1(M) = \{1\}$ and M has metrics with holonomy G_2, as we want.

By considering different groups Γ acting on T^7, and also by finding topologically distinct resolutions $M_1, \ldots, M_k$ of the same orbifold T^7/Γ, we can construct many compact Riemannian 7-manifolds with holonomy G_2. A good number of examples are given in [**9**, §12]. Figure 1 displays the Betti numbers of compact, simply-connected 7-manifolds with holonomy G_2 constructed there. There are 252 different sets of Betti numbers.

Examples are also known [**9**, §12.4] of compact 7-manifolds with holonomy G_2 with finite, nontrivial fundamental group. It seems likely to the author that the Betti numbers given in Figure 1 are only a small proportion of the Betti numbers of all compact, simply-connected 7-manifolds with holonomy G_2.

2.5. Other constructions of compact G_2-manifolds.

Here are two other methods, taken from [**9**, §11.9], that may be used to construct compact 7-manifolds with holonomy G_2. The first was outlined in [**6**, §4.3].

Method 1. Let (Y, J, h) be a Calabi–Yau 3-fold, with Kähler form ω and holomorphic volume form θ. Suppose $\sigma : Y \to Y$ is an involution, satisfying $\sigma^*(h) = h$, $\sigma^*(J) = -J$ and $\sigma^*(\theta) = \bar{\theta}$. We call σ a *real structure* on Y. Let N be the fixed point set of σ in Y. Then N is a real 3-dimensional submanifold of Y, and is in fact a special Lagrangian 3-fold.

Let $\mathcal{S}^1 = \mathbb{R}/\mathbb{Z}$, and define a torsion-free G_2-structure (φ, g) on $\mathcal{S}^1 \times Y$ as in Proposition 1.8. Then $\varphi = \mathrm{d}x \wedge \omega + \mathrm{Re}\,\theta$, where $x \in \mathbb{R}/\mathbb{Z}$ is the coordinate on $\mathcal{S}^1$. Define $\hat\sigma : \mathcal{S}^1 \times Y \to \mathcal{S}^1 \times Y$ by $\hat\sigma\big((x,y)\big) = \big(-x, \sigma(y)\big)$. Then $\hat\sigma$ preserves (φ, g) and $\hat\sigma^2 = 1$. The fixed points of $\hat\sigma$ in $\mathcal{S}^1 \times Y$ are $\{\mathbb{Z}, \frac{1}{2} + \mathbb{Z}\} \times N$. Thus $(\mathcal{S}^1 \times Y)/\langle\hat\sigma\rangle$ is an orbifold. Its singular set is 2 copies of N, and each singular point is modelled on $\mathbb{R}^3 \times \mathbb{R}^4/\{\pm 1\}$.

We aim to resolve $(\mathcal{S}^1 \times Y)/\langle\hat\sigma\rangle$ to get a compact 7-manifold M with holonomy G_2. Locally, each singular point should be resolved like $\mathbb{R}^3 \times X$, where X is an ALE Calabi–Yau 2-fold asymptotic to $\mathbb{C}^2/\{\pm 1\}$. There is a 3-dimensional family of such X, and we need to choose one member of this family for each singular point in the singular set.

Calculations by the author indicate that the data needed to do this is a closed, coclosed 1-form α on N that is nonzero at every point of N. The existence of a suitable 1-form α depends on the metric on N, which is the restriction of the metric g on Y. But g comes from the solution of the Calabi Conjecture, so we know little about it. This may make the method difficult to apply in practice.

The second method is studied by Alexei Kovalev [**12**], and is based on an idea due to Simon Donaldson.

Method 2. Let X be a projective complex 3-fold with canonical bundle K_X, and s a holomorphic section of K_X^{-1} which vanishes to order 1 on a smooth divisor D in X. Then D has trivial canonical bundle, so D is T^4 or $K3$. Suppose D is a $K3$ surface. Define $Y = X \setminus D$, and suppose Y is simply-connected.

Then Y is a noncompact complex 3-fold with K_Y trivial, and one infinite end modelled on $D \times \mathcal{S}^1 \times [0, \infty)$. Using a version of the proof of the Calabi Conjecture for noncompact manifolds one constructs a complete Calabi–Yau metric h on Y, which is asymptotic to the product on $D \times \mathcal{S}^1 \times [0,\infty)$ of a Calabi–Yau metric on D, and Euclidean metrics on $\mathcal{S}^1$ and $[0,\infty)$. We call such metrics *Asymptotically Cylindrical*.

Suppose we have such a metric on Y. Define a torsion-free G_2-structure (φ, g) on $\mathcal{S}^1 \times Y$ as in Proposition 1.8. Then $\mathcal{S}^1 \times Y$ is a noncompact G_2-manifold with one end modelled on $D \times T^2 \times [0,\infty)$, whose metric is asymptotic to the product on $D \times T^2 \times [0,\infty)$ of a Calabi–Yau metric on D, and Euclidean metrics on T^2 and $[0,\infty)$.

Donaldson and Kovalev's idea is to take two such products $\mathcal{S}^1 \times Y_1$ and $\mathcal{S}^1 \times Y_2$ whose infinite ends are isomorphic in a suitable way, and glue them together to get a compact 7-manifold M with holonomy G_2. The gluing process swaps round the $\mathcal{S}^1$ factors. That is, the $\mathcal{S}^1$ factor in $\mathcal{S}^1 \times Y_1$ is identified with the asymptotic $\mathcal{S}^1$ factor in $Y_2 \sim D_2 \times \mathcal{S}^1 \times [0,\infty)$, and vice versa.

3. Compact Spin(7)-manifolds from Calabi–Yau 4-orbifolds

In a very similar way to the G_2 case, one can construct compact 8-manifolds with holonomy Spin(7) by resolving the singularities of torus orbifolds T^8/Γ. This is done in [**7**] and [**9**, §13–§14]. In [**9**, §14], examples are constructed realizing 181 different sets of Betti numbers. Two compact 8-manifolds with holonomy Spin(7) and the same Betti numbers may be distinguished by the cup products on their cohomologies (examples of this are given in [**7**, §3.4]), so they probably represent rather more than 181 topologically distinct 8-manifolds.

The main differences from the G_2 case are, firstly, that the technical details of the analysis are different and harder, and secondly, that the singularities that arise are typically more complicated and more tricky to resolve. One reason for this is that in the G_2 case the singular set is made up of 1 and 3-dimensional pieces in a 7-dimensional space, so one can often arrange for the pieces to avoid each other, and resolve them independently.

But in the Spin(7) case the singular set is typically made up of 4-dimensional pieces in an 8-dimensional space, so they nearly always intersect. There are also topological constraints arising from the $\hat{A}$-genus, which do not apply in the G_2 case. The moral appears to be that when you increase the dimension, things become more difficult.

Anyway, we will not discuss this further, as the principles are very similar to the G_2 case above. Instead, we will discuss an entirely different construction of compact 8-manifolds with holonomy Spin(7) developed by the author in [**8**] and [**9**, §15], a little like Method 1 of § 2.5. In this we start from a *Calabi–Yau* 4*-orbifold* rather than from T^8. The construction can be divided into five steps.

Step 1. Find a compact, complex 4-orbifold (Y, J) satisfying the conditions:
- (a) Y has only finitely many singular points $p_1, \dots, p_k$, for $k \geqslant 1$.
- (b) Y is modelled on $\mathbb{C}^4/\langle i\rangle$ near each p_j, where i acts on $\mathbb{C}^4$ by complex multiplication.
- (c) There exists an antiholomorphic involution $\sigma : Y \to Y$ whose fixed point set is $\{p_1, \dots, p_k\}$.
- (d) $Y \setminus \{p_1, \dots, p_k\}$ is simply-connected, and $h^{2,0}(Y) = 0$.

Step 2. Choose a σ-invariant Kähler class on Y. Then by Theorem 1.7 there exists a unique σ-invariant Ricci-flat Kähler metric g in this Kähler class. Let ω be the Kähler form of g. Let θ be a holomorphic volume form for (Y, J, g). By multiplying θ by $\mathrm{e}^{i\phi}$ if necessary, we can arrange that $\sigma^*(\theta) = \bar{\theta}$.

Define $\Omega = \frac{1}{2}\omega \wedge \omega + \operatorname{Re}\theta$. Then (Ω, g) is a torsion-free Spin(7)-structure on Y, by Proposition 1.10. Also, (Ω, g) is σ-invariant, as $\sigma^*(\omega) = -\omega$ and $\sigma^*(\theta) = \bar{\theta}$. Define $Z = Y/\langle\sigma\rangle$. Then Z is a compact real 8-orbifold with isolated singular points $p_1, \dots, p_k$, and (Ω, g) pushes down to a torsion-free Spin(7)-structure (Ω, g) on Z.

Step 3. Z is modelled on $\mathbb{R}^8/G$ near each p_j, where G is a certain finite subgroup of Spin(7) with $|G| = 8$. We can write down two explicit, topologically distinct ALE Spin(7)-manifolds X_1, X_2 asymptotic to $\mathbb{R}^8/G$. Each carries a 1-parameter family of homothetic ALE metrics h_t for $t > 0$ with $\mathrm{Hol}(h_t) = \mathbb{Z}_2 \ltimes \mathrm{SU}(4) \subset \mathrm{Spin}(7)$.

For $j = 1, \dots, k$ we choose $i_j = 1$ or 2, and resolve the singularities of Z by gluing in X_{i_j} at the singular point p_j for $j = 1, \dots, k$, to get a compact, nonsingular 8-manifold M, with projection $\pi : M \to Z$.

Step 4. On M, we explicitly write down a 1-parameter family of Spin(7)-structures (Ω_t, g_t) depending on $t \in (0, \epsilon)$. They are not torsion-free, but have small torsion when t is small. As $t \to 0$, the Spin(7)-structure (Ω_t, g_t) converges to the singular Spin(7)-structure $\pi^*(\Omega_0, g_0)$.

Step 5. We prove using analysis that for sufficiently small t, the Spin(7)-structure (Ω_t, g_t) on M, with small torsion, can be deformed to a Spin(7)-structure $(\tilde{\Omega}_t, \tilde{g}_t)$, with zero torsion.

It turns out that if $i_j = 1$ for $j = 1, \ldots, k$ we have $\pi_1(M) \cong \mathbb{Z}_2$ and $\mathrm{Hol}(\tilde{g}_t) = \mathbb{Z}_2 \ltimes \mathrm{SU}(4)$, and for the other $2^k - 1$ choices of $i_1, \ldots, i_k$ we have $\pi_1(M) = \{1\}$ and $\mathrm{Hol}(\tilde{g}_t) = \mathrm{Spin}(7)$. So $\tilde{g}_t$ is a metric with holonomy Spin(7) on the compact 8-manifold M for $(i_1, \ldots, i_k) \neq (1, \ldots, 1)$.

Once we have completed Step 1, Step 2 is immediate. Steps 4 and 5 are analogous to Steps 3 and 4 of § 2, and can be done using the techniques and analytic results developed by the author for the first T^8/Γ construction of compact Spin(7)-manifolds, [**7**], [**9**, §13]. So the really new material is in Steps 1 and 3, and we will discuss only these.

3.1. Step 1: An example.

We do Step 1 using complex algebraic geometry. The problem is that conditions (a)–(d) above are very restrictive, so it is not that easy to find *any* Y satisfying all four conditions. All the examples Y the author has found are constructed using *weighted projective spaces*, an important class of complex orbifolds.

Definition 3.1. Let $m \geqslant 1$ be an integer, and $a_0, a_1, \ldots, a_m$ positive integers with highest common factor 1. Let $\mathbb{C}^{m+1}$ have complex coordinates $(z_0, \ldots, z_m)$, and define an action of the complex Lie group $\mathbb{C}^*$ on $\mathbb{C}^{m+1}$ by

$$(z_0, \ldots, z_m) \stackrel{u}{\longmapsto} (u^{a_0} z_0, \ldots, u^{a_m} z_m), \qquad \text{for } u \in \mathbb{C}^*.$$

The *weighted projective space* $\mathbb{CP}^m_{a_0,\ldots,a_m}$ is $\bigl(\mathbb{C}^{m+1} \setminus \{0\}\bigr)/\mathbb{C}^*$. The $\mathbb{C}^*$-orbit of $(z_0, \ldots, z_m)$ is written $[z_0, \ldots, z_m]$.

Here is the simplest example the author knows.

Example 3.2. Let Y be the hypersurface of degree 12 in $\mathbb{CP}^5_{1,1,1,1,4,4}$ given by

$$Y = \bigl\{[z_0, \ldots, z_5] \in \mathbb{CP}^5_{1,1,1,1,4,4} : z_0^{12} + z_1^{12} + z_2^{12} + z_3^{12} + z_4^3 + z_5^3 = 0\bigr\}.$$

Calculation shows that Y has trivial canonical bundle and three singular points $p_1 = [0,0,0,0,1,-1]$, $p_2 = [0,0,0,0,1,e^{\pi i/3}]$ and $p_3 = [0,0,0,0,1,e^{-\pi i/3}]$, all modelled on $\mathbb{C}^4/\langle i \rangle$.

Now define a map $\sigma : Y \to Y$ by

$$\sigma : [z_0, \ldots, z_5] \longmapsto [\bar{z}_1, -\bar{z}_0, \bar{z}_3, -\bar{z}_2, \bar{z}_5, \bar{z}_4].$$

Note that $\sigma^2 = 1$, though this is not immediately obvious, because of the geometry of $\mathbb{CP}^5_{1,1,1,1,4,4}$. It can be shown that conditions (a)–(d) of Step 1 above hold for Y and σ.

More suitable 4-folds Y may be found by taking hypersurfaces or complete intersections in other weighted projective spaces, possibly also dividing by a finite group, and then doing a crepant resolution to get rid of any singularities that we don't want. Examples are given in [**8**], [**9**, §15].

3.2. Step 3: Resolving $\mathbb{R}^8/G$.

Define $\alpha, \beta : \mathbb{R}^8 \to \mathbb{R}^8$ by

$$\begin{aligned}
\alpha &: (x_1, \ldots, x_8) \longmapsto (-x_2, x_1, -x_4, x_3, -x_6, x_5, -x_8, x_7), \\
\beta &: (x_1, \ldots, x_8) \longmapsto (x_3, -x_4, -x_1, x_2, x_7, -x_8, -x_5, x_6).
\end{aligned}$$

Then α, β preserve Ω_0 given in (2), so they lie in Spin(7). Also $\alpha^4 = \beta^4 = 1$, $\alpha^2 = \beta^2$ and $\alpha\beta = \beta\alpha^3$. Let $G = \langle \alpha, \beta \rangle$. Then G is a finite non-Abelian subgroup of Spin(7) of order 8, which acts freely on $\mathbb{R}^8 \setminus \{0\}$. One can show that if Z is the compact Spin(7)-orbifold constructed in Step 2 above, then $T_{p_j} Z$ is isomorphic to $\mathbb{R}^8/G$ for $j = 1, \dots, k$, with an isomorphism identifying the Spin(7)-structures (Ω, g) on Z and (Ω_0, g_0) on $\mathbb{R}^8/G$, such that β corresponds to the σ-action on Y.

In the next two examples we construct *two different* ALE Spin(7)-manifolds (X_1, Ω_1, g_1) and (X_2, Ω_2, g_2) asymptotic to $\mathbb{R}^8/G$.

Example 3.3. Define complex coordinates $(z_1, \dots, z_4)$ on $\mathbb{R}^8$ by

$$(z_1, z_2, z_3, z_4) = (x_1 + ix_2, x_3 + ix_4, x_5 + ix_6, x_7 + ix_8),$$

Then $g_0 = |\mathrm{d}z_1|^2 + \cdots + |\mathrm{d}z_4|^2$, and $\Omega_0 = \frac{1}{2}\omega_0 \wedge \omega_0 + \mathrm{Re}(\theta_0)$, where ω_0 and θ_0 are the usual Kähler form and complex volume form on $\mathbb{C}^4$. In these coordinates, α and β are given by

$$\begin{aligned} \alpha &: (z_1, \dots, z_4) \longmapsto (iz_1, iz_2, iz_3, iz_4), \\ \beta &: (z_1, \dots, z_4) \longmapsto (\bar{z}_2, -\bar{z}_1, \bar{z}_4, -\bar{z}_3). \end{aligned} \tag{15}$$

Now $\mathbb{C}^4/\langle \alpha \rangle$ is a complex singularity, as $\alpha \in \mathrm{SU}(4)$. Let (Y_1, π_1) be the blow-up of $\mathbb{C}^4/\langle \alpha \rangle$ at 0. Then Y_1 is the unique crepant resolution of $\mathbb{C}^4/\langle \alpha \rangle$. The action of β on $\mathbb{C}^4/\langle \alpha \rangle$ lifts to a *free* antiholomorphic map $\beta : Y_1 \to Y_1$ with $\beta^2 = 1$. Define $X_1 = Y_1/\langle \beta \rangle$. Then X_1 is a nonsingular 8-manifold, and the projection $\pi_1 : Y_1 \to \mathbb{C}^4/\langle \alpha \rangle$ pushes down to $\pi_1 : X_1 \to \mathbb{R}^8/G$.

There exist ALE Calabi–Yau metrics g_1 on Y_1, which were written down explicitly by Calabi [**4**, p. 285], and are invariant under the action of β on Y_1. Let ω_1 be the Kähler form of g_1, and $\theta_1 = \pi_1^*(\theta_0)$ the holomorphic volume form on Y_1. Define $\Omega_1 = \frac{1}{2}\omega_1 \wedge \omega_1 + \mathrm{Re}(\theta_1)$. Then (Ω_1, g_1) is a torsion-free Spin(7)-structure on Y_1, as in Proposition 1.10.

As $\beta^*(\omega_1) = -\omega_1$ and $\beta^*(\theta_1) = \bar{\theta}_1$, we see that β preserves (Ω_1, g_1). Thus (Ω_1, g_1) pushes down to a torsion-free Spin(7)-structure (Ω_1, g_1) on X_1. Then (X_1, Ω_1, g_1) is an *ALE* Spin(7)*-manifold* asymptotic to $\mathbb{R}^8/G$.

Example 3.4. Define new complex coordinates $(w_1, \dots, w_4)$ on $\mathbb{R}^8$ by

$$(w_1, w_2, w_3, w_4) = (-x_1 + ix_3, x_2 + ix_4, -x_5 + ix_7, x_6 + ix_8).$$

Again we find that $g_0 = |\mathrm{d}w_1|^2 + \cdots + |\mathrm{d}w_4|^2$ and $\Omega_0 = \frac{1}{2}\omega_0 \wedge \omega_0 + \mathrm{Re}(\theta_0)$. In these coordinates, α and β are given by

$$\begin{aligned} \alpha &: (w_1, \dots, w_4) \longmapsto (\bar{w}_2, -\bar{w}_1, \bar{w}_4, -\bar{w}_3), \\ \beta &: (w_1, \dots, w_4) \longmapsto (iw_1, iw_2, iw_3, iw_4). \end{aligned} \tag{16}$$

Observe that (15) and (16) are the same, except that the rôles of α, β are reversed. Therefore we can use the ideas of Example 3.3 again.

Let Y_2 be the crepant resolution of $\mathbb{C}^4/\langle \beta \rangle$. The action of α on $\mathbb{C}^4/\langle \beta \rangle$ lifts to a free antiholomorphic involution of Y_2. Let $X_2 = Y_2/\langle \alpha \rangle$. Then X_2 is nonsingular, and carries a torsion-free Spin(7)-structure (Ω_2, g_2), making (X_2, Ω_2, g_2) into an ALE Spin(7)-manifold asymptotic to $\mathbb{R}^8/G$.

We can now explain the remarks on holonomy groups at the end of Step 5. The holonomy groups $\mathrm{Hol}(g_i)$ of the metrics g_1, g_2 in Examples 3.3 and 3.4 are both isomorphic to $\mathbb{Z}_2 \ltimes \mathrm{SU}(4)$, a subgroup of Spin(7). However, they are two *different* inclusions of $\mathbb{Z}_2 \ltimes \mathrm{SU}(4)$ in Spin(7), as in the first case the complex structure is α and in the second β.

The Spin(7)-structure (Ω, g) on Z also has holonomy $\mathrm{Hol}(g) = \mathbb{Z}_2 \ltimes \mathrm{SU}(4)$. Under the natural identifications we have $\mathrm{Hol}(g_1) = \mathrm{Hol}(g)$ but $\mathrm{Hol}(g_2) \neq \mathrm{Hol}(g)$ as subgroups of Spin(7). Therefore, if we choose $i_j = 1$ for all $j = 1, \ldots, k$, then Z and X_{i_j} all have the same holonomy group $\mathbb{Z}_2 \ltimes \mathrm{SU}(4)$, so they combine to give metrics $\tilde{g}_t$ on M with $\mathrm{Hol}(\tilde{g}_t) = \mathbb{Z}_2 \ltimes \mathrm{SU}(4)$.

However, if $i_j = 2$ for some j then the holonomy of g on Z and g_{i_j} on X_{i_j} are *different* $\mathbb{Z}_2 \ltimes \mathrm{SU}(4)$ subgroups of Spin(7), which together generate the whole group Spin(7). Thus they combine to give metrics $\tilde{g}_t$ on M with $\mathrm{Hol}(\tilde{g}_t) = \mathrm{Spin}(7)$.

3.3. Conclusions.

The author was able in [**8**] and [**9**, Ch. 15] to construct compact 8-manifolds with holonomy Spin(7) realizing 14 distinct sets of Betti numbers, which are given in Table 1. Probably there are many other examples which can be produced by similar methods.

Table 1. Betti numbers (b^2, b^3, b^4) of compact Spin(7)-manifolds

$(4, 33, 200)$	$(3, 33, 202)$	$(2, 33, 204)$	$(1, 33, 206)$	$(0, 33, 208)$
$(1, 0, 908)$	$(0, 0, 910)$	$(1, 0, 1292)$	$(0, 0, 1294)$	$(1, 0, 2444)$
$(0, 0, 2446)$	$(0, 6, 3730)$	$(0, 0, 4750)$	$(0, 0, 11\,662)$	

Comparing these Betti numbers with those of the compact 8-manifolds constructed in [**9**, Ch. 14] by resolving torus orbifolds T^8/Γ, we see that in these examples the middle Betti number b^4 is much bigger, as much as 11 662 in one case.

Given that the two constructions of compact 8-manifolds with holonomy Spin(7) that we know appear to produce sets of 8-manifolds with rather different 'geography', it is tempting to speculate that the set of all compact 8-manifolds with holonomy Spin(7) may be rather large, and that those constructed so far are a small sample with atypical behaviour.

References

1. M. Berger, *Sur les groupes d'holonomie homogène des variétés à connexion affines et des variétés Riemanniennes*, Bull. Soc. Math. France **83** (1955), 279–330.
2. R.L. Bryant, *Metrics with exceptional holonomy*, Ann. Math. **126** (1987), 525–576.
3. R.L. Bryant and S.M. Salamon, *On the construction of some complete metrics with exceptional holonomy*, Duke Math. J. **58** (1989), 829–850.
4. E. Calabi, *Métriques kählériennes et fibrés holomorphes*, Ann. Scient. Éc. Norm. Sup. **12** (1979), 269–294.
5. D.D. Joyce, *Compact Riemannian 7-manifolds with holonomy G_2. I*, J. Diff. Geom. **43** (1996), 291–328.
6. ______, *Compact Riemannian 7-manifolds with holonomy G_2. II*, J. Diff. Geom. **43** (1996), 329–375.
7. ______, *Compact Riemannian 8-manifolds with holonomy* Spin(7), Invent. Math. **123** (1996), 507–552.
8. ______, *A new construction of compact 8-manifolds with holonomy* Spin(7), J. Diff. Geom. **53** (1999), 89–130, math.DG/9910002.
9. ______, *Compact manifolds with special holonomy*, Oxford Mathematical Monographs, Oxford University Press, Oxford, 2000, reprinted 2003.
10. ______, *Asymptotically Locally Euclidean metrics with holonomy* SU(m), Ann. Global Anal. Geom. **19** (2001), 55–73, math.AG/9905041.

11. ———, *Quasi-ALE metrics with holonomy* $\mathrm{SU}(m)$ *and* $\mathrm{Sp}(m)$, Ann. Global Anal. Geom. **19** (2001), 103–132, math.AG/9905043.
12. A.G. Kovalev, *Twisted connected sums and special Riemannian holonomy*, J. Reine Angew. Math **565** (2003) 125-160.
13. P.B. Kronheimer, *The construction of ALE spaces as hyperkähler quotients*, J. Diff. Geom. **29** (1989), 665–683.
14. ———, *A Torelli-type theorem for gravitational instantons*, J. Diff. Geom. **29** (1989), 685–697.
15. S.-S. Roan, *Minimal resolution of Gorenstein orbifolds*, Topology **35** (1996), 489–508.
16. S.-T. Yau, *On the Ricci curvature of a compact Kähler manifold and the complex Monge–Ampère equations. I*, Comm. Pure Appl. Math. **31** (1978), 339–411.

Lincoln College, Oxford, OX1 3DR
E-mail address: dominic.joyce@lincoln.ox.ac.uk

Clay Mathematics Proceedings
Volume 3, 2004

From Fano Threefolds to Compact G_2-Manifolds

Alexei Kovalev

Abstract. G_2-manifolds are 7-dimensional Riemannian manifolds whose metrics have holonomy group G_2; these are necessarily Ricci-flat. We explain a systematic way to construct examples of compact G_2-manifolds by gluing a pair of asymptotically cylindrical manifolds of holonomy $SU(3)$ at their cylindrical ends. To obtain the latter $SU(3)$-manifolds one starts from complex 3-dimensional projective manifolds with $c_1 > 0$ (Fano threefolds) endowed with an appropriate choice of the anticanonical K3 divisor. The resulting G_2-manifolds are topologically distinct from those previously constructed by Joyce.

This article is an informal, introductory account of the 'generalized connected sum' construction of compact Riemannian manifolds with holonomy G_2. Full details and proofs the results can be found in the author's paper [**6**]. A good reference on the Riemannian holonomy groups, including G_2 and the previously known construction of compact G_2-manifolds, is the book by Joyce [**4**].

We briefly review in Section 1 the background results on G_2 holonomy. The method of construction of manifolds of holonomy G_2 is explained in Section 2. Section 3 explains how to obtain examples of this construction using the theory of Fano threefolds and K3 surfaces, and contains a discussion of the results.

1. Synopsis on the holonomy group G_2

The holonomy group $\mathrm{Hol}(g)$ of a Riemannian manifold (M, g) is defined as the group of isometries of the tangent space T_xM generated by parallel transport, using the Levi–Civita connection of g, over closed loops based at x. Up to conjugation, the holonomy group is well-defined as a subgroup of $O(n)$, $n = \dim M$. If M is an oriented simply-connected Riemannian manifold, which is not locally isometric to a Riemannian product or to a Riemannian symmetric space, then there are very few groups which may occur as the holonomy of M, according to Berger's classification theorem. In fact, if in addition one assumes that the dimension of M is *odd* then there are just two possibilities: either $\mathrm{Hol}(g) = SO(n)$ or $\dim M = 7$ and $\mathrm{Hol}(g) = G_2$.

2000 *Mathematics Subject Classification.* Primary 53C29, 14J45, 32Q25.

The group G_2 may be defined as the group of automorphisms of the cross-product algebra on $\mathbb{R}^7$ arising from the identification of $\mathbb{R}^7$ with the purely imaginary octonions. It is a compact Lie group and a (proper) subgroup of $SO(7)$. The cross-product multiplication may be encoded by a 3-form $\varphi_0 \in \Lambda^3(\mathbb{R}^7)^*$,

$$\varphi_0(a,b,c) = \langle a \times b, c\rangle,$$

or, explicitly,

$$\begin{aligned}\varphi_0 = & e_5 \wedge e_6 \wedge e_7 + (e_1 \wedge e_2 + e_3 \wedge e_4) \wedge e_7 \\ & + (e_1 \wedge e_3 - e_2 \wedge e_4) \wedge e_6 + (e_1 \wedge e_4 + e_2 \wedge e_3) \wedge e_5,\end{aligned} \tag{1.1}$$

where e_i denote an orthonormal basis of $(\mathbb{R}^7)^*$. Conversely, the formula

$$6\langle a,b\rangle\, d\mathrm{vol}_7 = (a \lrcorner \varphi_0) \wedge (b \lrcorner \varphi_0) \wedge \varphi_0. \tag{1.2}$$

expresses the Euclidean inner product in terms of φ_0 and the volume form of $\mathbb{R}^7$. The group G_2 is thus identified as the stabilizer of φ_0 in the natural action of $GL(7,\mathbb{R})$ on $\Lambda^3(\mathbb{R}^7)^*$. The form φ_0 is stable, in the sense of Hitchin [**2**] —the $GL(7,\mathbb{R})$-orbit of φ_0 is open in $\Lambda^3(\mathbb{R}^7)^*$.

A G_2-structure on a 7-manifold, M say, may be given by a 3-form φ such that at each point $p \in M$, $\varphi(p)$ is the image of φ_0 induced by a linear isomorphism $T_pM \to \mathbb{R}^7$. Denote by $\Omega^3_+(M)$ the subset of 3-forms point-wise modelled on φ_0 in the latter sense; elements of $\Omega^3_+(M)$ will sometimes be referred to as the G_2-structure 3-forms. Note that $\Omega^3_+(M)$ is an open subset of $\Omega^3(M)$ in the sup-norm topology, a direct consequence of the stablity property of φ_0.

Every 3-form $\varphi \in \Omega^3_+(M)$ defines an orientation and a Riemannian metric $g = g(\varphi)$ on M, as any G_2-structure is an instance of an $SO(7)$-structure. The formula (1.2) determines $g(\varphi)$ explicitly, up to a conformal factor. The holonomy group of $g(\varphi)$ will be a subgroup of G_2 if and only if the form φ is parallel, $\nabla\varphi = 0$, with respect to the Levi–Civita connection of g. The latter condition is equivalent to the system of partial differential equations on φ [**9**, Lemma 11.5],

$$d\varphi = 0 \qquad \text{and} \qquad d *_\varphi \varphi = 0. \tag{1.3}$$

The second equation in (1.3) is non-linear as the Hodge star $*_\varphi$ is taken in the metric $g(\varphi)$ and depends on φ. The holonomy reduction $\mathrm{Hol}(g(\varphi)) \subseteq G_2$ implies that $g(\varphi)$ is Ricci-flat.

Proposition 1.1 ([**4**, pp.244–245]). *Suppose that a 7-manifold M is compact and let $\varphi \in \Omega^3_+(M)$. Then* $\mathrm{Hol}(g(\varphi)) = G_2$ *if and only if $\varphi \in \Omega^3_+(M)$ is a solution to* (1.3) and *the fundamental group of M is finite.*

We shall say that a Riemannian 7-manifold (M, g) is a G_2-*manifold* if $\mathrm{Hol}(g) = G_2$ and shall use a similar terminology for other holonomy groups.

The first examples of compact G_2-manifolds were constructed in 1994-5 by Joyce, using a generalized Kummer construction and resolution of singularities. The most elaborate form of this construction can be found in [**4**]. Recently the author obtained different examples of compact G_2-manifolds by a different method [**6**] which we shall now describe.

2. The generalized connected sum construction

Our compact G_2-manifolds are constructed by forming a carefully chosen generalized connected sum of two non-compact Riemannian manifolds with asymptotically cylindrical ends. The construction develops an idea due to Donaldson.

Firstly, we produce a class of complete Ricci-flat Kähler threefolds W of holonomy $SU(3)$ with an infinite cylindrical end asymptotic to the Riemannian product $D \times S^1 \times \mathbb{R}_{>0}$, where D is a K3 surface with a hyper-Kähler metric. This step requires a proof of a non-compact version of the Calabi conjecture, which may be of independent interest.

The Riemannian product $W \times S^1$ carries a solution to (1.3). We consider a pair of such 7-manifolds $W_1 \times S^1$ and $W_2 \times S^1$. For certain pairs of hyper-Kähler K3 surfaces D_i 'at the infinity of W_i' ($i = 1, 2$), there is a way to join the two 7-manifolds $W_i \times S^1$ at their ends to obtain a compact 7-manifold M having finite fundamental group and a 1-parameter family of G_2-structures φ_T compatible with those on $W_i \times S^1$.

A G_2-structure φ_T on M is obtained using cut-off functions, which introduce error terms in the equations (1.3). These error terms are exponentially small in T. We use 'stretching the neck' analysis to prove a gluing theorem, obtaining a solution to (1.3) on M from the solutions on $W_i \times S^1$.

The three parts of the construction are described in more detail below.

2.1. Asymptotically cylindrical Calabi–Yau manifolds. The equations (1.3) define a metric $g(\varphi)$ whose holonomy group is only *contained* in G_2. In particular, the holonomy may be $SU(3)$, a maximal subgroup of G_2.

We begin by introducing the holonomy $SU(n)$ which will be needed in the cases $n = 3$ and $n = 2$. The group $SU(n)$ consists of all the complex linear isomorphisms of $\mathbb{C}^n$ preserving the standard Hermitian inner product and the complex volume. So $SU(n)$ is the stabilizer of the pair of forms on $\mathbb{C}^n$

$$\omega_0 = \tfrac{i}{2}\,(dz_1 \wedge d\bar{z}_1 + \ldots + dz_n \wedge d\bar{z}_n) \qquad \text{and} \qquad \Omega_0 = dz_1 \wedge \cdots \wedge dz_n$$

under the action of $GL(n, \mathbb{C})$. Note that both ω_0 and Ω_0 are stable differential forms (have open orbits under the action of $GL(2n, \mathbb{R})$). A metric g on a real $2n$-manifold Z will have holonomy contained in $SU(n)$ if and only if Z has an $SU(n)$-structure (I, ω, Ω) parallel with respect to g. Here I is an orthogonal complex structure with respect to g, and ω and Ω are differential forms which are point-wise modelled on ω_0 and Ω_0, via a $\mathbb{C}$-linear identification of tangent spaces to Z with $\mathbb{C}^n$. That is to say, a g-parallel $SU(n)$-structure makes Z into a *Kähler complex n-fold* with the Kähler form $\omega \in \Omega^{1,1}(Z)$, and Z has a nowhere vanishing holomorphic form $\Omega \in \Omega^{n,0}(Z)$ such that $\Omega \wedge \Omega^*$ is a constant multiple of ω^n. Such a Ω is sometimes called a holomorphic volume form. In particular, Z has trivial canonical bundle of $(n, 0)$-forms and $c_1(Z) = 0$ and the Kähler metric is Ricci-flat.

Conversely, the following is a direct consequence of Yau's proof of the Calabi conjecture [**12**].

Theorem 2.1. *Let Z be a Kähler complex n-fold with ω_Z the Kähler form on Z and suppose that $c_1(Z) = 0$. Then there exists on Z a unique Ricci-flat Kähler metric such that its Kähler form is given by $\omega_Z + i\partial\bar{\partial}u$ for some smooth real function u on Z. If Z is simply-connected then the holonomy of this Ricci-flat Kähler metric is contained in $SU(n)$.*

Kähler manifolds with holonomy in $SU(n)$ are often called *Calabi–Yau manifolds*. An important example is a K3 surface; recall that it may be defined as a simply-connected complex surface with $c_1 = 0$. By Yau's theorem, a K3 surface admits a unique Ricci-flat Kähler metric in every Kähler class.

Now let $n = 3$. The group $SU(3)$ is a subgroup of G_2 consisting of all those elements of G_2 which fix a particular one-dimensional subspace in $\mathbb{R}^7$, thus it determines a decomposition $\mathbb{R}^7 \cong \mathbb{C}^3 \oplus \mathbb{R}$. Consider a Kähler threefold W with holonomy in $SU(3)$ and let ω, Ω be respectively the Kähler form and a holomorphic volume form on W. Then on the 7-manifold $W \times S^1$ the 3-form

$$\varphi = \omega \wedge d\theta + \operatorname{Im} \Omega \tag{2.1}$$

is in $\Omega^3_+(W \times S^1)$ and defines a product metric, so $W \times S^1$ has the same holonomy as W. In this metric, one has $*\varphi = \frac{1}{2}\omega \wedge \omega - \operatorname{Re} \Omega \wedge d\theta$ and φ is a solution to (1.3) on $W \times S^1$.

We are now ready to introduce the class of complete $SU(3)$-manifolds that we need. Let $\overline{W}$ be a compact simply-connected Kähler threefold, with $\omega' \in \Omega^{1,1}(\overline{W})$ the Kähler form. Let D be a K3 surface in $\overline{W}$ such that there is a holomorphic section s of the anticanonical bundle $K^{-1}_{\overline{W}}$ vanishing to order 1 on D. It is easy to see that the complement $W = \overline{W} \setminus D$ has trivial canonical bundle.

Assume further that the normal bundle of D in $\overline{W}$ is trivial. Then W can be written as the union of two pieces,

$$W \simeq W_{\text{cpt}} \cup (D \times S^1 \times \mathbb{R}_+) \tag{2.2}$$

a compact manifold W_{cpt} with boundary and a cylindrical end attached along the boundary $D \times S^1$. Note that the relation (2.2) is only a diffeomorphism of the underlying real manifolds. The complex structure on the end of W is *not* isomorphic, but only asymptotic to the 'obvious' product complex structure on $D \times S^1 \times \mathbb{R}_+$.

Let g_D denote the Ricci-flat Kähler metric on D in the Kähler class $[\omega'|_D]$ determined by the embedding in $\overline{W}$. We prove that the following non-compact version of the Calabi conjecture is true.

THEOREM 2.2. *Let $\overline{W}$ and D be as above, so a K3 surface D is an anticanonical divisor and has trivial normal bundle in $\overline{W}$. Suppose also that $\overline{W}$ is simply-connected and the fundamental group of $W = \overline{W} \setminus D$ is finite.*

Then W admits a complete Ricci-flat Kähler metric g_W. The Kähler form and holomorphic volume form of g_W are exponentially asymptotic, along the cylindrical end of W, to those of the product Ricci-flat Kähler structure on $D \times S^1 \times \mathbb{R}_+$ defined using the metric g_D on D. The holonomy of g_W is $SU(3)$.

There is nothing special to threefolds in the proof of Theorem 2.2 and the result extends, with only minor modifications, to Kähler manifolds of arbitrary dimension.

We also remark at this point that previously a number of other non-compact versions of the Calabi conjecture were proved by Tian and Yau, Bando and Kobayashi, and Joyce. These authors construct complete Ricci-flat Kähler metrics asymptotic at infinity to the quotient $\mathbb{C}^n/\Gamma$ of Hermitian $\mathbb{C}^n$ by a finite subgroup Γ of $SU(n)$.

The main novelty of Theorem 2.2 is that it deals with the class of *asymptotically cylindrical* manifolds. We build up on Theorem 5.2 in [**11**] using analysis on exponentially weighted Sobolev spaces to work out the details of asymptotic behaviour and provide control on the boundary data at infinity.

2.2. K3 surfaces and a hyper-Kähler rotation. A remarkable property of a Ricci-flat Kähler metric on a complex surface D is that such a metric is *hyper-Kähler*: the underlying real 4-manifold admits, in addition to the given complex structure I, another complex structure J, such that $IJ = -JI$ and the metric is Kähler with respect to J too. Further, $K = IJ$ is also a complex structure on D, and I, J, K satisfy the quaternionic relations (and define an identification of each tangent space of D with the quaternions). The three respective Kähler forms κ_I,κ_J,κ_K satisfy $\kappa_I^2 = \kappa_J^2 = \kappa_K^2$. There is a complete $SO(3)$ symmetry between I, J, K, in particular, they generate a 2-sphere of complex structures $aI + bJ + cK$ on D, where $a^2 + b^2 + c^2 = 1$, and the metric is Kähler with respect to each of these.

Let D be a Ricci-flat Kähler K3 surface and let κ_I be the Kähler form on D. Then $\kappa_J + i\kappa_K$ defines a holomorphic volume form on D. Considering on D the complex structure J we obtain in general a different Ricci-flat Kähler K3 surface D_J. It has Kähler form κ_J and holomorphic volume form $\kappa_I - i\kappa_K$ and is sometimes called a *hyper-Kähler rotation* of D. Note that there is an S^1-ambiguity in choosing J, as one may take any $bJ + cK$ instead with $b^2 + c^2 = 1$.

Consider two asymptotically cylindrical $SU(3)$-manifolds W_1 and W_2 satisfying the assertions of Theorem 2.2 and, respectively, let D_1,D_2 be the Ricci-flat Kähler K3 surfaces which determine the asymptotic model on the cylindrical ends of W_1, W_2. For $i = 1, 2$, let $t_i \geq 0$ be the real parameter along the cylindrical end of W_i, as defined by (2.2). Cut off at $t_i = T - 1$ the Kähler and holomorphic volume form on each W_i to their asymptotic model on the cylindrical end and consider $W_i(T) \simeq W_{\rm cpt} \cup (D_i \times S^1 \times [0, T])$. Then $W_1(T) \times S^1$ is a manifold with boundary $D_1 \times S^1 \times S^1$ and with a G_2-structure form which on a collar neighbourhood of the boundary is given by

$$\varphi_{(D_1)} = \kappa'_I \wedge d\theta_1 + \kappa'_J \wedge d\theta_2 + \kappa'_K \wedge dt + d\theta_1 \wedge d\theta_2 \wedge dt. \tag{2.3}$$

Here we used (2.1) and the cylindrical asymptotic model $\omega = \kappa'_I + d\theta_2 \wedge dt$, $\Omega = (\kappa'_J + i\kappa'_K) \wedge (d\theta_2 + idt)$ of the Kähler and holomorphic volume forms on the end of W_1. In particular, $\varphi_{(D_1)}$ is a solution to (1.3) on the cylinder $(D_1 \times S^1 \times \mathbb{R}) \times S^1$. Similar expressions hold for $W_2 \times S^1$.

Now *assume* that the Ricci-flat Kähler K3 surface D_2 is isomorphic to a hyper-Kähler rotation of D_1. Let $f : D_{1,J} \to D_2$ denote the isomorphism. Then the pull-back action of f on the Kähler forms is given by

$$f^* : \quad \kappa''_I \mapsto \kappa'_J, \quad \kappa''_J \mapsto \kappa'_I, \quad \kappa''_K \mapsto (-\kappa'_K).$$

Define

$$\begin{aligned} F : \ (y, \theta_1, \theta_2, t) \in D_1 \times S^1 \times S^1 \times [T-1, T] \to \\ (f(y), \theta_2, \theta_1, 2T - 1 - t) \in D_2 \times S^1 \times S^1 \times [T-1, T] \end{aligned} \tag{2.4}$$

and join the two manifolds with boundary to construct a closed oriented 7-manifold

$$M = (W_1(T) \times S^1) \cup_F (W_2(T) \times S^1),$$

using the map F to identify collar neighbourhoods of the boundaries. The compact 7-manifold M is a generalized connected sum with the neck having the cross-section $D \times S^1 \times S^1$. We have $F^*\varphi_{(D_1)} = \varphi_{(D_2)}$, by the construction of F, therefore there is a well-defined 1-parameter family of G_2-structures φ_T on M induced from those on $W_i(T) \times S^1$, defined above using cut-off functions. Here the parameter T is approximately half the length of the neck of M, measured by $g(\varphi_T)$.

The fundamental group of M is *finite*. This is because in the construction of M the circle factor in $W_1(T) \times S^1$ is identified with a circle in W_2 and the circle factor in $W_2(T) \times S^1$ is identified with a circle in W_1, and we assumed that $\pi_1(W_i)$ are finite. Therefore, by Proposition 1.1 any solution $\varphi \in \Omega^3_+(M)$ of the equations (1.3) will define on M a metric $g(\varphi)$ of holonomy G_2.

2.3. The gluing theorem. A G_2-structure form φ_T is constructed by patching the solutions of (1.3) when joining the two pieces of M. This uses cut-off functions which introduce 'error terms' in the equations. In fact we can achieve $d\varphi_T = 0$ for all T and satisfy one of the two equations in (1.3), but the term $d{*_T}\varphi_T$ in general will not vanish. The error terms depend on the difference between the $SU(3)$-structures on the end of W_i and on its cylindrical asymptotic model, and we have an estimate

$$\|d *_T \varphi_T\|_{L^p_k} < C_{p,k} e^{-\lambda T},$$

where $0 < \lambda < 1$. Here $*_T$ denotes the Hodge star of the metric $g(\varphi_T)$.

We prove the following.

THEOREM 2.3. *There exists $T_0 \in \mathbb{R}$ and for every $T \geq T_0$ a unique smooth 2-form η_T on M so that the following holds.*
(1) $\|\eta_T\|_{C^1} < const\cdot e^{-\mu T}$, for some $0 < \mu < 1$, where the C^1-norm is defined using the metric $g(\varphi_T)$. In particular, $\varphi_T + d\eta_T$ is in $\Omega^3_+(M)$.
(2) The closed 3-form $\varphi_T + d\eta_T$ satisfies

$$d *_{\varphi_T + d\eta_T} (\varphi_T + d\eta_T) = 0. \tag{2.5}$$

and so $\varphi_T + d\eta_T$ defines a metric of holonomy G_2 on M.

The equation (2.5) can be rewritten, using the results of [**4**, §10.3] as a non-linear elliptic PDE for η. For small η, this PDE has the form $a(\eta) = a_0 + A\eta + Q(\eta) = 0$, where $a_0 = d *_T \varphi_T$, the linear elliptic operator $A = A_T$ is a compact perturbation of the Hodge Laplacian of the form $dd^* + d^*d + O(e^{-\delta T})$, and $Q(\eta) = O(|d\eta|^2)$.

The central idea in the proof of Theorem 2.3 may be informally stated as follows. For small η, the map $a(\eta)$ is approximated by its linearization and so there is a unique small solution η to the equation $a(\eta) = 0$, for every small a_0 in the range of A. This perturbative approach requires the invertibility of A and a suitable upper bound on the operator norm $\|A_T^{-1}\|$, as $T \to \infty$. This bound determines what is meant by 'small' a_0 in this paragraph.

As we actually need the value of $d\eta$ rather than η, there is no loss in restricting the equation (2.5) for η to the orthogonal complement of harmonic 2-forms on M where the Laplacian is invertible. We use the technique of [**5**, §4.1] based on Fredholm theory for the asymptotically cylindrical manifolds and weighted Sobolev spaces to find an upper bound $\|A_T^{-1}\| < Ge^{\delta T}$. Here the constant G is independent of T and $\delta > 0$ can be taken arbitrary small. So, for large T, the growth of $\|A_T^{-1}\|$ is negligible compared to the decay of $\|d{*_T}\varphi_T\|$ and the 'inverse function theorem' strategy applies to give the required small solution η_T. Standard elliptic methods show that this η_T is in fact smooth.

3. Examples arising from Fano threefolds

For applications of the construction given in Section 2 we need, as a start, to find Kähler threefolds satisfying the hypotheses of Theorem 2.2.

3.1. Introduction to Fano threefolds. The following example is classical in algebraic geometry.

The intersection of three generically chosen quadric hypersurfaces in $\mathbb{C}P^6$ defines a smooth Kähler threefold X_8. It is simply-connected and the characteristic class $c_1(X_8)$ of its anticanonical bundle is the pull-back to X_8 of the positive generator of the cohomology ring $H^*(\mathbb{C}P^6)$. That is to say, the anticanonical bundle $K_{X_8}^{-1}$ is the restriction to X_8 of the tautological line bundle $\mathcal{O}(1)$ over $\mathbb{C}P^6$. It follows that any anticanonical divisor D on X_8 is obtained by taking an intersection $D = X_8 \cap H$ with a hyperplane H in $\mathbb{C}P^6$. A generic such hyperplane section D is a complex surface, isomorphic to a smooth complete intersection of three quadrics in $\mathbb{C}P^5$. This is a well-known example of a K3 surface.

We next look at the normal bundle of X_8 in D. An adjunction-type argument shows that the normal bundle will be trivial if we can find another anticanonical divisor D' on X_8 such that D' does not meet D. But $D' = X_8 \cap H'$ and the second hyperplane section $C = D \cap D' = X_8 \cap H \cap H'$ is *never* empty—it is an algebraic curve (intersection of three quadrics) in $\mathbb{C}P^4$. Fortunately, a suitable threefold can be obtained by blowing up the curve C. The K3 divisor D lifts via the blow-up map $\tilde{X}_8 \to X$ to an isomorphic K3 surface $\tilde{D}$ which is an anticanonical divisor in $\tilde{X}_8$ and has trivial normal bundle. Moreover, a Kähler metric on $\tilde{X}_8$ may be chosen so that $\tilde{D}$ and D are isometric Kähler manifolds.

Finally, as both $\tilde{D}$ and X_8 are simply-connected we find that the only possibility for a nontrivial generator of $\pi_1(X_8 \setminus D)$ would be a circle around $\tilde{D}$. But this circle contracts in an exceptional curve as this curve meets $\tilde{D}$ in exactly one point. Hence $\tilde{X}_8 \setminus \tilde{D}$ is simply connected. The pair $\tilde{X}_8$, $\tilde{D}$ now satisfies all the hypotheses of Theorem 2.2, and so the quasiprojective threefold $W = \tilde{X}_8 \setminus \tilde{D}$ admits an asymptotically cylindrical Ricci-flat Kähler metric of holonomy $SU(3)$.

The threefold X_8 in the above example can be replaced by an arbitrary (smooth) projective-algebraic threefold V with $c_1(V) > 0$, i.e. a *Fano threefold*. Fano threefolds have been extensively studied over the past few decades and a lot is known about them. In particular, they are simply-connected and a generic anticanonical divisor D on a Fano threefold is a K3 surface [**10**]. It can be shown that the threefolds $\tilde{V} \setminus \tilde{D}$ are again simply-connected and we obtain, by application of Theorem 2.2 the following.

PROPOSITION 3.1. *Let V be a Fano 3-fold, $D \in |-K_V|$ a K3 surface, and $\tilde{V}$ the blow-up of V along a self-intersection curve $D \cdot D$, and $\tilde{D}$ the proper transform of D. Then $\tilde{V} \setminus \tilde{D}$ has a complete Ricci-flat Kähler metric with holonomy $SU(3)$. This metric is asymptotic to the Riemannian product $D \times S^1 \times \mathbb{R}_{>0}$, where the Ricci-flat Kähler metric on D is in the Kähler class induced by the embedding in V.*

3.2. Matching the K3 divisors. Let V_1, V_2 be Fano threefolds and D_1, D_2, respectively, anticanonical K3 divisors on these. Recall that by Yau's theorem each of the Kähler K3 surfaces D_i has a uniquely determined Ricci-flat Kähler metric in its Kähler class. If the two Ricci-flat Kähler structures in the Kähler classes of $D_i \subset V_i$ are hyper-Kähler rotations of each other then, in view of Proposition 3.1, we can proceed to the construction of a generalized connected sum M from $\tilde{V}_i \setminus \tilde{D}_i$, as described in Section 2.2. The 7-manifold M admits metrics of holonomy G_2 by Theorem 2.3. In this case, we shall say that a compact G_2-manifold M is *constructed from the pair of Fano threefolds* V_1 *and* V_2. Can we choose D_1 and D_2 so as to satisfy the required hyper-Kähler rotation condition?

We have the freedom to move a K3 surface in the anticanonical linear system of V and to deform V in its algebraic family of Fano threefolds. Recall also from Section 2.2 that there is an S^1-family of choices for the second complex structure J on each D_i. It turns out that, with this freedom, a pair of 'matching divisors' D_1, D_2 can always be found. We now briefly explain, ignoring some important technical points, the ideas in the solution of the matching problem.

All the K3 surfaces are deformations of each other and are diffeomorphic as real 4-manifolds. In particular, their second cohomology lattices are isomorphic to the (unique) even unimodular lattice L of signature $(3, 19)$, known as the K3 lattice. Respectively, $L \otimes \mathbb{C}$ is isomorphic to the second cohomology with complex coefficients and inherits the Hodge decomposition. A Kähler isometry between two K3 surfaces induces a so-called effective Hodge isometry between their second cohomology lattices, preserving the Hodge decomposition and mapping the Kähler class of one K3 to the Kähler class of the other. Surprisingly, the global Torelli theorem for K3 surfaces asserts that the converse is also true: any effective Hodge isometry between second cohomology of two K3 surfaces arises as the pull-back of a unique biholomorphic map between these K3 surfaces [**1**, Ch.VIII]. The latter map will necessarily be an isometry between *Ricci-flat* Kähler K3 surfaces because of the uniqueness of a Ricci-flat Kähler metric in a Kähler class.

We can identify, using a version of the Kodaira–Spencer–Kuranishi deformation theory, the data of Hodge decomposition and Kähler class which occurs in the anticanonical K3 divisors in a given algebraic family of Fano threefolds. The problem of choosing a matching pair of K3 divisors D_i in Fano threefolds V_i then reduces to a problem in the arithmetic of the K3 lattice.

The solution of this problem gives us the following general result.

THEOREM 3.2. *For any pair of algebraic families* $\mathcal{V}_1$,$\mathcal{V}_2$ *of Fano threefolds there exists (smooth)* $V_1 \in \mathcal{V}_1$, $V_2 \in \mathcal{V}_2$ *such that a compact* G_2*-manifold* M *can be constructed from* V_1,V_2.

This G_2*-manifold satisfies*

$$0 \leq b_2(M) \leq \max\{b_2(V_1), b_2(V_2)\} - 1$$

and

$$b_2(M) + b_3(M) = b_3(V_1) - K_{V_1}^3 + b_3(V_2) - K_{V_2}^3 + 27.$$

EXAMPLE 3.3. A smooth complete intersection X_8 of three quadrics in $\mathbb{C}P^6$ has $b^2 = 1$, $b^3 = 28$, and $-K^3 = 8$. According to Theorem 3.2, an appropriate choice of two such complete intersections $V_8^{(1)}, V_8^{(2)} \subset \mathbb{C}P^6$ and of a hyperplane section D_i in each of the $V_8^{(i)}$ provides data for the construction of a compact G_2-manifold M. We obtain $b_2(M) = 0$ and $b_3(M) = 99$. Also M is simply-connected. This G_2-manifold is not homeomorphic to any of the examples constructed in [**4**].

3.3. Discussion of the results. There is a complete classification of smooth Fano threefolds into 104 algebraic families [**3, 8**]. This provides 5,460 different pairs to form the generalized connected sums, leading to examples of compact G_2-manifolds. As any Fano threefold has $1 \leq b^2 \leq 10$, we have

$$b_2(M) \leq 9,$$

for any G_2-manifold M constructed from a pair of Fanos. Further inspecting the classification list of Fanos, we find that

$$39 \leq b_2(M) + b_3(M) \leq 239$$

and, in particular, $b_3(M) \geq 30$. (Recall that $b^1 = 0$ for any G_2-manifold, therefore b^2, b^3 determine all the Betti numbers of M.)

It is an interesting question to identify the most general class of Kähler threefolds $\overline{W}$ for which the hypotheses of Theorem 2.2 hold. If the anticanonical linear systems on a pair of such $\overline{W}$ are 'large enough' then the connected sum defined in Section 2 can be formed and will admit G_2-metrics. It seems that the blow-ups $\tilde{V}$ of smooth Fano threefolds discussed in this section can be generalized to include at least manifolds obtained by resolution of singularities in some singular Fano varieties.

A pair of Fano threefolds in general yields several topologically distinct compact G_2-manifolds. Example in [**6**, §8] shows two topologically distinct G_2-manifolds constructed from a pair of $\mathbb{C}P^2 \times \mathbb{C}P^1$'s, realizing both of the values $b_2(M) = 0$ and $b_2(M) = 1$ allowed in this case by Theorem 3.2. Of course, the counting of pairs of Betti numbers (b^2, b^3) only gives a lower estimate of the actual number of topological types realized by our examples of compact G_2-manifolds.

In any event, the majority of smooth Fano threefolds have the Betti number $b^2 \leq 4$ and respectively the G_2-manifolds constructed from these have $b^2 \leq 3$. On the other hand, a majority of the compact G_2-manifolds constructed in [**4**] have $b^2 > 3$. Thus most of the compact G_2-manifolds constructed from smooth Fano threefolds can be easily identified as new examples, topologically distinct from those previously known.

Another interesting property of the connected sum construction is that it exhibits a new type of boundary point in the moduli space of all G_2-metrics on the given compact 7-manifold M. Any 1-parameter family $\varphi_T + d\eta_T$ of G_2-metrics given by the gluing Theorem 2.3 defines a path in the moduli space. The boundary point attained as $T \to \infty$ corresponds to pulling apart a G_2-manifold at a cross-section K3$\times$(2-torus), obtaining a pair of asymptotically cylindrical pieces. The approach to the boundary of the moduli space in this case involves no development of singularities, nor a curvature growth. This becomes important in [**7**] where we construct the first examples of fibrations of compact G_2-manifolds by certain *minimal* submanifolds called coassociative calibrated submanifolds. The fibrations are an odd-dimensional non-holomorphic analogue of the well-known elliptic fibrations of K3 surfaces.

References

1. W. Barth, C. Peters, and A. van de Ven, *Compact complex surfaces.* Springer-Verlag, 1984.
2. N. Hitchin, *Stable forms and special metrics*, Global differential geometry: the mathematical legacy of Alfred Gray (Bilbao, 2000), Contemp. Math., vol. 288, Amer. Math. Soc., 2001, pp. 70–89.
3. V.A. Iskovskih, *Fano threefolds. I. II,* Izv. Akad. Nauk SSSR Ser. Mat. **41** (1977), 516–562,717, and **42** (1978), 506–549. English translation: Math. USSR Izvestia **11** (1977), 485–527, and **12** (1978), 469–506.

4. D.D. Joyce, *Compact manifolds with special holonomy*, OUP Mathematical Monographs series, 2000.
5. A.G. Kovalev and M.A. Singer, *Gluing theorems for complete anti-self-dual spaces*, Geom. Func. Anal. **11** (2001), 1229–1281.
6. A.G. Kovalev, *Twisted connected sums and special Riemannian holonomy*, J. Reine Angew. Math. **565** (2003) 125-160.
7. A.G. Kovalev, *Coassociative K3 fibrations of compact G_2-manifolds*, in preparation.
8. S. Mukai and S. Mori, *Classification of Fano threefolds with $B_2 \geq 2$*, Manuscr. Math. **36** (1981/1982), 147–162.
9. S.M. Salamon, *Riemannian geometry and holonomy groups*, Pitman Res. Notes in Math. 201, Longman, Harlow, 1989.
10. V.V. Shokurov, *Smoothness of a general anticanonical divisor on a Fano variety.* Izv. Akad. Nauk SSSR Ser. Mat. **43** (1979), 430–441. English translation: Math. USSR-Izv. **14** (1980), 395–405.
11. G. Tian and S.-T. Yau, *Complete Kähler manifolds with zero Ricci curvature. I*, J. Amer. Math. Soc. **3** (1990), 579–609.
12. S.-T. Yau. *On the Ricci curvature of a compact Kähler manifold and the complex Monge-Ampère equation. I*, Comm. Pure Appl. Math. **31** (1978), 339–411.

DPMMS, University of Cambridge, Centre for Mathematical Sciences, Wilberforce Road, Cambridge CB3 0WB, U.K.

E-mail address: a.g.kovalev@dpmms.cam.ac.uk

Clay Mathematics Proceedings
Volume 3, 2004

An introduction to motivic integration

Alastair Craw

Abstract. By associating a *motivic integral* to every complex projective variety X with at worst Gorenstein canonical singularities, Kontsevich [**20**] proved that, when a crepant resolution of singularities $\varphi\colon Y \to X$ exists, the Hodge numbers of Y do not depend upon the choice of the crepant resolution. In this article we provide an elementary introduction to the theory of motivic integration, leading to a proof of the result described above. We calculate the motivic integral of several quotient singularities and discuss these calculations in the context of the cohomological McKay correspondence.

1. Introduction

The string-theoretic Mirror Symmetry Conjecture states that there exist mirror pairs of Calabi–Yau varieties with certain compatibilities. For instance, if (X, X^*) is a smooth, projective mirror pair of dimension n, then we expect the relation

$$h^{p,q}(X) = h^{n-p,q}(X^*) \tag{1.1}$$

to hold between their Hodge numbers. Mirror pairs are not smooth in general and the compatibility relation (1.1) can fail to hold if either X or its mirror have singularities. In this case, if there exist crepant resolutions $Y \to X$ and $Y^* \to X^*$ then we expect the relation

$$h^{p,q}(Y) = h^{n-p,q}(Y^*) \tag{1.2}$$

to hold between the Hodge numbers of the smooth varieties Y and Y^* (recall that a resolution $\varphi\colon Y \to X$ is said to be *crepant* if $K_Y = \varphi^* K_X$). However, it is not obvious that the revised relation (1.2) is well defined: even if a crepant resolution exists it is not necessarily unique. In particular, given two crepant resolutions $Y_1 \to X$ and $Y_2 \to X$, it is not clear a priori that the Hodge numbers of Y_1 and Y_2 are equal.

Nevertheless, the consistency of string theory led Batyrev and Dais [**5**] to conjecture that for a variety X with only mild Gorenstein singularities, the Hodge numbers of Y_1 and Y_2 are equal. In a subsequent paper [**2**], Batyrev used methods of p-adic integration to prove that the Betti numbers of Y_1 and Y_2 are equal.

2000 *Mathematics Subject Classification.* Primary 14E15, 32S45 Secondary 14E30, 14M25.
Key words and phrases. Motivic integration, Space of formal arcs, McKay correspondence.

Kontsevich [**20**] later proved that the Hodge numbers are equal by introducing the notion of motivic integration.

This article provides an elementary introduction to Kontsevich's theory of motivic integration. The first step is to construct the motivic integral of a pair (Y, D), for a complex manifold Y and an effective divisor D on Y with simple normal crossings. We define the *space of formal arcs* $J_\infty(Y)$ of Y and associate a function F_D defined on $J_\infty(Y)$ to the divisor D. The *motivic integral* of the pair (Y, D) is the integral of F_D over $J_\infty(Y)$ with respect to a certain measure μ on $J_\infty(Y)$. This measure is not real-valued. Indeed the subtlety in the construction is in defining the ring R in which μ takes values: R is a completion of the polynomial ring in the formal variable $\mathbb{L}^{-1}$ with values in the Grothendieck ring of algebraic varieties over $\mathbb{C}$ (see Definition 2.11). We adopt the structure of the proof of Theorem 6.28 from Batyrev [**1**] to establish the following user-friendly formula:

THEOREM 1.1 (formula for the motivic integral). *Let Y be a complex manifold of dimension n and $D = \sum_{i=1}^r a_i D_i$ an effective divisor on Y with simple normal crossings. The motivic integral of the pair (Y, D) is*

$$\int_{J_\infty(Y)} F_D \,\mathrm{d}\mu = \sum_{J \subseteq \{1,\dots,r\}} [D_J^\circ] \cdot \left(\prod_{j \in J} \frac{\mathbb{L} - 1}{\mathbb{L}^{a_j+1} - 1} \right) \cdot \mathbb{L}^{-n} \tag{1.3}$$

where we sum over all subsets $J \subseteq \{1, \dots, r\}$ including $J = \emptyset$.

The motivic integral of a complex algebraic variety X with Gorenstein canonical singularities is defined to be the motivic integral of a pair (Y, D), where $Y \to X$ is a resolution of singularities for which the discrepancy divisor D has simple normal crossings. Crucially, this is well defined independent of the choice of resolution.

The motivic integral induces a *stringy E-function*

$$E_{\mathrm{st}}(X) := \sum_{J \subseteq \{1,\dots,r\}} E(D_J^\circ) \cdot \left(\prod_{j \in J} \frac{uv - 1}{(uv)^{a_j+1} - 1} \right) \tag{1.4}$$

which is also independent of the choice of resolution (see Warning 3.4). The E-polynomials $E(D_J^\circ)$ encode the Hodge–Deligne numbers of open strata $D_J^\circ \subset Y$, and the stringy E-function records these numbers with certain 'correction terms' written in parentheses in formula (1.4). When $Y \to X$ is a crepant resolution the correction terms disappear leaving simply the terms $E(D_J^\circ)$ whose sum is the E-polynomial of Y. As a result, when a crepant resolution $Y \to X$ exists, the function $E_{\mathrm{st}}(X)$ encodes the Hodge numbers of Y, thereby establishing Kontsevich's result on the equality of Hodge numbers. It is important to note however that crepant resolutions do not exist in general. To get a better feeling for the stringy E-function of varieties admitting no crepant resolution we calculate $E_{\mathrm{st}}(X)$ for several 4- and 6-dimensional Gorenstein terminal cyclic quotient singularities.

We conclude this article with an application of motivic integration, namely the proof by Batyrev [**3, 4**] of the cohomological McKay correspondence conjecture which we now recall. A celebrated result of John McKay [**22**] states that the graph of ADE type associated to a Kleinian singularity $\mathbb{C}^2/G$ can be constructed using only the representation theory of the finite subgroup $G \subset \mathrm{SL}(2, \mathbb{C})$. This establishes a one-to-one correspondence between a basis for the cohomology of the minimal resolution Y of $\mathbb{C}^2/G$ and the irreducible representations of G. In particular, the Euler number $e(Y)$ is equal to the number of irreducible representations (or conjugacy classes) of G. Reid [**27**] proposed the following generalisation:

CONJECTURE 1.2 (The McKay correspondence). *For $G \subset \mathrm{SL}(n, \mathbb{C})$ a finite subgroup, suppose that the quotient variety $X := \mathbb{C}^n/G$ admits a crepant resolution $\varphi\colon Y \to X$. Then $H^*(Y, \mathbb{Q})$ has a basis consisting of algebraic cycles corresponding one-to-one with conjugacy classes of G. In particular, the Euler number of Y equals the number of conjugacy classes of G.*

Batyrev proved this result using the theory of motivic integration. The key step in the proof is to show that the stringy E-function of the quotient $\mathbb{C}^n/G$ coincides with the 'orbifold E-function' of the pair $(\mathbb{C}^n, G)$ (see §5.2 for the definition). We present a simple, direct proof of the conjecture for a finite Abelian subgroup $G \subset \mathrm{SL}(n, \mathbb{C})$ to illustrate the simplicity of Batyrev's approach in this case.

The original references on motivic integration are by Batyrev [**1**, §6] and Denef and Loeser [**14**]. The more recent article by Looijenga [**21**] provides a detailed survey of motivic integration.

Acknowledgements I'd like to thank the organisers of the 2002 Clay Mathematics Institute School on Geometry and String Theory held at the Isaac Newton Institute in Cambridge, especially Miles Reid who originally encouraged me to write this article (which has existed in roughly the present form since November 1998). Also, thanks to Willem Veys and to Victor Batyrev for useful comments.

2. Construction of the motivic integral

2.1. The space of formal arcs of a complex manifold.

DEFINITION 2.1. Let Y be a complex manifold of dimension n. A *k-jet* over a point $y \in Y$ is a morphism

$$\gamma_y\colon \operatorname{Spec} \mathbb{C}[z]/\langle z^{k+1}\rangle \longrightarrow Y$$

with $\gamma_y(\operatorname{Spec} \mathbb{C}) = y$. Once local coordinates are chosen, the space of k-jets over $y \in Y$ can be viewed as the space of n-tuples of polynomials of degree k whose constant terms are zero. Let $J_k(Y)$ denote the bundle over Y whose fibre over $y \in Y$ is the space of k-jets over y. A *formal arc* over $y \in Y$ is a morphism

$$\gamma_y\colon \operatorname{Spec} \mathbb{C}\,[\![z]\!] \longrightarrow Y$$

with $\gamma_y(\operatorname{Spec} \mathbb{C}) = y$. Once local coordinates are chosen, the space of formal arcs over $y \in Y$ can be viewed as the space of n-tuples of power series whose constant terms are zero. Let $J_\infty(Y)$ denote the bundle whose fibre over $y \in Y$ is the space of formal arcs over y. For each $k \in \mathbb{Z}_{\geq 0}$ the inclusion $\mathbb{C}[z]/\langle z^{k+1}\rangle \hookrightarrow \mathbb{C}\,[\![z]\!]$ induces a surjective map

$$\pi_k\colon J_\infty(Y) \longrightarrow J_k(Y),$$

where $\pi_0\colon J_\infty(Y) \to Y$ sends γ_y to the point y.

Recall that a subset of a variety is *constructible* if it is a finite, disjoint union of (Zariski) locally closed subvarieties.

DEFINITION 2.2. A subset $C \subseteq J_\infty(Y)$ of the space of formal arcs is called a *cylinder set* if $C = \pi_k^{-1}(B_k)$ for $k \in \mathbb{Z}_{\geq 0}$ and $B_k \subseteq J_k(Y)$ a constructible subset.

It's clear that the collection of cylinder sets forms an algebra of sets (see [**28**, p. 10]), i.e., $J_\infty(Y) = \pi_0^{-1}(Y)$ is a cylinder set, as are finite unions and complements (and hence finite intersections) of cylinder sets.

2.2. The function F_D associated to an effective divisor.

Definition 2.3. Let D be an effective divisor on Y and g a local defining equation for D on a neighbourhood U of a point $y \in Y$. For an arc γ_u over a point $u \in U$, define the *intersection number* $\gamma_u \cdot D$ to be the order of vanishing of the formal power series $g(\gamma_u(z))$ at $z = 0$. Let

$$F_D \colon J_\infty(Y) \longrightarrow \mathbb{Z}_{\geq 0} \cup \infty$$

be the function defined by $F_D(\gamma_u) = \gamma_u \cdot D$. Write $D = \sum_{i=1}^r a_i D_i$ as a linear combination of prime divisors. Then g decomposes as a product $g = \prod_{i=1}^r g_i^{a_i}$ of defining equations for D_i, hence $F_D = \sum_{i=1}^r a_i F_{D_i}$. Furthermore

$$F_{D_i}(\gamma_u) = 0 \iff u \notin D_i \quad \text{and} \quad F_{D_i}(\gamma_u) = \infty \iff \gamma_u \subseteq D_i. \tag{2.1}$$

Our ultimate goal is to integrate the function F_D over $J_\infty(Y)$, so we must understand the nature of the level set $F_D^{-1}(s) \subseteq J_\infty(Y)$ for each $s \in \mathbb{Z}_{\geq 0} \cup \infty$. With this goal in mind, we introduce a partition of $F_D^{-1}(s)$.

Definition 2.4. For $D = \sum_{i=1}^r a_i D_i$ and $J \subseteq \{1, \dots, r\}$ any subset, define

$$D_J := \begin{cases} \bigcap_{j \in J} D_j & \text{if } J \neq \emptyset \\ Y & \text{if } J = \emptyset \end{cases} \qquad \text{and} \qquad D_J^\circ := D_J \setminus \bigcup_{i \in \{1,\dots,r\} \setminus J} D_i.$$

These subvarieties stratify Y and define a partition of the space of arcs into cylinder sets:

$$Y = \bigsqcup_{J \subseteq \{1,\dots,r\}} D_J^\circ \quad \text{and} \quad J_\infty(Y) = \bigsqcup_{J \subseteq \{1,\dots,r\}} \pi_0^{-1}(D_J^\circ).$$

For any $s \in \mathbb{Z}_{\geq 0}$ and any subset $J \subseteq \{1, \dots, r\}$, define

$$M_{J,s} := \left\{ (m_1, \dots, m_r) \in \mathbb{Z}_{\geq 0}^r \mid \sum a_i m_i = s \text{ with } m_j > 0 \Leftrightarrow j \in J \right\}.$$

It now follows from (2.1) that

$$\gamma_u \in \pi_0^{-1}(D_J^\circ) \cap F_D^{-1}(s) \iff (F_{D_1}(\gamma_u), \dots, F_{D_r}(\gamma_u)) \in M_{J,s}.$$

As a result we produce a finite partition of the level set

$$F_D^{-1}(s) = \bigsqcup_{J \subseteq \{1,\dots,r\}} \bigsqcup_{(m_1,\dots,m_r) \in M_{J,s}} \left(\bigcap\nolimits_{i=1,\dots r} F_{D_i}^{-1}(m_i) \right). \tag{2.2}$$

Proposition 2.5. *If D is an effective divisor with simple normal crossings then $F_D^{-1}(s)$ is a cylinder set (see Definition 2.2) for each $s \in \mathbb{Z}_{\geq 0}$.*

Recall (see [**19**, p. 25]) that a divisor $D = \sum_{i=1}^r a_i D_i$ on Y has only simple normal crossings if at each point $y \in Y$ there is a neighbourhood U of y with coordinates $z_1, \dots, z_n$ for which a local defining equation for D is

$$g = z_1^{a_1} \cdots z_{j_y}^{a_{j_y}} \quad \text{for some } j_y \leq n. \tag{2.3}$$

Proof of Proposition 2.5. A finite union of cylinder sets is cylinder and we have a partition (2.2) of $F_D^{-1}(s)$, so it is enough to prove that $\bigcap_{i=1,\dots r} F_{D_i}^{-1}(m_i)$ is a cylinder set[1] for some $J \subseteq \{1, \ldots, r\}$ and $(m_1, \ldots, m_r) \in M_{J,s}$. Cover Y by finitely many charts U on which D has a local equation of the form (2.3), and lift to cover $J_\infty(Y) = \bigcup \pi_0^{-1}(U)$. We need only prove that the set

$$U_{m_1,\ldots,m_r} := \bigcap_{i=1,\ldots r} F_{D_i}^{-1}(m_i) \cap \pi_0^{-1}(U)$$

is cylinder. In the notation of (2.3), if $J \not\subseteq \{1, \ldots, j_y\}$ then $D_J^\circ \cap U = \emptyset$ which forces $U_{m_1,\ldots,m_r} \subset \pi_0^{-1}(D_J^\circ \cap U)$ to be empty, and hence a cylinder set. We suppose therefore that $J \subseteq \{1, \ldots, j_y\}$, thus $|J| \leq n$ holds by (2.3).

The key observation is that by regarding arcs γ_u as n-tuples $(p_1(z), \ldots, p_n(z))$ of formal power series with zero constant terms, the equality $F_{D_i}(\gamma_u) = m_i$ is equivalent to a condition on the truncation of the power series $p_i(z)$ to degree m_i. Indeed, since D_i is cut out by $z_i = 0$ on U, it follows that $F_{D_i}(\gamma_u)$ equals the order of vanishing of $p_i(z)$ at $z = 0$. Thus $\gamma_u \in F_{D_i}^{-1}(m_i)$ if and only if the truncation of $p_i(z)$ to degree m_i is of the form $c_{m_i} z^{m_i}$, with $c_{m_i} \neq 0$. Truncating all n of the power series to degree $t := \max\{m_j \mid j \in J\}$ produces $n - |J|$ polynomials of degree t with zero constant term and, for each $j \in J$, a polynomial of the form

$$\pi_t(p_j(z)) = 0 + \cdots + 0 + c_{m_j} z^{m_j} + c_{(m_j+1)} z^{m_j+1} + \cdots + c_t z^t,$$

where $c_{m_j} \in \mathbb{C}^*$ and $c_k \in \mathbb{C}$ for all $k > m_j$. The space of all such n-tuples is isomorphic to $\mathbb{C}^{t(n-|J|)} \times (\mathbb{C}^*)^{|J|} \times \mathbb{C}^{t|J| - \sum_{j \in J} m_j}$, hence

$$U_{m_1,\ldots,m_r} = \pi_t^{-1}\left((U \cap D_J^\circ) \times \mathbb{C}^{tn - \sum_{j\in J} m_j} \times (\mathbb{C}^*)^{|J|}\right). \tag{2.4}$$

The set $(U \cap D_J^\circ) \times \mathbb{C}^{tn - \sum_{j\in J} m_j} \times (\mathbb{C}^*)^{|J|}$ is constructible, so $U_{m_1,\ldots,m_r}$ is a cylinder set. This completes the proof of the proposition. □

It is worth noting that $F_D^{-1}(\infty)$ is not a cylinder set. Indeed, suppose otherwise, so there exists a constructible subset $B_k \subseteq J_k(Y)$ for which $F_D^{-1}(\infty) = \pi_k^{-1}(B_k)$. Each arc $\gamma_y \in F_D^{-1}(\infty)$ is an n-tuple of power series, at least one of which is identically zero, whereas each $\gamma_y \in \pi_k^{-1}(B_k)$ is an n-tuple of power series whose terms of degree higher than k may take any complex value, a contradiction.

Proposition 2.6. *$F_D^{-1}(\infty)$ is a countable intersection of cylinder sets.*

Proof. Observe that

$$F_D^{-1}(\infty) = \bigcap_{k \in \mathbb{Z}_{\geq 0}} \pi_k^{-1} \pi_k\big(F_D^{-1}(\infty)\big) \tag{2.5}$$

because a power series is identically zero if and only if its truncation to degree k is the zero polynomial, for all $k \in \mathbb{Z}_{\geq 0}$. It is easy to see that the sets $\pi_k(F_D^{-1}(\infty)) \subset J_k(Y)$ are constructible. □

2.3. A measure μ on the space of formal arcs. In this section we define a measure μ on $J_\infty(Y)$ with respect to which the function F_D is measurable. The measure is not real-valued, so we begin by constructing the ring in which μ takes values.

[1]Finite intersections of cylinder sets are cylinder, so we could reduce to proving the result for $F_{D_i}^{-1}(m_i)$. However we require (2.4) in §2.4.

Definition 2.7. Let $\mathcal{V}_{\mathbb{C}}$ denote the category of complex algebraic varieties. The Grothendieck group of $\mathcal{V}_{\mathbb{C}}$ is the free Abelian group on the isomorphism classes $[V]$ of complex algebraic varieties modulo the subgroup generated by elements of the form $[V]-[V']-[V\setminus V']$ for a closed subset $V'\subseteq V$. The product of varieties induces a ring structure $[V]\cdot[V']=[V\times V']$, and the resulting ring denoted by $K_0(\mathcal{V}_{\mathbb{C}})$ is the *Grothendieck ring of complex algebraic varieties.* Let

$$[\quad]\colon \operatorname{Ob}\mathcal{V}_{\mathbb{C}}\longrightarrow K_0(\mathcal{V}_{\mathbb{C}})$$

denote the natural map sending V to its class $[V]$ in the Grothendieck ring. This map is universal with respect to maps which are additive on disjoint unions of constructible subsets, and which respect products.

Write $1:=[\text{point}]$ and $\mathbb{L}:=[\mathbb{C}]$ (see Appendix A: the class of $\mathbb{C}$ in $K_0(\mathcal{V}_{\mathbb{C}})$ corresponds to the Lefschetz motive $\mathbb{L}$). Then

$$[\mathbb{C}^*]=[\mathbb{C}]-[\{0\}]=\mathbb{L}-1.$$

Also, if $f\colon Y\to X$ is a locally trivial fibration w.r.t. the Zariski topology and F is the fibre over a (closed) point then $[Y]=[F\times X]$.

Definition 2.8. Let $K_0(\mathcal{V}_{\mathbb{C}})[\mathbb{L}^{-1}]:=S^{-1}K_0(\mathcal{V}_{\mathbb{C}})$ denote the ring of fractions of $K_0(\mathcal{V}_{\mathbb{C}})$ with respect to the multiplicative set $S:=\{1,\mathbb{L},\mathbb{L}^2,\dots\}$.

Definition 2.9. Recall that cylinder sets in $J_\infty(Y)$ are subsets $\pi_k^{-1}(B_k)\subset J_\infty(Y)$ for $k\in\mathbb{Z}_{\geq 0}$ and for $B_k\subseteq J_k(Y)$ a constructible subset. The function

$$\widetilde{\mu}\colon \Big\{\text{cylinder sets in } J_\infty(Y)\Big\}\longrightarrow K_0(\mathcal{V}_{\mathbb{C}})[\mathbb{L}^{-1}]$$

defined by

$$\pi_k^{-1}(B_k)\mapsto [B_k]\cdot\mathbb{L}^{-n(k+1)}$$

assigns a 'measure' to each cylinder set.

It is straightforward to show that

$$\widetilde{\mu}\Big(\textstyle\bigsqcup_{i=1}^{l}C_i\Big)=\displaystyle\sum_{i=1}^{l}\widetilde{\mu}(C_i)\quad\text{for cylinder sets } C_1,\dots,C_l$$

because the map $[\quad]$ introduced in Definition 2.7 is additive on disjoint unions of constructible sets. For this reason we call $\widetilde{\mu}$ a finitely additive measure.

Remark 2.10. Proposition 2.5 states that for $s\in\mathbb{Z}_{\geq 0}$, the level set $F_D^{-1}(s)$ is a cylinder set, and is therefore $\widetilde{\mu}$-measurable, i.e., $\widetilde{\mu}(F_D^{-1}(s))$ is well defined. However, F_D is not $\widetilde{\mu}$-measurable because $F_D^{-1}(\infty)$ is not cylinder. To proceed, we extend $\widetilde{\mu}$ to a measure μ with respect to which $F_D^{-1}(\infty)$ is measurable.

The following discussion is intended to motivate the definition of μ (see Definition 2.12 to follow). The set $J_\infty(Y)\setminus F_D^{-1}(\infty)$ is a countable disjoint union of cylinder sets

$$J_\infty(Y)\setminus\pi_0^{-1}\pi_0(F_D^{-1}(\infty))\sqcup\bigsqcup_{k\in\mathbb{Z}_{\geq 0}}\big(\pi_k^{-1}\pi_k(F_D^{-1}(\infty))\setminus\pi_{k+1}^{-1}\pi_{k+1}(F_D^{-1}(\infty))\big).\tag{2.6}$$

To see this, take complements in equation (2.5). Our goal is to extend $\widetilde{\mu}$ to a measure μ defined on the collection of countable disjoint unions of cylinder sets so that the set $J_\infty(Y) \setminus F_D^{-1}(\infty)$, and hence its complement $F_D^{-1}(\infty)$, is μ-measurable. One would like to define

$$\mu\Big(\bigsqcup_{i\in\mathbb{N}} C_i\Big) := \sum_{i\in\mathbb{N}} \mu(C_i) = \sum_{i\in\mathbb{N}} \widetilde{\mu}(C_i) \quad \text{for cylinder sets } C_1, \dots, C_l. \tag{2.7}$$

However, countable sums are not defined in $K_0(\mathcal{V}_\mathbb{C})[\mathbb{L}^{-1}]$. Furthermore, given a countable disjoint union $C = \bigsqcup_{i\in\mathbb{N}} C_i$, it is not clear a priori that $\mu(C)$ defined by formula (2.7) is independent of the choice of the C_i.

Kontsevich [**20**] solved both of these problems at once by completing the ring $K_0(\mathcal{V}_\mathbb{C})[\mathbb{L}^{-1}]$, thereby allowing appropriate countable sums, in such a way that the measure of the set $C = \bigsqcup_{i\in\mathbb{N}} C_i$ is independent of the choice of the C_i, assuming that $\mu(C_i) \to 0$ as $i \to \infty$.

Definition 2.11. Let R denote the completion of the ring $K_0(\mathcal{V}_\mathbb{C})[\mathbb{L}^{-1}]$ with respect to the filtration

$$\cdots \supseteq F^{-1}K_0(\mathcal{V}_\mathbb{C})[\mathbb{L}^{-1}] \supseteq F^{0}K_0(\mathcal{V}_\mathbb{C})[\mathbb{L}^{-1}] \supseteq F^{1}K_0(\mathcal{V}_\mathbb{C})[\mathbb{L}^{-1}] \supseteq \cdots$$

where for each $m \in \mathbb{Z}$, $F^m K_0(\mathcal{V}_\mathbb{C})[\mathbb{L}^{-1}]$ is the subgroup of $K_0(\mathcal{V}_\mathbb{C})[\mathbb{L}^{-1}]$ generated by elements of the form

$$[V] \cdot \mathbb{L}^{-i} \quad \text{for} \quad i - \dim V \geq m.$$

The natural completion map is denoted $\phi\colon K_0(\mathcal{V}_\mathbb{C})[\mathbb{L}^{-1}] \to R$.

By composing $\widetilde{\mu}$ with the natural completion map ϕ, we produce a finitely additive measure with values in the ring R, namely $\phi \circ \widetilde{\mu}$ (which we will also denote $\widetilde{\mu}$) that sends

$$\pi_k^{-1}(B_k) \mapsto \phi\left([B_k] \cdot \mathbb{L}^{-n(k+1)}\right).$$

Given a sequence of cylinder sets $\{C_i\}$ one may now ask whether or not $\widetilde{\mu}(C_i) \to 0$ as $i \to \infty$. We are finally in a position to define the measure μ on the space of formal arcs.

Definition 2.12. Let $\mathcal{C}$ denote the collection of countable disjoint unions of cylinder sets $\bigsqcup_{i\in\mathbb{N}} C_i$ for which $\widetilde{\mu}(C_i) \to 0$ as $i \to \infty$, together with the complements of such sets. Extend $\widetilde{\mu}$ to a measure

$$\mu\colon \mathcal{C} \longrightarrow R$$

by defining

$$\bigsqcup_{i\in\mathbb{N}} C_i \longrightarrow \sum_{i\in\mathbb{N}} \widetilde{\mu}(C_i).$$

It is nontrivial to show (see [**14**, §3.2] or [**1**, §6.18]) that this definition is independent of the choice of the C_i.

Proposition 2.13. *F_D is μ-measurable, and $\mu(F_D^{-1}(\infty)) = 0$.*

Proof. We prove that $F_D^{-1}(\infty)$ (in fact its complement) lies in $\mathcal{C}$. It's clear from (2.6) that we need only prove that $\mu(\pi_k^{-1}\pi_k(F_D^{-1}(\infty))) \to 0$ as $k \to \infty$. Lemma 2.14 below reveals that $\mu(\pi_k^{-1}\pi_k(F_D^{-1}(\infty))) \in \phi(F^{k+1}K_0(\mathcal{V}_\mathbb{C})[\mathbb{L}^{-1}])$ which, by the nature of the topology on R, tends to zero as k tends to infinity. This proves the first statement. Using (2.6) we calculate

$$\begin{aligned}(2.8)\quad \mu\big(J_\infty(Y)\setminus F_D^{-1}(\infty)\big) &= \widetilde{\mu}\big(J_\infty(Y)\setminus \pi_0^{-1}\pi_0(F_D^{-1}(\infty))\big)\\ &\quad + \sum_{k\in\mathbb{Z}_{\geq 0}} \widetilde{\mu}\left(\pi_k^{-1}\pi_k(F_D^{-1}(\infty))\setminus \pi_{k+1}^{-1}\pi_{k+1}(F_D^{-1}(\infty))\right).\end{aligned}$$

This equals $\mu(J_\infty(Y)) - \lim_{k\to\infty}\mu\left(\pi_k^{-1}\pi_k(F_D^{-1}(\infty))\right)$. By the above remark, this is simply $\mu(J_\infty(Y))$, so $\mu(F_D^{-1}(\infty)) = 0$ as required. □

Lemma 2.14. $\widetilde{\mu}(\pi_k^{-1}\pi_k(F_D^{-1}(\infty))) \in F^{k+1}K_0(\mathcal{V}_{\mathbb{C}})[\mathbb{L}^{-1}]$

Proof. It is enough to prove the result for a prime divisor D, since $F_D^{-1}(\infty)$ is the union of sets $F_{D_i}^{-1}(\infty)$. Choose coordinates on a chart U in which D is $(z_1 = 0)$. Each $\gamma_y \in F_D^{-1}(\infty)\cap\pi_0^{-1}(U)$ is an n-tuple $(p_1(z),\dots,p_n(z))$ of power series over $y \in U\cap D$ such that $p_1(z)$ is identically zero. Truncating these power series to degree k leaves $n-1$ polynomials of degree k with zero constant term, and the zero polynomial $\pi_k(p_1(z))$. The space of all such polynomials is isomorphic to $\mathbb{C}^{(n-1)k}$, so that $\pi_k(F_D^{-1}(\infty)\cap\pi_0^{-1}(U)) \simeq (U\cap D)\times\mathbb{C}^{(n-1)k}$. Thus $[\pi_k(F_D^{-1}(\infty)] = [D]\cdot[\mathbb{C}^{(n-1)k}]$ and

$$\begin{aligned}\widetilde{\mu}(\pi_k^{-1}\pi_k(F_D^{-1}(\infty)) &= [\pi_k(F_D^{-1}(\infty)]\cdot\mathbb{L}^{-n(k+1)}\\ &= [D]\cdot\mathbb{L}^{(n-1)k}\cdot\mathbb{L}^{-n(k+1)}\\ &= [D]\cdot\mathbb{L}^{-(n+k)}\end{aligned}$$

This lies in $F^{k+1}K_0(\mathcal{V}_{\mathbb{C}})[\mathbb{L}^{-1}]$ since D has dimension $n-1$. □

2.4. The motivic integral of a pair (Y, D).

Definition 2.15. Let Y be a complex manifold of dimension n, and choose an effective divisor $D = \sum_{i=1}^r a_i D_i$ on Y with only simple normal crossings. The *motivic integral of the pair* (Y, D) is

$$\int_{J_\infty(Y)} F_D\,\mathrm{d}\mu := \sum_{s\in\mathbb{Z}_{\geq 0}\cup\infty} \mu\left(F_D^{-1}(s)\right)\cdot\mathbb{L}^{-s}.$$

Since the set $F_D^{-1}(\infty)\subset J_\infty(Y)$ has measure zero (see Proposition 2.13), we need only integrate over $J_\infty(Y)\setminus F_D^{-1}(\infty)$, so we need only sum over $s\in\mathbb{Z}_{\geq 0}$.

We now show that the motivic integral converges in the ring R introduced in Definition 2.11. In doing so, we establish a user-friendly formula.

Theorem 2.16 (formula for the motivic integral). *Let Y be a complex manifold of dimension n and $D = \sum_{i=1}^r a_i D_i$ an effective divisor on Y with only simple normal crossings. The motivic integral of the pair (Y, D) is*

$$\int_{J_\infty(Y)} F_D\,\mathrm{d}\mu = \sum_{J\subseteq\{1,\dots,r\}} [D_J^\circ]\cdot\left(\prod_{j\in J}\frac{\mathbb{L}-1}{\mathbb{L}^{a_j+1}-1}\right)\cdot\mathbb{L}^{-n}$$

where we sum over all subsets $J\subseteq\{1,\dots,r\}$ including $J=\emptyset$.

PROOF. In the proof of Proposition 2.5 we cover Y by sets $\{U\}$ and prove that $\bigcap_{i=1,\dots r} F_{D_i}^{-1}(m_i) \cap \pi_0^{-1}(U)$ is a cylinder set of the form

$$\pi_t^{-1}\left((U \cap D_J^\circ) \times \mathbb{C}^{tn-\sum_{j\in J} m_j} \times (\mathbb{C}^*)^{|J|}\right).$$

Since the map $[\ \ ]$ introduced in Definition 2.7 is additive on a disjoint union of constructible subsets, take the union over the cover $\{U\}$ of Y to see that $\bigcap_{i=1,\dots r} F_{D_i}^{-1}(m_i) = \pi_t^{-1}(B_t)$ where

$$[B_t] = \left[D_J^\circ \times \mathbb{C}^{tn-\sum_{j\in J} m_j} \times (\mathbb{C}^*)^{|J|}\right] = [D_J^\circ] \cdot \mathbb{L}^{tn-\sum_{j\in J} m_j} \cdot (\mathbb{L}-1)^{|J|}.$$

Since $\mu\left(\pi_t^{-1}(B_t)\right) = [B_t] \cdot \mathbb{L}^{-(n+nt)}$, we have

$$\mu\left(\bigcap_{i=1,\dots r} F_{D_i}^{-1}(m_i)\right) = [D_J^\circ] \cdot \mathbb{L}^{-\sum_{j\in J} m_j} \cdot (\mathbb{L}-1)^{|J|} \cdot \mathbb{L}^{-n}.$$

Now use the partition (2.2) of $F_D^{-1}(s)$ to compute the motivic integral:

$$\begin{aligned}
&\sum_{s\in\mathbb{Z}_{\geq 0}} \mu\left(F_D^{-1}(s)\right) \cdot \mathbb{L}^{-s} \\
&\quad = \sum_{s\in\mathbb{Z}_{\geq 0}} \sum_{J\subset\{1,\dots,r\}} \sum_{(m_1,\dots,m_r)\in M_{J,s}} \mu\left(\bigcap_{i=1,\dots r} F_{D_i}^{-1}(m_i)\right) \cdot \mathbb{L}^{-\sum_{j\in J} a_j m_j} \\
&\quad = \sum_{s\in\mathbb{Z}_{\geq 0}} \sum_{J\subset\{1,\dots,r\}} \sum_{(m_1,\dots,m_r)\in M_{J,s}} [D_J^\circ] \cdot (\mathbb{L}-1)^{|J|} \cdot \mathbb{L}^{-n} \cdot \prod_{j\in J} \mathbb{L}^{-(a_j+1)m_j} \\
&\quad = \sum_{J\subset\{1,\dots,r\}} [D_J^\circ] \cdot \prod_{j\in J} \left((\mathbb{L}-1) \cdot \textstyle\sum_{m_j>0} \mathbb{L}^{-(a_j+1)m_j}\right) \cdot \mathbb{L}^{-n} \\
&\quad = \sum_{J\subset\{1,\dots,r\}} [D_J^\circ] \cdot \prod_{j\in J} \left((\mathbb{L}-1) \cdot \left(\frac{1}{1-\mathbb{L}^{-(a_j+1)}} - 1\right)\right) \cdot \mathbb{L}^{-n} \\
&\quad = \sum_{J\subset\{1,\dots,r\}} [D_J^\circ] \cdot \left(\prod_{j\in J} \frac{\mathbb{L}-1}{\mathbb{L}^{a_j+1}-1}\right) \cdot \mathbb{L}^{-n}.
\end{aligned}$$

□

WARNING 2.17. There is a small error in the proof of the corresponding result in Batyrev [**1**, §6.28] which leads to the omission of the $\mathbb{L}^{-n}$ term.

COROLLARY 2.18. *The motivic integral of the pair* (Y, D) *lies in the subring*

$$\phi\big(K_0(\mathcal{V}_\mathbb{C})[\mathbb{L}^{-1}]\big)\left[\left\{\frac{1}{\mathbb{L}^i - 1}\right\}_{i\in\mathbb{N}}\right]$$

of the ring R *introduced in Definition 2.11.*

2.5. The transformation rule for the integral. The discrepancy divisor $W := K_{Y'} - \alpha^* K_Y$ of a proper birational morphism $\alpha\colon Y' \to Y$ between smooth varieties is the divisor of the Jacobian determinant of α. The next result may therefore be viewed as the 'change of variables formula' for the motivic integral.

THEOREM 2.19. *Let* $\alpha\colon Y' \longrightarrow Y$ *be a proper birational morphism between smooth varieties and let* $W := K_{Y'} - \alpha^* K_Y$ *be the discrepancy divisor. Then*

$$\int_{J_\infty(Y)} F_D \, \mathrm{d}\mu = \int_{J_\infty(Y')} F_{\alpha^* D + W} \, \mathrm{d}\mu\,.$$

PROOF. Composition defines maps $\alpha_t\colon J_t(Y') \to J_t(Y)$ for each $t \in \mathbb{Z}_{\geq 0} \cup \infty$. An arc in Y which is not contained in the locus of indeterminacy of α^{-1} has a birational transform as an arc in Y'. In light of (2.1) and Proposition 2.13, α_∞ is bijective off a subset of measure zero.

The sets $F_W^{-1}(k)$, for $k \in \mathbb{Z}_{\geq 0}$, partition $J_\infty(Y') \setminus F_W^{-1}(\infty)$. Thus, for any $s \in \mathbb{Z}_{\geq 0}$ we have, modulo the set $F_W^{-1}(\infty)$ of measure zero, a partition

$$F_D^{-1}(s) = \bigsqcup_{k\in\mathbb{Z}_{\geq 0}} \alpha_\infty(C_{k,s}) \quad \text{where} \quad C_{k,s} := F_W^{-1}(k) \cap F_{\alpha^* D}^{-1}(s). \tag{2.9}$$

The set $C_{k,s}$ is cylinder and, since the image of a constructible set is constructible ([**24**, p. 72]), the set $\alpha_\infty(C_{k,s})$ is cylinder. Lemma 2.20 below states that $\mu\big(C_{k,s}\big) = \mu\big(\alpha_\infty(C_{k,s})\big) \cdot \mathbb{L}^k$. We use this identity and the partition (2.9) to calculate

$$\int_{J_\infty(Y)} F_D \,\mathrm{d}\mu = \sum_{k,s\in\mathbb{Z}_{\geq 0}} \mu\big(\alpha_\infty(C_{k,s})\big) \cdot \mathbb{L}^{-s} = \sum_{k,s\in\mathbb{Z}_{\geq 0}} \mu\big(C_{k,s}\big) \cdot \mathbb{L}^{-(s+k)}.$$

Set $s' := s+k$. Clearly $\bigsqcup_{0\leq k\leq s'} C_{k,s'-k} = F_{\alpha^*D+W}^{-1}(s')$. Substituting this into the above leaves

$$\int_{J_\infty(Y)} F_D \,\mathrm{d}\mu = \sum_{s'\in\mathbb{Z}_{\geq 0}} \mu\big(F_{\alpha^*D+W}^{-1}(s')\big) \cdot \mathbb{L}^{-s'} = \int_{J_\infty(Y')} F_{\alpha^*D+W} \,\mathrm{d}\mu,$$

as required. ☐

LEMMA 2.20. $\mu\big(C_{k,s}\big) = \mu\big(\alpha_\infty(C_{k,s})\big) \cdot \mathbb{L}^k$.

DISCUSSION OF PROOF. Both $C_{k,s}$ and $\alpha_\infty(C_{k,s})$ are cylinder sets so there exists $t \in \mathbb{Z}_{\geq 0}$ and constructible sets B_t' and B_t in $J_\infty(Y')$ and $J_\infty(Y)$ respectively such that the following diagram commutes:

$$\begin{array}{ccc} C_{k,s} \subset J_\infty(Y') & \xrightarrow{\alpha_\infty} & \alpha_\infty(C_{k,s}) \subset J_\infty(Y) \\ \pi_t \big\downarrow & & \big\downarrow \pi_t \\ B_t' \subset J_t(Y') & \xrightarrow{\alpha_t} & B_t \subset J_t(Y). \end{array}$$

We claim that the restriction of α_t to B_t' is a $\mathbb{C}^k$-bundle over B_t. It follows that $[B_t'] = [\mathbb{C}^k] \cdot [B_t]$ and we have

$$\mu\big(C_{k,s}\big) = [B_t'] \cdot \mathbb{L}^{-(n+nt)} = [B_t] \cdot \mathbb{L}^k \cdot \mathbb{L}^{-(n+nt)} = \mu\big(\alpha_\infty(C_{k,s})\big) \cdot \mathbb{L}^k$$

as required. The proof of the claim is a local calculation which is carried out in [**14**, Lemma 3.4(b)]. The key observation is that the order of vanishing of the Jacobian determinant of α at $\gamma_y \in C_{k,s}$ is $F_W(\gamma_y) = k$. ☐

DEFINITION 2.21. Let X denote a complex algebraic variety with at worst Gorenstein canonical singularities. The *motivic integral* of X is defined to be the motivic integral of the pair (Y, D), where $\varphi\colon Y \to X$ is any resolution of singularities for which the discrepancy divisor $D = K_Y - \varphi^* K_X$ has only simple normal crossings.

Note first that the discrepancy divisor D is effective because X has at worst Gorenstein canonical singularities. The crucial point however is that the motivic integral of (Y, D) is independent of the choice of resolution:

PROPOSITION 2.22. *Let $\varphi_1\colon Y_1 \longrightarrow X$ and $\varphi_2\colon Y_2 \longrightarrow X$ be resolutions of X with discrepancy divisors D_1 and D_2 respectively. Then the motivic integrals of the pairs (Y_1, D_1) and (Y_2, D_2) are equal.*

PROOF. Form a 'Hironaka hut'

$$\begin{array}{ccc} Y_0 & \xrightarrow{\psi_2} & Y_2 \\ \psi_1 \Big\downarrow & \searrow & \Big\downarrow \varphi_2 \\ Y_1 & \xrightarrow{\varphi_1} & X \end{array}$$

and let D_0 denote the discrepancy divisor of $\varphi_0 \colon Y_0 \longrightarrow X$. The discrepancy divisor of ψ_i is $D_0 - \psi_i^* D_i$. Indeed

$$\begin{aligned} K_{Y_0} &= \varphi_0^*(K_X) + D_0 = \psi_i^* \circ \varphi_i^*(K_X) + D_0 = \psi_i^*(K_{Y_i} - D_i) + D_0 \\ &= \psi_i^*(K_{Y_i}) + (D_0 - \psi_i^* D_i). \end{aligned}$$

The maps $\psi_i \colon Y_0 \longrightarrow Y_i$ are proper birational morphisms between smooth projective varieties so Theorem 2.19 applies:

$$\int_{J_\infty(Y_i)} F_{D_i} \,\mathrm{d}\mu = \int_{J_\infty(Y_0)} F_{\psi_i^* D_i + (D_0 - \psi_i^* D_i)} \,\mathrm{d}\mu = \int_{J_\infty(Y_0)} F_{D_0} \,\mathrm{d}\mu\,.$$

This proves the result. □

3. Hodge numbers via motivic integration

This section describes how the motivic integral of X gives rise to the so-called "stringy E-function" which encodes the Hodge–Deligne numbers of a resolution $Y \to X$.

3.1. Encoding Hodge–Deligne numbers. Deligne [**8, 9**] showed that the cohomology groups $H^k(X, \mathbb{Q})$ of a complex algebraic variety X carry a natural mixed Hodge structure. This consists of an increasing weight filtration

$$0 = W_{-1} \subseteq W_0 \subseteq \cdots \subseteq W_{2k} = H^k(X, \mathbb{Q})$$

on the rational cohomology of X and a decreasing Hodge filtration

$$H^k(X, \mathbb{C}) = F^0 \supseteq F^1 \supseteq \cdots \supseteq F^k \supseteq F^{k+1} = 0$$

on the complex cohomology of X such that the filtration induced by $F^\bullet$ on the graded quotient $\mathrm{Gr}_l^W H^k(X) := W_l / W_{l-1}$ is a pure Hodge structure of weight l. Thus

$$\mathrm{Gr}_l H^k(X) \otimes \mathbb{C} = F^p \mathrm{Gr}_l^W H^k(X) \oplus \overline{F^{l-p+1} \mathrm{Gr}_l^W H^k(X)}$$

where $F^p \mathrm{Gr}_l^W H^k(X)$ denotes the complexified image of $F^p \cap W_l$ in the quotient $W_l / W_{l-1} \otimes \mathbb{C}$. The integers

$$h^{p,q}\big(H^k(X, \mathbb{C})\big) := \dim_{\mathbb{C}} \Big(F^p \mathrm{Gr}_{p+q}^W H^k(X) \cap \overline{F^q \mathrm{Gr}_{p+q}^W H^k(X)} \Big)$$

are called the *Hodge–Deligne numbers* of X. For a smooth projective variety X over $\mathbb{C}$, $\mathrm{Gr}_l^W H^k(X, \mathbb{Q}) = 0$ unless $l = k$ in which case the Hodge–Deligne numbers are the classical Hodge numbers $h^{p,q}(X)$.

Danilov and Khovanskii [**12**] observed that cohomology with compact support $H_c^k(X, \mathbb{Q})$ also admits a mixed Hodge structure and they encode the corresponding Hodge–Deligne numbers in a single polynomial:

DEFINITION 3.1. The *E-polynomial* $E(X) \in \mathbb{Z}[u, v]$ of a complex algebraic variety X of dimension n is defined to be

$$E(X) := \sum_{0 \le p,q \le n} \sum_{0 \le k \le 2n} (-1)^k h^{p,q} \left(H_c^k(X, \mathbb{C})\right) u^p\, v^q.$$

Evaluating $E(X)$ at $u = v = 1$ produces the standard topological Euler number $e_c(X) = e(X)$.

THEOREM 3.2 ([**12**]). *Let X, Y be complex algebraic varieties. Then*

(i) *if $X = \bigsqcup X_i$ is stratified by a disjoint union of locally closed subvarieties then the E-polynomial is additive, i.e., $E(X) = \sum E(X_i)$.*
(ii) *the E-polynomial is multiplicative, i.e., $E(X \times Y) = E(X) \cdot E(Y)$.*
(iii) *if $f\colon Y \to X$ is a locally trivial fibration w.r.t. the Zariski topology and F is the fibre over a closed point then $E(Y) = E(F) \cdot E(X)$.*

See Danilov and Khovanskii [**12**] for a proof.

3.2. Kontsevich's theorem. Theorem 3.2 asserts that the map $E\colon \mathcal{V}_{\mathbb{C}} \longrightarrow \mathbb{Z}[u, v]$ associating to each complex variety X its E-polynomial is additive on a disjoint union of locally closed subvarieties, and satisfies $E(X \times Y) = E(X) \cdot E(Y)$. It follows from the universality of the map $[\ \ ]$ introduced in Definition 2.7 that E factors through the Grothendieck ring of algebraic varieties, inducing a function $E\colon K_0(\mathcal{V}_{\mathbb{C}}) \to \mathbb{Z}[u, v]$. By defining $E(\mathbb{L}^{-1}) := (uv)^{-1}$, this extends to[2]

$$E\colon K_0(\mathcal{V}_{\mathbb{C}})[\mathbb{L}^{-1}] \to \mathbb{Z}[u, v, (uv)^{-1}].$$

PROPOSITION 3.3. *The map E can be extended uniquely to the subring*

$$\phi\big(K_0(\mathcal{V}_{\mathbb{C}})[\mathbb{L}^{-1}]\big) \left[\left\{\frac{1}{\mathbb{L}^i - 1}\right\}_{i \in \mathbb{N}}\right]$$

of the ring R introduced in Definition 2.11.

PROOF. The kernel of the completion map $\phi\colon K_0(\mathcal{V}_{\mathbb{C}})[\mathbb{L}^{-1}] \to R$ is

$$\bigcap_{m \in \mathbb{Z}} F^m K_0(\mathcal{V}_{\mathbb{C}})[\mathbb{L}^{-1}]. \tag{3.1}$$

For $[V] \cdot \mathbb{L}^{-i} \in F^m K_0(\mathcal{V}_{\mathbb{C}})[\mathbb{L}^{-1}]$, the degree of the E-polynomial $E\left([V] \cdot \mathbb{L}^{-i}\right)$ is $2 \dim V - 2i \leq -2m$. The E-polynomial of an element Z in the intersection (3.1) must therefore be $-\infty$; that is, $E(Z) = 0$. Thus E annihilates $\ker \phi$ and hence factors through $\phi\left(K_0(\mathcal{V}_{\mathbb{C}})[\mathbb{L}^{-1}]\right)$. Defining $E(1/(\mathbb{L}^i - 1)) := 1/((uv)^i - 1)$ for $i \in \mathbb{N}$ establishes the result. $\square$

By Corollary 2.18 the motivic integral of the pair (Y, D) lies in the subring of Proposition 3.3. We now consider the image of the integral under E.

WARNING 3.4. As Warning 2.17 states, the derivation of the motivic integral in [**1**] contains a small error which leads to the omission of an $\mathbb{L}^{-n}$ term. However, in practise it is extremely convenient to omit this term (!). As a result, we define the stringy E-function to be the image under E of the motivic integral times $\mathbb{L}^n$. In short, our stringy E-function agrees with that in [**1**], even though our calculation of the motivic integral differs.

[2] One can use this function to define a finitely additive $\mathbb{Z}[u, v, (uv)^{-1}]$–valued measure $\widetilde{\mu}_E := E \circ \widetilde{\mu}$ on cylinder sets given by $\pi_k^{-1}(B_k) \to E(B_k) \cdot (uv)^{-n(k+1)}$, then construct the stringy E-function directly. This is the approach adopted by Batyrev [**1**, §6].

DEFINITION 3.5. Let X be a complex algebraic variety of dimension n with at worst Gorenstein canonical singularities. Let $\varphi\colon Y \to X$ be a resolution of singularities for which the discrepancy divisor $D = \sum_{i=1}^{r} a_i D_i$ has only simple normal crossings. The *stringy E-function* of X is

$$E_{\mathrm{st}}(X) := E\left(\int_{J_\infty(Y)} F_D \,\mathrm{d}\mu \cdot \mathbb{L}^n\right) = \sum_{J\subseteq\{1,\dots,r\}} E(D_J^\circ)\cdot\left(\prod_{j\in J}\frac{uv-1}{(uv)^{a_j+1}-1}\right),$$

where we sum over all subsets $J \subseteq \{1,\dots,r\}$ including $J = \emptyset$.

THEOREM 3.6 ([**20**]). *Let X be a complex projective variety with at worst Gorenstein canonical singularities. If X admits a crepant resolution $\varphi\colon Y \to X$ then the Hodge numbers of Y are independent of the choice of crepant resolution.*

PROOF. The discrepancy divisor $D = \sum_{i=1}^{r} a_i D_i$ of the crepant resolution $\varphi\colon Y \to X$ is by definition zero, so the motivic integral of X is the motivic integral of the pair $(Y, 0)$. Since each $a_i = 0$ it's clear that

$$E_{\mathrm{st}}(X) = \textstyle\sum_{J\subseteq\{1,\dots,r\}} E(D_J^\circ) = E(Y).$$

The stringy E-function is independent of the choice of the resolution φ. In particular, $E(Y) = E_{\mathrm{st}}(X) = E(Y_2)$ for $\varphi_2 : Y_2 \to X$ another crepant resolution. It remains to note that $E(Y)$ determines the Hodge–Deligne numbers of Y, and hence the Hodge numbers since Y is smooth and projective. □

4. Calculating the motivic integral

To perform nontrivial calculations of the stringy E-function we must choose varieties which admit no crepant resolution. A nice family of examples is provided by Gorenstein terminal cyclic quotient singularities.

4.1. Toric construction of cyclic quotient singularities. Consider the action of the cyclic group $G = \mathbb{Z}/r \subset \mathrm{GL}(n,\mathbb{C})$ generated by the diagonal matrix[3]

$$g = \mathrm{diag}\left(e^{2\pi i\alpha_1/r},\dots,e^{2\pi i\alpha_n/r}\right) \quad \text{with} \quad 0 \le \alpha_j < r,$$

where $i = \sqrt{-1}$. The quotient $\mathbb{C}^n/G$ is the cyclic quotient singularity of type $\frac{1}{r}(\alpha_1,\dots,\alpha_n)$. This fractional notation derives from the construction of $\mathbb{C}^n/G$ as an affine toric variety as we now describe (see Reid [**26**, §4] for more details).

Write $\overline{M} \cong \mathbb{Z}^n$ for the lattice of Laurent monomials in $x_1,\dots,x_n$, and $\overline{N}$ for the dual lattice with basis $e_1,\dots,e_n$. Let $\sigma = \mathbb{R}_{\ge 0}e_1 + \dots + \mathbb{R}_{\ge 0}e_n$ denote the positive orthant in $\overline{N}\otimes\mathbb{R}$ with dual cone $\sigma^\vee \subset \overline{M}\otimes\mathbb{R}$. The overlattice

$$N := \overline{N} + \mathbb{Z}\cdot\tfrac{1}{r}(\alpha_1,\dots,\alpha_n) \tag{4.1}$$

is dual to $M := \mathrm{Hom}(N,\mathbb{Z})$. A Laurent monomial in $x_1,\dots,x_n$ lies in the sublattice $M \subset \overline{M}$ if and only if it is invariant under the action of the group G. Restricting to Laurent monomials with only nonnegative powers leads to the equality $\mathbb{C}[x_1,\dots,x_n]^G = \mathbb{C}[\sigma^\vee \cap M]$, and hence

$$\mathbb{C}^n/G = \mathrm{Spec}\ \mathbb{C}[x_1,\dots,x_n]^G = \mathrm{Spec}\ \mathbb{C}[\sigma^\vee\cap M] =: U_\sigma.$$

[3]It is convenient to assume that $\gcd(r,\alpha_1,\dots,\widehat{\alpha_j},\dots,\alpha_n) = 1$ for all $j = 1,\dots,n$ to ensure that the group action is ‘small’. The notation $\widehat{\alpha_j}$ means that α_j is omitted.

In order to consider only Gorenstein terminal cyclic quotient singularities we impose certain restrictions on the type $\frac{1}{r}(\alpha_1,\dots,\alpha_n)$. Watanabe [**31**] showed that for a small subgroup $G\subset \mathrm{GL}(n,\mathbb{C})$, the quotient $\mathbb{C}^n/G$ is Gorenstein if and only if $G\subset \mathrm{SL}(n,\mathbb{C})$. Thus,

$$U_\sigma \text{ is Gorenstein } \iff \textstyle\sum_{j=1}^n \alpha_j \equiv 0 \mod r.$$

To determine when U_σ is terminal we recall the discrepancy calculation for cyclic quotients following Reid [**25, 26**] (see also Craw [**7**]). Write $\Box$ for the unit box in $N\otimes\mathbb{R}\cong\mathbb{R}^n$, i.e., the unit cell of the sublattice $\overline{N}\cong\mathbb{Z}^n$. Each element $g\in G\cong N/\overline{N}$ has a unique representative

$$v_g = \tfrac{1}{r(g)}\big(\alpha_1(g),\dots,\alpha_n(g)\big)\in N\cap\Box;$$

here v_g denotes both the vector in $N\otimes\mathbb{R}$ and the lattice point in N. For each primitive vector $v_g\in N\cap\Box$, the simplicial subdivision of σ at v_g determines a toric blow-up $\varphi\colon Y=X_\Sigma\to U_\sigma=X$ of the cyclic quotient. The exceptional divisor is the closed toric stratum[4] $V(\tau)\subset Y$, where τ is the ray with primitive generator v_g. Adjunction for the toric blow-up φ is

$$K_Y=\varphi^*K_X+\left(\frac{1}{r(g)}\sum_{j=1}^n\alpha_j(g)-1\right)V(\tau), \tag{4.2}$$

where the equality here denotes numerical equivalence. Therefore

$$U_\sigma \text{ is terminal } \iff \sum_{j=1}^n\frac{1}{r(g)}\alpha_j(g)>1 \text{ for each } g\in G.$$

That is, U_σ is terminal when every point $v_g\in N\cap\Box$ lies above the hyperplane $\sum x_i=1$.

Before proceeding to the examples we recall the simple formula that computes the E-polynomial of the toric variety X_Σ determined by a fan Σ in $N\otimes\mathbb{R}$.

PROPOSITION 4.1. *For a toric variety X_Σ of dimension n we have*

$$E\big(X_\Sigma\big)=\sum_{k=0}^n d_k\cdot\big(uv-1\big)^{n-k}, \tag{4.3}$$

where d_k is the number of cones of dimension k in Σ.

PROOF. The Hodge numbers of $\mathbb{P}^1$ are well known and, by Theorem 3.2, we compute $E(\mathbb{C}^*)=E(\mathbb{P}^1)-E(\{0\})-E(\{\infty\})=uv-1$. The E-polynomial is multiplicative so $E\big((\mathbb{C}^*)^{n-k}\big)=\big(uv-1\big)^{n-k}$. The action of the torus $\mathbb{T}^n\simeq(\mathbb{C}^*)^n$ on X_Σ induces a stratification of X_Σ into orbits of the torus action $O_\tau\cong(\mathbb{C}^*)^{n-\dim\tau}$, one for each cone $\tau\in\Sigma$. The result follows from Theorem 3.2(i). $\Box$

---

[4]The cones in Σ containing τ as a face define a fan denoted $\mathrm{Star}(\tau)$ in the vector space $N(\tau)\otimes\mathbb{R}$, where $N(\tau):=N/(\tau\cap N)$. The toric variety $V(\tau):=X_{\mathrm{Star}(\tau)}$ is the closure of the orbit O_τ. See Fulton [**16**, p. 52] for a nice picture.

4.2. The examples. For a finite subgroup $G \subset \mathrm{SL}(n,\mathbb{C})$ with $n = 2$ or 3, the Gorenstein quotient $\mathbb{C}^n/G$ admits a crepant resolution[5], so the stringy E-function of $\mathbb{C}^n/G$ is simply the E-polynomial of the crepant resolution. We therefore begin by considering 4-dimensional quotient singularities of type $\frac{1}{r}(1, r-1, a, r-a)$ with $\gcd(r,a) = 1$. Morrison and Stevens [**23**, Theorem 2.4(ii)] prove that these are the only Gorenstein terminal 4-fold cyclic quotient singularities.

Remark 4.2. In each example below we calculate both $E(Y)$ and $E_{\mathrm{st}}(\mathbb{C}^n/G)$ after resolving the singularity $\varphi\colon Y \to \mathbb{C}^n/G$. Note that $E(Y)$ is not equal to the E-polynomial of the exceptional fibre $D = \varphi^{-1}(\pi(0))$, for $\pi\colon \mathbb{C}^n \to \mathbb{C}^n/G$ the quotient map. Indeed,

$$E(Y) = E(Y \setminus D) + E(D) = (uv)^n - 1 + E(D).$$

The point is that the E-polynomial encodes the Hodge–Deligne numbers of compactly supported cohomology, yet

$$H^*_c(D,\mathbb{C}) = H^*(D,\mathbb{C}) \cong H^*(Y,\mathbb{C}) \neq H^*_c(Y,\mathbb{C});$$

the first equality holds because D is compact, and the isomorphism is induced by a deformation retraction of Y onto $D \subset Y$.

Example 4.3. Write $X = U_\sigma$ for the quotient singularity of type $\frac{1}{2}(1,1,1,1)$, i.e., for the quotient of $\mathbb{C}^4$ by the action of $G = \mathbb{Z}/2$, where the nontrivial element acts diaginally as -1. Add the ray τ generated by the vector $v = \frac{1}{2}(1,1,1,1)$ to the cone σ, then take the simplicial subdivision of σ. This determines a toric resolution $\varphi\colon Y \to X$ with a single exceptional divisor $D = V(\tau) \cong \mathbb{P}^3$. The discrepancy of D is 1 by (4.2). Using Proposition 4.1 we calculate

$$\begin{aligned} E(Y) &= E(Y \setminus D) + E(\mathbb{P}^3) \\ &= \big((uv)^4 - 1\big) + \big((uv)^3 + (uv)^2 + uv + 1\big) \\ &= (uv)^4 + (uv)^3 + (uv)^2 + uv. \end{aligned}$$

Compare this with the stringy E-function:

$$E_{\mathrm{st}}(X) = E(Y \setminus D) + E(\mathbb{P}^3) \cdot \frac{uv - 1}{(uv)^2 - 1} = (uv)^4 + (uv)^2.$$

Example 4.4. Write $X = U_\sigma$ for the quotient singularity of type $\frac{1}{3}(1,2,1,2)$. Add rays τ_1 and τ_2 generated by the vectors $v_1 = \frac{1}{3}(1,2,1,2)$ and $v_2 = \frac{1}{3}(2,1,2,1)$ respectively to the cone σ, then take the simplicial subdivision of σ. The resulting fan Σ is determined by its cross-section $\Delta_2 := \sigma \cap (\sum x_i = 2)$ illustrated in Figure 1.

There are eight 3-dimensional simplices in Δ_2 (four contain a face of the tetrahedron and four contain the edge joining v_1 to v_2). Each of these simplices determines a 4-dimensional cone in Σ which is generated by a basis of the lattice N, so $Y = X_\Sigma \to U_\sigma$ is a resolution. The union of all eight 3-dimensional simplices in Δ_2 contains eighteen faces, fifteen edges and six vertices. Write d_k for the number of cones of dimension k in Σ, so

$$d_4 = 8; \quad d_3 = 18; \quad d_2 = 15; \quad d_1 = 6; \quad d_0 = 1 \text{ (the origin in } N \otimes \mathbb{R}\text{).}$$

[5] The moduli space $Y = G\text{-Hilb}(\mathbb{C}^n)$ of G-clusters in $\mathbb{C}^n$ is a crepant resolution for $n = 2, 3$. See Bridgeland, King and Reid [**6**] for definitions and details.

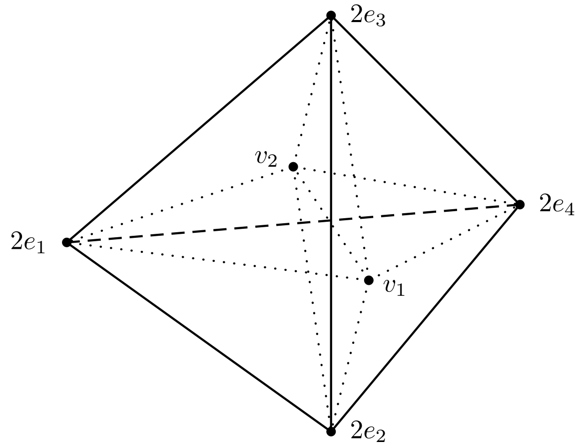


FIGURE 1. The simplex Δ_2 for $\frac{1}{3}(1,2,1,2)$

Apply Proposition 4.1 to compute

$$E(Y) = (uv)^4 + 2(uv)^3 + 3(uv)^2 + 2uv.$$

To compute $E_{\rm st}(X)$ observe that for $j = 1, 2$ the exceptional divisor $D_j := V(\tau_j)$ has discrepancy 1 by (4.2). Write $d_k(\tau_j)$ for the number of cones of dimension k in the fan $\mathrm{Star}(\tau_j)$ defining $V(\tau_j)$, so

$$d_3(\tau_j) = 6; \quad d_2(\tau_j) = 9; \quad d_1(\tau_j) = 5; \quad d_0(\tau_j) = 1 \text{ (the origin in } N(\tau_j)).$$

Proposition 4.1 gives

$$E(D_j) = (uv)^3 + 2(uv)^2 + 2(uv) + 1 \quad \text{for } j = 1, 2.$$

Similarly, the fan $\mathrm{Star}(\langle\tau_1, \tau_2\rangle)$ contains four faces, four edges and one vertex so Proposition 4.1 gives $E(D_1 \cap D_2) = (uv)^2 + 2(uv) + 1$. As a result

- $E(D^\circ_\emptyset) = E(Y \setminus (D_1 \cup D_2)) = (uv)^4 - 1$.
- $E(D^\circ_{\{1\}}) = E(D^\circ_{\{2\}}) = E(D_j) - E(D_1 \cap D_2) = (uv)^3 + (uv)^2$.
- $E(D^\circ_{\{1,2\}}) = E(D_1 \cap D_2) = (uv)^2 + 2(uv) + 1$.

Now compute the stringy E-function using formula (3.2):

$$\begin{aligned} E_{\rm st}(X) &= (uv)^4 - 1 + E(D^\circ_{\{1\}}) \cdot \left(\frac{uv-1}{(uv)^2-1}\right) + E(D^\circ_{\{2\}}) \cdot \left(\frac{uv-1}{(uv)^2-1}\right) \\ &\quad + E(D^\circ_{\{1,2\}}) \cdot \left(\frac{uv-1}{(uv)^2-1}\right)^2 \\ &= (uv)^4 + 2(uv)^2. \end{aligned}$$

EXAMPLE 4.5. Let $X = U_\sigma$ denote the cyclic quotient singularity of type $\frac{1}{4}(1,3,1,3)$. Add rays τ_1, τ_2 and τ_3 generated by the vectors $v_1 = \frac{1}{4}(1,3,1,3)$ and $v_2 = \frac{1}{4}(2,2,2,2)$ and $v_3 = \frac{1}{4}(3,1,3,1)$ respectively to the cone σ, then take the simplicial subdivision of σ. The cross-section Δ_2 of the resulting fan Σ has three colinear points in the interior of the tetrahedron but is otherwise similar to that shown in Figure 1. There are twelve 3-dimensional simplices in Δ_2 containing 26 faces, 20 edges and 7 vertices. Proposition 4.1 calculates

$$E(Y) = (uv)^4 + 3(uv)^3 + 5(uv)^2 + 3uv.$$

For $j = 1, 2, 3$ the divisors $D_j := V(\tau_j)$ have discrepancy 1 by (4.2). Following the method of Example 4.4 we calculate

$$E(D_1) = E(D_3) = (uv)^3 + 2(uv)^2 + 2(uv) + 1$$

and $E(D_1 \cap D_2) = E(D_2 \cap D_3) = (uv)^2 + 2(uv) + 1$. To compute the E-polynomial of D_2 observe that

$$d_3(\tau_2) = 8; \quad d_2(\tau_2) = 12; \quad d_1(\tau_2) = 6; \quad d_0(\tau_2) = 1 \text{ (the origin in } N(\tau_2)),$$

where $d_k(\tau_2)$ denotes the number of cones of dimension k in $\mathrm{Star}(\tau_2)$. It follows from Proposition 4.1 that

$$E(D_2) = (uv)^3 + 3(uv)^2 + 3(uv) + 1.$$

Finally, since $D_1 \cap D_3 = \emptyset$ we have $E(D_1 \cap D_3) = E(D_1 \cap D_2 \cap D_3) = 0$. As a result

- $E(D^\circ_\emptyset) = E(Y \setminus (D_1 \cup D_2 \cup D_3)) = (uv)^4 - 1$.
- $E(D^\circ_{\{1\}}) = E(D^\circ_{\{3\}}) = (uv)^3 + (uv)^2$.
- $E(D^\circ_{\{2\}}) = (uv)^3 + (uv)^2 - (uv) - 1$.
- $E(D^\circ_{\{1,2\}}) = E(D^\circ_{\{2,3\}}) = (uv)^2 + 2(uv) + 1$.
- $E(D^\circ_{\{1,3\}}) = E(D^\circ_{\{1,2,3\}}) = 0$.

Apply formula (3.2) to compute $E_{\mathrm{st}}(X) = (uv)^4 + 3(uv)^2$.

Remark 4.6. The above examples feature only exceptional divisors with discrepancy 1. To obtain examples of Gorenstein terminal cyclic quotient singularities which admit resolutions containing divisors having discrepancy larger than one we must work in dimension higher than four.

Example 4.7. Let $X = U_\sigma$ denote the cyclic quotient singularity of type $\frac{1}{r}(1, 1, 1, \ldots, 1)$ where $n := \dim X = kr$ for some $k \in \mathbb{Z}$ (by assuming that r divides n we ensure that X is Gorenstein). Add a single ray τ generated by the vector $v_1 = \frac{1}{r}(1, 1, 1, \ldots, 1)$ to the cone σ, then take the simplicial subdivision of σ. This determines a toric resolution $\varphi\colon Y \to X$ with a single exceptional divisor $D = V(\tau) \cong \mathbb{P}^{n-1}$. The discrepancy of D is $k - 1$ by (4.2). Using Proposition 4.1 we calculate

$$\begin{aligned} E(Y) &= E(Y \setminus D) + E(\mathbb{P}^{n-1}) \\ &= \big((uv)^n - 1\big) + \big((uv)^{n-1} + (uv)^{n-2} + \cdots + uv + 1\big) \\ &= (uv)^n + (uv)^{n-1} + \cdots + uv. \end{aligned}$$

Compare this with the stringy E-function:

$$\begin{aligned} E_{\mathrm{st}}(X) &= E(Y \setminus D) + E(\mathbb{P}^{n-1}) \cdot \frac{uv - 1}{(uv)^k - 1} \\ &= (uv)^n + (uv)^{n-k} + \cdots + (uv)^{2k} + (uv)^k. \end{aligned}$$

Example 4.8. Let $X = U_\sigma$ denote the cyclic quotient singularity of type $\frac{1}{3}(1, 2, 1, 2, 1, 2)$ (compare Example 4.4). Add rays τ_1 and τ_2 generated by the vectors $v_1 = \frac{1}{3}(1, 2, 1, 2, 1, 2)$ and $v_2 = \frac{1}{3}(2, 1, 2, 1, 2, 1)$ respectively to the cone σ, then take the simplicial subdivision of σ. Both v_1 and v_2 lie in the simplex

$\Delta_3 := \sigma \cap (\sum x_i = 0)$ of the resulting fan Σ so the corresponding exceptional divisors D_1 and D_2 each have discrepancy 2 by (4.2). The cross-section Δ_3 is difficult to draw (it is 5-dimensional!) but, using Figure 1 as a guide, one can show that

$$d_6 = 15; \quad d_5 = 48; \quad d_4 = 68; \quad d_3 = 56; \quad d_2 = 28; \quad d_1 = 8; \quad d_0 = 1,$$

where d_k denotes the number of cones of dimension k in Σ. Hence

$$E(Y) = (uv)^6 + 2(uv)^5 + 3(uv)^4 + 4(uv)^3 + 3(uv)^2 + 2(uv).$$

As with Example 4.4, for $j = 1, 2$ write $d_k(\tau_j)$ for the number of cones of dimension k in $\mathrm{Star}(\tau_j)$, so

$$d_5(\tau_j) = 12;\ d_4(\tau_j) = 30;\ d_3(\tau_j) = 34;\ d_2(\tau_j) = 21;\ d_1(\tau_j) = 7;\ d_0(\tau_j) = 1.$$

Proposition 4.1 gives

$$E(D_j) = (uv)^5 + 2(uv)^4 + 3(uv)^3 + 3(uv)^2 + 2(uv) + 1 \quad \text{for } j = 1, 2.$$

Similarly, counting simplices in the fan $\mathrm{Star}(\langle \tau_1, \tau_2 \rangle)$ gives

$$E(D_1 \cap D_2) = (uv)^4 + 2(uv)^3 + 3(uv)^2 + 2(uv) + 1.$$

Now compute the stringy E-function using formula (3.2):

$$\begin{aligned} E_{\mathrm{st}}(X) &= (uv)^6 - 1 + E(D^\circ_{\{1\}}) \cdot \left(\frac{uv - 1}{(uv)^3 - 1} \right) + E(D^\circ_{\{2\}}) \cdot \left(\frac{uv - 1}{(uv)^3 - 1} \right) \\ &\quad + E(D^\circ_{\{1,2\}}) \cdot \left(\frac{uv - 1}{(uv)^3 - 1} \right)^2 \\ &= (uv)^6 + 2(uv)^3. \end{aligned}$$

Remark 4.9. The stringy E-function of a Gorenstein canonical quotient singularity $\mathbb{C}^n/G$ can be calculated in terms of the representation theory of the finite subgroup $G \subset \mathrm{SL}(n, \mathbb{C})$ using a simple formula due to Batyrev [**3, 4**] (see also Denef and Loeser [**15**]). To state the formula, note that each $g \in G$ is conjugate to a diagonal matrix

$$g = \mathrm{diag}\left(e^{2\pi i \alpha_1(g)/r(g)}, \ldots, e^{2\pi i \alpha_n(g)/r(g)} \right) \quad \text{with} \quad 0 \le \alpha_j(g) < r(g), \tag{4.4}$$

where $r(g)$ is the order of g and $i = \sqrt{-1}$. To each conjugacy class $[g]$ of the group G we associate an integer in the range $0 \le \mathrm{age}[g] \le n - 1$ defined by

$$\mathrm{age}[g] := \frac{1}{r(g)} \sum_{j=1}^{n} \alpha_j(g).$$

In particular, for the diagonal action introduced in §4.1, the age grading on the cyclic group G corresponds to the slicing of the unit box $\square \subset \overline{N}_{\mathbb{R}} \cong \mathbb{R}^n$ into polytopes $\Delta_k := \sigma \cap (\sum x_i = k)$ for $k = 0, \ldots, n-1$. The formula for the stringy E-function of the quotient $\mathbb{C}^n/G$ is simply

$$E_{\mathrm{st}}(\mathbb{C}^n/G) = \sum_{[g] \in \mathrm{Conj}(G)} (uv)^{n - \mathrm{age}[g]}, \tag{4.5}$$

where we sum over the conjugacy classes of G. For example, the nontrivial element g of the group $G = \mathbb{Z}/2$ acting on $\mathbb{C}^4$ in Example 4.3 has age two because $v_g = \frac{1}{2}(1,1,1,1) \in \Delta_2$. Formula (4.5) gives $E_{\mathrm{st}}(\mathbb{C}^4/G) = (uv)^4 + (uv)^2$ as shown in §4.2. Observe that the same holds for the other examples of §4.2.

5. The McKay correspondence

In this section we discuss formula (4.5) in the wider context of the cohomological McKay correspondence.

5.1. The McKay correspondence conjecture. Motivated by string theory, Dixon et al. [**10**] introduced the *orbifold Euler number* for a finite group G acting on a manifold M. This number can be written in the form

$$e(M,G) = \sum_{[g]\in\mathrm{Conj}(G)} e\big(M^g/C(g)\big),$$

where the sum runs over the conjugacy classes of G, e denotes the topological Euler number, M^g is the fixed point set of g and $C(g)$ is the centraliser of g (this version of the formula is due to Hirzebruch and Höfer [**18**]). Observe that the orbifold Euler number is the standard topological Euler number of the disjoint union

$$\widetilde{M/G} = \bigsqcup_{[g]\in\mathrm{Conj}(G)} M^g/C(g).$$

Dixon et al. [**11**] formulated what became known as the "physicists' Euler number conjecture":

Conjecture 5.1. *If M/G is a Gorenstein Calabi–Yau variety which admits a crepant resolution $Y \to M/G$ then $e(Y) = e(M,G)$.*

Hirzebruch and Höfer [**18**] observed that for a finite subgroup $G \subset \mathrm{U}(n)$ acting on $M = \mathbb{C}^n$, the orbifold Euler number is equal to the number of conjugacy classes of G because every fixed point set M^g is contractible. For $G \subset \mathrm{SU}(2,\mathbb{C})$, the classical McKay correspondence states that the second Betti number $b_2(Y)$ of the minimal resolution $Y \to \mathbb{C}^2/G$ equals the number of nontrivial conjugacy classes of G. As a result, the equality

$$e(Y) = b_2(Y) + 1 = \#\Big\{\text{conjugacy classes of } G\Big\} = e(\mathbb{C}^2, G) \tag{5.1}$$

can be viewed as a version of the McKay correspondence. Inspired by this observation, Reid [**27**] formulated Conjecture 1.2.

5.2. The orbifold E-function. Following the introduction of the orbifold Euler number, Vafa [**30**] and Zaslow [**32**] considered the *orbifold Hodge numbers*

$$h^{p,q}(M,G) := \sum_{[g]\in\mathrm{Conj}(G)} \sum_{i=1}^{m(g)} \dim_{\mathbb{C}} H_c^{p-\mathrm{age}[g],q-\mathrm{age}[g]}(M_i(g)/C(g)), \tag{5.2}$$

where $M^g = M_1(g) \cup \cdots \cup M_{m(g)}(g)$ are the smooth connected components of the fixed-point set. These numbers are the standard (compactly supported) Hodge numbers of $\widetilde{M/G}$ shifted according to the age[6] of the appropriate conjugacy class.

It is natural to introduce the orbifold analogue of the E-polynomial.

[6] In this context, the *age* of a conjugacy class is called the *Fermionic shift number*.

DEFINITION 5.2. The *orbifold E-function* of the pair (M, G) is

$$\begin{aligned} E_{\mathrm{orb}}(M,G) &:= \sum_{p,q} (-1)^{p+q}\, h^{p,q}(M,G)\, u^p\, v^q \\ &= \sum_{[g]\in\mathrm{Conj}(G)} \sum_{i=1}^{m(g)} E(M_i(g)/C(g)) \cdot (uv)^{\mathrm{age}[g]}, \end{aligned}$$

where E denotes the standard E-polynomial. When $E_{\mathrm{orb}}(M, G)$ is evaluated at $u = v = 1$ we produce the orbifold Euler number $e(M, G)$.

REMARK 5.3. Note that the second formula for $E_{\mathrm{orb}}(M, G)$ given in Definition 5.2 follows from a straightforward substitution. In the special case $M = \mathbb{C}^n$ and $G \subset \mathrm{GL}(n, \mathbb{C})$, the fixed-point set M^g is an affine subspace of dimension equal to $n - \mathrm{rank}(g - \mathrm{id}) = n - \mathrm{age}[g] - \mathrm{age}[g^{-1}]$. As a result,

$$E_{\mathrm{orb}}(\mathbb{C}^n, G) = \sum_{[g]\in\mathrm{Conj}(G)} (uv)^{n-\mathrm{age}[g^{-1}]} = \sum_{[g]\in\mathrm{Conj}(G)} (uv)^{n-\mathrm{age}[g]},$$

where the final equality follows from summing over conjugacy classes $[g^{-1}]$.

5.3. McKay correspondence via motivic integration. Batyrev proved Conjecture 1.2 by first establishing that the stringy E-function coincides with the orbifold E-function. That is, he proved formula (4.5):

THEOREM 5.4 ([**3**]). *Let $G \subset \mathrm{SL}(n, \mathbb{C})$ be a finite subgroup. Then*

$$E_{\mathrm{st}}(\mathbb{C}^n/G) = E_{\mathrm{orb}}(\mathbb{C}^n, G) = \sum_{[g]\in\mathrm{Conj}(G)} (uv)^{n-\mathrm{age}[g]},$$

where the sum runs over conjugacy classes of G.

PROOF OF THE ABELIAN CASE. Choose coordinates on $\mathbb{C}^n$ so that every matrix $g \in G$ takes the form given in (4.4). The cyclic construction of §4.1 can be adapted to show that $\mathbb{C}^n/G$ is the toric variety U_σ corresponding to the cone $\sigma = \mathbb{R}_{\geq 0}e_1 + \cdots + \mathbb{R}_{\geq 0}e_n$ and the lattice

$$N := \mathbb{Z}e_1 + \cdots + \mathbb{Z}e_n + \sum_{g\in G} \mathbb{Z}\cdot v_g, \tag{5.3}$$

for $v_g = \frac{1}{r(g)}\big(\alpha_1(g), \ldots, \alpha_n(g)\big)$. Now, U_σ is Gorenstein because $G \subset \mathrm{SL}(n, \mathbb{C})$, so there exists a continuous linear function $\psi_K \colon N \otimes \mathbb{R} \to \mathbb{R}_{\geq 0}$ satisfying $\psi(e_i) = 1$ for $i = 1, \ldots, n$. A straightforward computation due to Batyrev [**1**, Theorem 4.3] gives

$$E_{\mathrm{st}}(U_\sigma) = (uv - 1)^n \sum_{v\in N\cap\sigma} (uv)^{-\psi_K(v)}. \tag{5.4}$$

For $v_g \in \Box$, $\psi_K(v_g) = \mathrm{age}(g)$. In fact, for any lattice point $v \in N \cap \sigma$ we have $\psi_K(v) = k \iff v \in \Delta_k = \sigma \cap (\sum x_i = k)$. Thus for each $v \in N \cap \sigma$ there exist unique $v_g \in \Box$ and $(x_1, \ldots, x_n) \in \mathbb{Z}^n_{\geq 0}$ such that v is the translate of v_g by $(x_1, \ldots, x_n)$, and $\psi_K(v) = \mathrm{age}(g) + \sum_{i=1}^n x_i$.

Covering the positive orthant σ by translations of $\Box$ gives

$$\begin{aligned}\sum_{v\in N\cap\sigma}(uv)^{-\psi_K(v)} &= \sum_{v_g\in\Box}\sum_{(x_1,\dots,x_n)\in\mathbb{Z}^n_{\geq 0}}(uv)^{-\operatorname{age}(g)-\sum x_i}\\ &= \sum_{v_g\in\Box}(uv)^{-\operatorname{age}(g)}\prod_{i=1}^{n}\sum_{x_i\in\mathbb{Z}_{\geq 0}}(uv)^{-x_i}\\ &= \sum_{g\in G}(uv)^{-\operatorname{age}(g)}\prod_{i=1}^{n}\frac{1}{1-(uv)^{-1}}.\end{aligned}$$

Substituting this into (5.4) gives

$$E_{\mathrm{st}}(U_\sigma)=\sum_{g\in G}(uv)^{-\operatorname{age}(g)}\prod_{i=1}^{n}\frac{uv-1}{1-(uv)^{-1}}=\sum_{g\in G}(uv)^{n-\operatorname{age}(g)}.$$

This proves the theorem for a finite Abelian subgroup $G\subset\mathrm{SL}(n,\mathbb{C})$. $\Box$

COROLLARY 5.5 (strong McKay correspondence). *Let $G\subset\mathrm{SL}(n,\mathbb{C})$ be a finite subgroup and suppose that the quotient $X=\mathbb{C}^n/G$ admits a crepant resolution $\varphi\colon Y\to X$. The nonzero Betti numbers of Y are*

$$\dim_{\mathbb{C}}H^{2k}(Y,\mathbb{C})=\#\Big\{\textit{age } k \textit{conjugacy classes of } G\Big\}$$

for $k=0,\dots,n-1$. In particular, the Euler number of Y equals the number of conjugacy classes of G.

PROOF. The Hodge structure in $H^i_c(Y,\mathbb{Q})$ is pure for each i and Poincaré duality $H^{2n-i}_c(Y,\mathbb{C})\otimes H^i(Y,\mathbb{C})\to H^{2n}_c(Y,\mathbb{C})$ respects the Hodge structure, so it is enough to show that the only nonzero Hodge–Deligne numbers of the compactly supported cohomology of Y are

$$h^{n-k,n-k}\big(H^{2n-2k}_c(Y,\mathbb{C})\big)=\#\Big\{\text{age } k \text{ conjugacy classes of } G\Big\}.$$

Now $h^{n-k,n-k}\big(H^{2n-2k}_c(Y,\mathbb{C})\big)$ is the coefficient of $(uv)^{n-k}$ in the E-polynomial of Y. Moreover, the resolution $\varphi\colon Y\to X$ is crepant so $E(Y)=E_{\mathrm{st}}(X)$ and the result follows from Theorem 5.4. $\Box$

Appendix A. The motivic nature of the integral

In this appendix we investigate the motivic nature of the integral and justify the notation $\mathbb{L}$ for the class of the complex line $\mathbb{C}$ in the Grothendieck ring of algebraic varieties.

The category $\mathcal{M}_{\mathbb{C}}$ of Chow motives over $\mathbb{C}$ is the category (see [**29**]) whose objects are triples (X,p,m), where X is a smooth complex projective variety of dimension d, p is an element of the Chow ring $A^d(X\times X)$ satisfying $p^2=p$ and $m\in\mathbb{Z}$. If (X,p,m) and (Y,q,n) are motives then

$$\operatorname{Hom}_{\mathcal{M}_{\mathbb{C}}}\left((X,p,m),(Y,q,n)\right)=qA^{d+n-m}(X,Y)p$$

where composition of morphisms is given by composition of correspondences. The category $\mathcal{M}_{\mathbb{C}}$ is additive, $\mathbb{Q}$-linear and pseudo-Abelian. Tensor product of motives is defined as $(X,p,m)\otimes(Y,q,n)=(X\times Y,p\otimes q,m+n)$. There is a functor

$$h\colon\mathcal{V}^{\circ}_{\mathbb{C}}\to\mathcal{M}_{\mathbb{C}}$$

from the (opposite) category of algbraic varieties over $\mathbb{C}$ to the category of Chow motives over $\mathbb{C}$, sending X to $(X, \Delta_X, 0)$, the Chow motive of X, where the diagonal $\Delta_X \subset X \times X$ is the identity in $A^*(X \times X)$. The motive of a point $1 = h(\operatorname{Spec}\mathbb{C})$ is the identity with respect to tensor product. The Lefschetz motive $\mathbb{L}$ is defined implicitly via the relation $h(\mathbb{P}^1_{\mathbb{C}}) = 1 \oplus \mathbb{L}$.

DEFINITION A.1. The Grothendieck group of $\mathcal{M}_{\mathbb{C}}$ is the free Abelian group generated by isomorphism classes of objects in $\mathcal{M}_{\mathbb{C}}$ modulo the subgroup generated by elements of the form $[(X,p,m)] - [(Y,q,n)] - [(Z,r,k)]$ whenever the relation $(X,p,m) \simeq (Y,q,n) \oplus (Z,r,k)$ holds. Tensor product of motives induces a ring structure and the resulting ring, denoted $K_0(\mathcal{M}_{\mathbb{C}})$, is the Grothendieck ring of Chow Motives (over $\mathbb{C}$).

Gillet and Soulé [**17**] exhibit a map $M\colon \mathcal{V}_{\mathbb{C}} \longrightarrow K_0(\mathcal{M}_{\mathbb{C}})$ sending a smooth projective variety X to the class $[h(X)]$ of the motive of X. Furthermore the map is additive on disjoint unions of locally closed subsets and satisfies $M(X \times Y) = M(X) \cdot M(Y)$.

We now play the same game as we did in §3. That is, M factors through $K_0(\mathcal{V}_{\mathbb{C}})$ inducing

$$M\colon K_0(\mathcal{V}_{\mathbb{C}}) \longrightarrow K_0(\mathcal{M}_{\mathbb{C}}).$$

Observe that the image of $[\mathbb{C}]$ under M is the class of the Lefschetz motive $\mathbb{L}$; this explains why we use the notation $\mathbb{L}$ to denote the class of $\mathbb{C}$ in $K_0(\mathcal{V}_{\mathbb{C}})$ in § 2. Sending $\mathbb{L}^{-1} \in K_0(\mathcal{V}_{\mathbb{C}})[\mathbb{L}^{-1}]$ to $\mathbb{L}^{-1} \in K_0(\mathcal{M}_{\mathbb{C}})$ produces a map

$$M\colon K_0(\mathcal{V}_{\mathbb{C}})[\mathbb{L}^{-1}] \longrightarrow K_0(\mathcal{M}_{\mathbb{C}}).$$

It is unknown whether or not M annihilates the kernel of the natural completion map $\phi\colon K_0(\mathcal{V}_{\mathbb{C}})[\mathbb{L}^{-1}] \to R$. Denef and Loeser conjecture that it does (see [**13**, Remark 1.2.3]). If this is true, extend M to a ring homomorphism

$$M_{\mathrm{st}}\colon \phi\left(K_0(\mathcal{V}_{\mathbb{C}})[\mathbb{L}^{-1}]\right)\left[\left\{\frac{1}{\mathbb{L}^i - 1}\right\}_{i\in\mathbb{N}}\right] \longrightarrow K_0(\mathcal{M}_{\mathbb{C}})\left[\left\{\frac{1}{\mathbb{L}^i - 1}\right\}_{i\in\mathbb{N}}\right]$$

such that the image of $[D_J^\circ]$ under M_{st} is equal to $M(D_J^\circ)$.

DEFINITION A.2. Let X denote a complex algebraic variety with at worst Gorenstein canonical singularities and let $\varphi\colon Y \to X$ be any resolution of singularities for which the discrepancy divisor $D = \sum a_i D_i$ has only simple normal crossings. The *stringy motive* of X is

$$M_{\mathrm{st}}(X) := M_{\mathrm{st}}\left(\int_{J_\infty(Y)} F_D \,\mathrm{d}\mu \cdot \mathbb{L}^n\right) = \sum_{J\subseteq\{1,\dots,r\}} M(D_J^\circ)\cdot\left(\prod_{j\in J}\frac{\mathbb{L}-1}{\mathbb{L}^{a_j+1}-1}\right),$$

where we sum over all subsets $J \subseteq \{1,\dots,r\}$ including $J = \emptyset$. As with the definition of the stringy E-function (see Definition 3.5) we multiply by $\mathbb{L}^n$ for convenience.

References

1. V. Batyrev, *Stringy Hodge numbers of varieties with Gorenstein canonical singularities*, Integrable systems and algebraic geometry (Kobe/Kyoto, 1997), (1998), pp. 1–32.
2. V. Batyrev, *Birational Calabi-Yau n-folds have equal Betti numbers*, New trends in algebraic geometry, Klaus Hulek et al. (editors), CUP, pages 1–11, (1999).
3. V. Batyrev, *Non-Archimedean integrals and stringy Euler numbers of log-terminal pairs*, J. Eur. Math. Soc. **1** (1999), 5–33.

4. V. Batyrev, *Canonical Abelianization of finite group actions* (correction to [Bat99b]), Preprint math.AG/0009043, (2000).
5. V. Batyrev and D. Dais, *Strong McKay correspondence, string-theoretic Hodge numbers and mirror symmetry*, Topology **35** (1996), 901–929.
6. T. Bridgeland, A. King, and M. Reid, *The McKay correspondence as an equivalence of derived categories*, J. Amer. Math. Soc. **14** (2001), 535–554.
7. A. Craw, Notes from a course on Toric Geometry at the University of Utah, available from `www.math.utah.edu/~craw/math7800`, (2002).
8. P. Deligne, *Théorie de Hodge II*, Publ. Math. IHES **40** (1971), 5–57.
9. P. Deligne, *Théorie de Hodge III*, Publ. Math. IHES **44** (1974), 5–77.
10. L. Dixon, J. Harvey, C. Vafa, and E. Witten, *Strings on Orbifolds I*, Nuclear Physics B **261** (1985), 678–686.
11. L. Dixon, J. Harvey, C. Vafa, and E. Witten, *Strings on Orbifolds II*, Nuclear Physics B **274** (1986), 285–314.
12. V. Danilov and A. Khovanskii, *Newton polyhedra and an algorithm for computing Hodge–Deligne numbers*, Math. USSR Izvestiya **29** (1987), 279–298.
13. J. Denef and F. Loeser, *Motivic Igusa zeta functions*, J. Alg. Geom. **7** (1998), 505–537.
14. J. Denef and F. Loeser, *Germs of arcs on singular algebraic varieties and motivic integration*, Invent. Math. **135** (1999), 201–232.
15. J. Denef and F. Loeser, *Motivic integration, quotient singularities and the McKay correspondence*, Compositio Math. **131** (2002), 267–290.
16. W. Fulton, *Introduction to toric varieties*, Annals of Math. Studies **131**, Princeton University Press, (1993).
17. H. Gillet and C. Soulé, *Descent, motives and K-theory*, J. Reine Angew. Math. **478** (1996), 127–176.
18. H. Hirzebruch and T. Höfer, *On the Euler number of an orbifold*, Math. Ann. **286** (1990), 255–260.
19. Y. Kawamata, K. Matsuda, and K. Matsuki, *Introduction to the Minimal Model Problem*, Adv. Studies in Pure Math. **10** (1987), 283–360.
20. M. Kontsevich, *Motivic integration*, Lecture at Orsay, (1995).
21. E. Looijenga, *Motivic measures*, Séminaire Bourbaki, Vol. 1999/2000. Astérisque **276** (2002), 267–297.
22. J. McKay, *Graphs, singularities and finite groups*, Proceedings of Symposia in Pure Mathematics **37**, 1980, pp. 183–186.
23. D. Morrison and G. Stevens, *Terminal quotient singularities in dimensions three and four*, Proceedings of the A.M.S. **90**, (1984), pp. 15–20.
24. D. Mumford, *The red book of varieties and schemes*, Lecture Notes in Mathematics **1358**, Springer-Verlag, (1988).
25. M. Reid, *Canonical 3-folds*, Journées de Géométrie Algébrique d'Angers, A. Beauville (editor), Sijthoff and Noordhoff, pages 273–310, (1980).
26. M. Reid, *Young person's guide to canonical singularities*, Proceedings of Symposia in Pure Mathematics **46**, (1987), pp. 345–414.
27. M. Reid, *The McKay correspondence and the physicists' Euler number conjecture*, Lecture notes given at the University of Utah (Sept.), and MSRI (Nov.), (1992).
28. W. Rudin, *Real and complex analysis*, McGraw–Hill Series in Higher Mathematics, (1987).
29. A. Scholl, *Classical motives*, Proceedings of Symposia in Pure Mathematics **55** Part 1, (1994), pp. 163–187.
30. C. Vafa, *String vacua and orbifoldized LG models*, Modern Phys. Lett. A **4** (1989), 1169–1185.
31. K. Watanabe, *Certain invariant subrings are Gorenstein I and II*, Osaka Journal **11** (1974), 1–8 and 379–388.
32. E. Zaslow, *Topological orbifold models and quantum cohomology rings*, Comm. Math. Phys. **156** (1993), 301–331.

Department of Mathematics, University of Utah, 155 South 1400 East, Salt Lake City, UT 84112

E-mail address: `craw@math.utah.edu`

Clay Mathematics Proceedings
Volume 3, 2004

Representation moduli of the McKay quiver for finite Abelian subgroups of $\mathrm{SL}(3,\mathbb{C})$

Akira Ishii

Abstract. The representation moduli of the McKay quiver, or the moduli of G-constellations, is known to be a crepant resolution of the quotient singularity $\mathbb{C}^3/G$ for a finite subgroup G of $\mathrm{SL}(3,\mathbb{C})$. The proof by Bridgeland, King and Reid [**1**] uses derived category technique. In this note, we give local (toric) coordinates of the moduli when G is Abelian, generalising Nakamura's coordinates on G-Hilb [**9**]. We also resume the chamber structure for stability from [**3**] briefly.

1. Introduction

Let G be a finite subgroup of $\mathrm{SL}(3,\mathbb{C})$. Nakamura [**9**] showed that the moduli space G-Hilb of G-clusters is a crepant resolution when G is Abelian. Bridgeland, King and Reid [**1**] proved that the same holds for any finite subgroup of $\mathrm{SL}(3,\mathbb{C})$ by using the Fourier-Mukai transform $\Phi\colon D(Y) \xrightarrow{\sim} D^G(\mathbb{C}^3)$, where $D(Y)$ and $D^G(\mathbb{C}^3)$ are the bounded derived categories of coherent sheaves on $Y = G$-Hilb and G-equivariant coherent sheaves on $\mathbb{C}^3$ respectively. Therefore there is a special choice G-Hilb among crepant resolutions of $\mathbb{C}^3/G$.

We can consider more general moduli spaces $\mathcal{M}_\theta$ introduced by Kronheimer [**7**] and further studied by Sardo Infirri [**11**]. These spaces are again crepant resolutions for generic parameters θ by the same arguments as in [**1**]. In [**3**], under the assumption that G is Abelian, we studied variation of the moduli spaces $\mathcal{M}_\theta$ (and the associated Fourier-Mukai transform Φ_θ) when the stability parameter θ moves. In particular, we saw that every projective crepant resolution is isomorphic to $\mathcal{M}_\theta$ for some θ.

In this note, we introduce local coordinates on the moduli space $\mathcal{M}_\theta$ when G is Abelian. This gives an elementary proof that $\mathcal{M}_\theta$ is a crepant resolution in the Abelian case. To describe the local coordinates, we use the *universal covering of the McKay quiver* and the *G-igsaw puzzle* as in [**9**] and [**10**]. Some of the ideas here were also used in [**3**] §10.

In the rest of this section, we give basic definitions. We introduce the notion of a G-constellation here which is regarded as a representation of the McKay quiver in

2000 *Mathematics Subject Classification.* 14E15, 14D20.

§3. For a finite subgroup G of $\mathrm{SL}(3,\mathbb{C})$, we denote by R its regular representation and by $R(G)$ the representation ring.

Definition 1.1. A *G-constellation* is a G-equivariant coherent sheaf F on $\mathbb{C}^n$ such that $H^0(F)$ is isomorphic as a $\mathbb{C}[G]$-module to R. Set

$$\Theta := \{\, \theta \in \mathrm{Hom}_{\mathbb{Z}}(R(G),\mathbb{Q}) \,\big|\, \theta(R) = 0 \,\}.$$

For $\theta \in \Theta$, a G-constellation is said to be *θ-stable* if every proper G-equivariant coherent subsheaf $0 \subset E \subset F$ satisfies $\theta(E) > 0$, i.e., $\theta(H^0(E)) > 0 = \theta(H^0(F))$. The notion of *$\theta$-semistable* is the same with $\geq$ replacing $>$.

We denote by $\mathcal{M}_\theta$ and $\overline{\mathcal{M}_\theta}$ the moduli spaces of θ-stable and θ-semistable G-constellations respectively. The existence of these moduli spaces are ensured by a work of King [**6**]. We recall the construction of these moduli spaces in §3. For a particular choice of parameters θ, $\mathcal{M}_\theta = \overline{\mathcal{M}_\theta}$ is G-Hilb (See [**5**] §3 and also [**3**] §9). $\mathcal{M}_\theta$ and the quotient singularity $X = \mathbb{C}^n/G$ are related by the following (set-theoretically) commutative diagram:

$$\begin{tikzcd}
\mathcal{M}_\theta \arrow[r, "\text{open immersion}"] \arrow[d] & \overline{\mathcal{M}_\theta} \arrow[d, "\text{projective}"] \\
X \arrow[r, "\text{injective}"'] & \overline{\mathcal{M}_0}
\end{tikzcd}$$

The first vertical arrow sends a G-constellation to its support and the second horizontal arrow sends a G-orbit to the class of G-constellations supported by it.

Definition 1.2. We say a parameter $\theta \in \Theta$ is *generic* if every θ-semistable G-constellation is θ-stable.

Arguments in [**1**] show that if $n = 3$ and if the parameter θ is generic, then $\mathcal{M}_\theta$ is a crepant resolution of $\mathbb{C}^3/G$ and we have a Fourier–Mukai transform $\Phi_\theta : D(\mathcal{M}_\theta) \xrightarrow{\sim} D^G(\mathbb{C}^3)$. In §3, we give an elementary proof that $\mathcal{M}_\theta$ is a crepant resolution if θ is generic.

Acknowledgements The author wishes to thank A. King, H. Nakajima, I. Nakamura and M. Reid for discussions and encouragements. He is also grateful to A. Craw for many discussions and useful suggestions. In particular, § 2 is a resumé of the author's joint work with Craw.

2. Chamber structure for stability

In this section, we briefly recall results from [**3**], a part of which the author talked about in the conference. These results are not needed in the following sections. In [**3**], we studied the variation of pairs $(\mathcal{M}_\theta, \Phi_\theta)$ as the parameters θ vary, assuming that G is Abelian. It is described by a chamber system in the parameter space Θ; for a generic θ, $(\mathcal{M}_\theta, \Phi_\theta)$ depends only on the chamber C which contains θ. Thus we can write $(\mathcal{M}_C, \Phi_C)$ instead of $(\mathcal{M}_\theta, \Phi_\theta)$. Our purpose in [**3**] was to describe this chamber system together with the variation of $(\mathcal{M}_C, \Phi_C)$. One natural question is whether every projective crepant resolution is isomorphic to $\mathcal{M}_C$ for some chamber C. We gave an affirmative answer to this question in [**3**]:

THEOREM 2.1. *For a finite Abelian subgroup $G \subset \mathrm{SL}(3,\mathbb{C})$, suppose $Y \to \mathbb{C}^3/G$ is a projective crepant resolution. Then there is a chamber $C \subset \Theta$ such that $Y \cong \mathcal{M}_C$.*

Another main result in [**3**] is the determination of the equivalences $\Phi_{C'}^{-1} \circ \Phi_C$ for adjacent chambers C, C' separated by a wall. Walls are classified into three types 0, I, III. (Types I and III correspond to the primitive birational contractions of types I and III , and type 0 is the case where no contraction occurs; see [**3**].)

THEOREM 2.2. *Let C and C' be adjacent chambers in Θ with $W = \overline{C} \cap \overline{C'}$. Then the relation between the moduli spaces $\mathcal{M}_C, \mathcal{M}_{C'}$ and the explicit form of the derived equivalence*

$$\Phi_{C'}^{-1} \circ \Phi_C \colon D(\mathcal{M}_C) \xrightarrow{\sim} D(\mathcal{M}_{C'})$$

is determined by the type of the wall W:

- **Type** 0**:** *$\mathcal{M}_C$ is isomorphic to $\mathcal{M}_{C'}$ and the derived equivalence $\Phi_{C'}^{-1} \circ \Phi_C$ is a Seidel–Thomas twist* [**12**] *(up to a tensor product with a line bundle).*
- **Type** I**:** *$\mathcal{M}_C$ is a flop of $\mathcal{M}_{C'}$ and $\Phi_{C'}^{-1} \circ \Phi_C$ is the induced derived equivalence described by Bondal–Orlov* [**2**, §3].
- **Type** III**:** *$\mathcal{M}_C$ is isomorphic to $\mathcal{M}_{C'}$ and $\Phi_{C'}^{-1} \circ \Phi_C$ is an EZ-transformation of Horja–Szendrői* [**4, 13**] *(up to a tensor product with a line bundle).*

The key step to these results is to understand a chamber C in terms of the K-theory of $\mathcal{M}_C$ via the Fourier-Mukai transform Φ_C. C is given by linear inequalities that correspond to exceptional curves or line bundles on divisors via Φ_C:

THEOREM 2.3. *Let $C \subset \Theta$ be a chamber. Then $\theta \in C$ if and only if*

(1) *for every exceptional curve ℓ we have $\theta([\Phi_C(\mathcal{O}_\ell)]) > 0$.*
(2) *for every compact reduced divisor D and irreducible representation ρ we have*

$$\theta([\Phi_C(\mathcal{R}_\rho^{-1} \otimes \omega_D)]) < 0 \quad \textit{and} \quad \theta([\Phi_C(\mathcal{R}_\rho^{-1}|_D)]) > 0.$$

Here $\mathcal{R}_\rho$ are the *tautological bundles* on $\mathcal{M}_\theta$ associated to irreducible representations $\rho \in \mathrm{Irr}(G)$. By identifying the Grothendieck group $K_0^G(\mathbb{C}^3)$ of coherent sheaves on $\mathbb{C}^3$ supported at the origin with the representation ring $R(G)$ of G, we can substitute the class $[\mathcal{F}]$ of $\mathcal{F} \in D_0^G(\mathbb{C}^3)$ into θ. For the details of these results, see [**3**].

3. Local coordinates on $\mathcal{M}_\theta$

Hereafter, we assume that θ is generic and we give local coordinates of the moduli space $\mathcal{M}_\theta$.

3.1. Representations of the McKay quiver. Let V be a three-dimensional vector space over $\mathbb{C}$ and $G \subset \mathrm{SL}(V)$ a finite subgroup. The moduli space $\mathcal{M}_\theta$ of G-constellations is constructed by regarding it as a moduli space of representations of the McKay quiver. In this section, we recall the construction.

Consider the affine scheme

$$\mathcal{N} = \{\, B \in \mathrm{Hom}_{\mathbb{C}[G]}(R, V \otimes R) \mid B \wedge B = 0 \,\},$$

where $B \wedge B$ lies in $\mathrm{Hom}_{\mathbb{C}[G]}(R, \wedge^2 V \otimes R)$. Each map $B \in \mathcal{N}$ determines an action of V^* on R and the condition $B \wedge B = 0$ ensures that this action is commutative. In this way, B endows R with a $\mathbb{C}[V]$-module structure, where $\mathbb{C}[V] = \oplus_{k=0}^{\infty} \mathrm{Sym}^k(V^*)$

is the polynomial ring of functions on V. Thus B determines a G-constellation on V.

Denote by $\mathrm{Aut}_{\mathbb{C}[G]}(R) \subset \mathrm{GL}(R)$ the group of G-equivariant automorphisms of R. If we decompose the regular representation $R = \oplus_{\rho \in \mathrm{Irr}(G)} R_\rho \otimes \rho$ as a G-module, we have

$$\mathrm{Aut}_{\mathbb{C}[G]}(R) = \prod_{\rho \in \mathrm{Irr}(G)} \mathrm{GL}(R_\rho)$$

This group acts on $\mathcal{N}$ by conjugation and the diagonal scalar subgroup $\mathbb{C}^*$ acts trivially, leaving a faithful action of

$$\mathrm{PAut}_{\mathbb{C}[G]}(R) := \mathrm{Aut}_{\mathbb{C}[G]}(R)/\mathbb{C}^*$$

on $\mathcal{N}$ by conjugation. Let $\theta \in \mathrm{Hom}_{\mathbb{Z}}(R(G), \mathbb{Z})$ satisfy $\theta(R) = 0$. Then θ determines a character χ_θ of $\mathrm{PAut}_{\mathbb{C}[G]}(R)$ mapping $[(g_\rho)_\rho]$ to $\prod_\rho \det(g_\rho)^{\theta(\rho)} \in \mathbb{C}^*$. This character determines a linearisation of the trivial line bundle on $\mathcal{N}$. We denote by

$$\mathcal{M}_\theta := \mathcal{N}_\theta / \mathrm{PAut}_{\mathbb{C}[G]}(R) \quad \text{and} \quad \overline{\mathcal{M}_\theta} := \overline{\mathcal{N}_\theta} /\!\!/ \mathrm{PAut}_{\mathbb{C}[G]}(R)$$

the geometric and the categorical quotients of the open subsets $\mathcal{N}_\theta$ and $\overline{\mathcal{N}_\theta}$ of χ_θ-stable and χ_θ-semistable points of $\mathcal{N}$. King's result (which holds in a more general context) asserts that $\mathcal{M}_\theta$ and $\overline{\mathcal{M}_\theta}$ are equivalently the moduli spaces of θ-stable and θ-semistable G-constellations respectively. There is an isomorphism $\mathcal{M}_{k\theta} \cong \mathcal{M}_\theta$ for $k \in \mathbb{Z}_{>0}$, so $\mathcal{M}_\theta$ is well defined even for parameters $\theta \in \mathrm{Hom}_{\mathbb{Z}}(R(G), \mathbb{Q})$, i.e., for $\theta \in \Theta$.

To let $R \otimes \mathcal{O}_{\mathcal{N}}$ and the universal homomorphism $R \otimes \mathcal{O}_{\mathcal{N}} \to V \otimes R \otimes \mathcal{O}_{\mathcal{N}}$ on $\mathcal{N}$ descend to $\mathcal{M}_\theta$, we have to determine an equivariant $\mathrm{PAut}_{\mathbb{C}[G]}(R)$-action on $R \otimes \mathcal{O}_{\mathcal{N}}$. Define a subgroup

$$\mathrm{Aut}'_{\mathbb{C}[G]}(R) = \{ \textstyle\prod_{\rho \in \mathrm{Irr}(G)} (g_\rho) \in \mathrm{Aut}_{\mathbb{C}[G]}(R) \,\big|\, g_{\rho_0} = 1 \}$$

of $\mathrm{Aut}_{\mathbb{C}[G]}(R)$, where ρ_0 is the trivial representation and hence $\dim R_{\rho_0} = 1$. Then the projection gives an isomorphism $\mathrm{Aut}'_{\mathbb{C}[G]}(R) \to \mathrm{PAut}_{\mathbb{C}[G]}(R)$. On the other hand, $\mathrm{Aut}'_{\mathbb{C}[G]}(R)$ has a natural equivariant action on $R \otimes \mathcal{O}_{\mathcal{N}}$. Descent theory of coherent sheaves for a faithfully flat morphism implies that $R \otimes \mathcal{O}_{\mathcal{N}}$ and $R_\rho \otimes \mathcal{O}_{\mathcal{N}}$ descend to locally free sheaves $\mathcal{R} := \mathcal{R}_\theta$ and $\mathcal{R}_\rho := (\mathcal{R}_\theta)_\rho$ respectively on $\mathcal{M}_\theta$ such that $\mathcal{R} = \oplus_{\rho \in \mathrm{Irr}(G)} \mathcal{R}_\rho \otimes \rho$. Moreover, we have a homomorphism $\mathcal{R} \to V \otimes \mathcal{R}$ determining the universal G-constellation $\mathcal{U}_\theta$ on $\mathcal{M}_\theta \times \mathbb{C}^3$. Note by the definition of $\mathrm{Aut}'_{\mathbb{C}[G]}(R)$ that the line bundle $\mathcal{R}_{\rho_0}$ is trivial.

From now on, we consider the case where G is Abelian. We denote by $G^* = \mathrm{Irr}(G)$ the character group of G. Decompose $V^* = \rho_1 \oplus \rho_2 \oplus \rho_3$ as a representation of G. An element $B \in \mathrm{Hom}_{\mathbb{C}[G]}(R, V \otimes R)$ is decomposed as

$$B = (b_1^\rho, b_2^\rho, b_3^\rho)_{\rho \in G^*} = (b_i^\rho)_{i,\rho}$$

where b_i^ρ is a linear map $R_\rho \to R_{\rho\rho_i}$. Then the commutativity condition $B \wedge B = 0$ is expressed as

$$b_j^{\rho\rho_i} b_i^\rho = b_i^{\rho\rho_j} b_j^\rho.$$

The action of $[(\alpha_\rho)_{\rho \in G^*}] \in \mathrm{PAut}_{\mathbb{C}[G]}(R)$ is written as

$$[(\alpha_\rho)_{\rho \in G^*}] \cdot (b_i^\rho)_{i,\rho} = (\alpha_{\rho\rho_i} \alpha_\rho^{-1} b_i^\rho)_{i,\rho}.$$

The three-dimensional torus $\widetilde{T} = \mathrm{Aut}_{\mathbb{C}[G]}(V)$ acts naturally on $\mathcal{N}$ by

$$(t_1, t_2, t_3) \cdot (b_1^\rho, b_2^\rho, b_3^\rho)_\rho = (t_1 b_1^\rho, t_2 b_2^\rho, t_3 b_3^\rho)_\rho.$$

We regard (b_i^ρ) as a representation of the McKay quiver with dimension vector $(1, 1, \ldots, 1)$ (see [**11**]). By definition, the McKay quiver is the quiver with vertex set G^* and arrows a_i^ρ from ρ to $\rho\rho_i$ for $i = 1, 2, 3$ and $\rho \in G^*$.

3.2. Universal covering of the McKay quiver. In the study of representations of the McKay quiver in our case, it is convenient to consider the universal covering of it ([**9**], [**10**]). By taking ρ_1, ρ_2, ρ_3 as generators of G^*, we have an exact sequence

$$0 \to M \to \mathbb{Z}^3 \to G^* \to 0. \tag{3.1}$$

Since G is in $\mathrm{SL}(V)$, M contains $(1, 1, 1)$. If we put $M' = M/\mathbb{Z}(1, 1, 1)$ and $H = \mathbb{Z}^3/\mathbb{Z}(1, 1, 1)$ then we have an exact sequence

$$0 \to M' \to H \to G^* \to 0.$$

We denote by f_1, f_2, f_3 the images of $(1, 0, 0)$, $(0, 1, 0)$, $(0, 0, 1)$ respectively.

Definition 3.1. Consider the quiver with vertex set H and arrows from $h \in H$ to $h + f_i$ for $i = 1, 2, 3$. We call it the universal covering of the McKay quiver.

The McKay quiver is the quotient of the universal covering by the action of M'. For later purposes, we identify $H_\mathbb{R} = H \otimes \mathbb{R}$ with the orthogonal complement of $(1, 1, 1)$ in $\mathbb{R}^3$ with respect to the Euclidean metric.

3.3. Torus fixed points and the G-igsaw puzzle. We look at torus fixed points of $\mathcal{M}_\theta$. Let $B \in \mathcal{N}_\theta$ give a T-fixed point of $\mathcal{M}_\theta$, where $T = \mathrm{Spec}(\mathbb{C}[M]) = \widetilde{T}/G$. Define a graph $\Gamma(B)$ whose vertex set is G^* and whose edges are a_i^ρ with $b_i^\rho \neq 0$, forgetting the orientation. We have a natural lift of $\Gamma(B)$ in $H_\mathbb{R}$ as follows. $[B] \in \mathcal{N}_\theta / \mathrm{Aut}'_{\mathbb{C}[G]}(R)$ is a $\widetilde{T}$-fixed point and $\mathrm{Aut}'_{\mathbb{C}[G]}(R)$ acts freely on $\mathcal{N}$ at B. Therefore, we have a homomorphism $\alpha : \widetilde{T} \to \mathrm{Aut}'_{\mathbb{C}[G]}(R)$ such that $t \cdot B = \alpha(t) \cdot B$. Write α as

$$(t_1, t_2, t_3) \mapsto (t_1^{-k_\rho} t_2^{-l_\rho} t_3^{-m_\rho})_\rho \in \prod_{\rho \in G^*} \mathbb{C}^* \tag{3.2}$$

where (k_ρ, l_ρ, m_ρ) are integers attached to $\rho \in G^*$ with $k_{\rho_0} = l_{\rho_0} = m_{\rho_0} = 0$. For $\rho \in G^*$, put

$$h(\rho) = k_\rho f_1 + l_\rho f_2 + m_\rho f_3 \in H. \tag{3.3}$$

Draw a line segment from $h(\rho)$ to $h(\rho) + f_i$ if $b_i^\rho \neq 0$. In this case, we can see that $h(\rho) + f_i = h(\rho\rho_i)$ and hence we can embed $\Gamma(B)$ in the plane $H_\mathbb{R}$ so that the sets of edges and vertices of $\Gamma(B)$ are subsets of those of the universal covering of the McKay quiver respectively. For $h \in H$ we put

$$\mathrm{Hex}(h) = \{\, P \in H_\mathbb{R} \mid d(P, h) \leq d(P, h') \text{ for } h' \in H \,\}$$

where $d(*, *)$ denotes the metric of $H_\mathbb{R}$. Then $\mathrm{Hex}(h)$ is a regular hexagon with center h and these hexagons form a honeycombed tiling of $H_\mathbb{R}$. Define $D(B)$ as

$$D(B) = \bigcup_{\rho \in G^*} \mathrm{Hex}(h(\rho)).$$

Then $D(B)$ is a fundamental domain for the action of M' on $H_\mathbb{R}$ and we have a tiling of $H_\mathbb{R}$ by the fundamental domains parallel to $D(B)$ as

$$H_\mathbb{R} = \bigcup_{m \in M'} (D(B) + m)\,. \tag{3.4}$$

By the commutativity condition, $\Gamma(B)$ has the following property: if a_i^ρ and $a_j^{\rho\rho_i}$ are edges of $\Gamma(B)$, then so are a_j^ρ and $a_i^{\rho\rho_j}$ for $\rho \in G^*$ and $i \neq j$. Let $C(\rho, i, j)$ denote the cycle formed by arrows a_i^ρ, $a_j^{\rho\rho_i}$, $(a_i^{\rho\rho_j})^{-1}$ and $(a_j^\rho)^{-1}$ which appear in the commutativity condition. We denote by $\mathrm{Cycle}(B)$ the set of such cycles in $\Gamma(B)$ with $i < j$. Note that if $C(\rho, i, j) \in \mathrm{Cycle}(B)$, then the diagonal $a_k^{\rho\rho_i\rho_j}$ is not contained in $\Gamma(B)$, where $\{i, j, k\} = \{1, 2, 3\}$.

LEMMA 3.2. *There exists a unique representation $\sigma \in G^*$ such that $b_1^\sigma = b_2^{\sigma\rho_1} = b_3^{\sigma\rho_1\rho_2} = 0$. Symmetrically, there also exists a unique representation $\tau \in G^*$ such that $b_1^\tau = b_3^{\tau\rho_1} = b_2^{\tau\rho_1\rho_3} = 0$. Moreover, the cycles $C(\rho, i, j) \in \mathrm{Cycle}(B)$ form a basis of $H_1(\Gamma(B), \mathbb{Z})$.*

PROOF. Let p(resp. q) denote the number of vertices $\rho \in G^*$ such that exactly two (resp. one) of b_1^σ, $b_2^{\sigma\rho_1}$, $b_3^{\sigma\rho_1\rho_2}$ are non-zero. Note that at least one of these three values is 0, since B gives a fixed point. Hence the assertion that σ exists uniquely is equivalent to the equality $n - (p + q) = 1$.

Since $\Gamma(B)$ has $2p + q$ edges and n vertices, the Euler characteristic of $\Gamma(B)$ is given by

$$\chi(\Gamma(B)) = n - (2p + q).$$

On the other hand, we have exactly p cycles $C(\rho, i, j) \in \mathrm{Cycle}(B)$ which are minimal loops of $\Gamma(B)$ in the plane $H_{\mathbb{R}}$. It follows that the first Betti number satisfies $b_1(\Gamma(B)) \geq p$. The stability implies that $\Gamma(B)$ is connected. Thus, $\chi(\Gamma(B)) \leq 1 - p$ and therefore $n - (p + q) \leq 1$. So if we show there exists at least one σ or τ, then every equality holds and all the assertions follow.

We embedded $\Gamma(B)$ in $H_{\mathbb{R}}$ and obtained a fundamental domain $D(B)$ determined by it. There is a point P on the boundary of $D(B)$ where $D(B)$ and two other fundamental domains of the tiling of $H_{\mathbb{R}}$ meet. The three hexagons containing P belong to different domains and one of the central points of these hexagons is the desired one. □

COROLLARY 3.3. *$D(B)$ contacts 6 fundamental domains of the tiling (3.4). The points where $D(B)$ and two of these domains meet are the following:*

$$\begin{array}{lll} h(\sigma) + \frac{2}{3}f_1 + \frac{1}{3}f_2, & h(\sigma\rho_1) + \frac{2}{3}f_2 + \frac{1}{3}f_3, & h(\sigma\rho_1\rho_2) + \frac{2}{3}f_3 + \frac{1}{3}f_1, \\ h(\tau) + \frac{2}{3}f_1 + \frac{1}{3}f_3, & h(\tau\rho_1) + \frac{2}{3}f_3 + \frac{1}{3}f_2, & h(\tau\rho_1\rho_3) + \frac{2}{3}f_2 + \frac{1}{3}f_1 \end{array}$$

Points in the first line and the second line above appear alternately on the boundary $\partial D(B)$.

PROOF. Only the last statement needs a proof. Assume Y and Y' are points on the first line adjacent in $\partial D(B)$. We denote by $\widehat{YY'}$ the connected component of $\partial D(B) \setminus \{Y, Y'\}$ on which there are no other points above. Since Y and Y' are congruent modulo M', there is $m \in M'$ such that $Y' = Y + m$. Put $L = \bigcup_{j \in \mathbb{Z}} (\widehat{YY'} + jm)$. It is a connected infinite graph without trivalent points, dividing $H_{\mathbb{R}}$ into two parts. The points where three fundamental domains meet on L are $Y + jm$ $(j \in \mathbb{Z})$. The opposite side of $D(B)$ with respect to L must be contained in a single domain parallel to $D(B)$, a contradiction. □

Let $\delta(B)$ denote the set of edges e of $\mathrm{Hex}(h(\rho))$ for some $\rho \in G^*$ such that e is not in the interior of any $C(\rho, i, j) \in \mathrm{Cycle}(B)$ and such that $e \cap \Gamma(B) = \emptyset$.

LEMMA 3.4. *Any $e \in \delta$ is contained in the boundary $\partial D(B)$ of $D(B)$ and we have*

$$\partial D(B) = \bigcup_{e \in \delta(B)} e.$$

PROOF. The inclusion $\partial D(B) \subseteq \cup_{e \in \delta(B)} e$ is obvious. If $\cup_{e \in \delta(B)} e$ has a trivalent point, it must be one of the 6 points in the above corollary. But at any of these points three fundamental domains contact and it is impossible that the three edges having this point as a vertex are contained in $D(B)$. The commutativity condition $B \wedge B = 0$ shows that $\cup_{e \in \delta(B)} e$ does not have an end point. Therefore, $\cup_{e \in \delta(B)} e$ is homeomorphic to the disjoint union of finitely many copies of S^1. Each S^1 must surround an integral point and hence it follows from the simply-connectedness of $D(B)$ that $\cup_{e \in \delta(B)} e$ is connected. □

We fix the orientation of $H_{\mathbb{R}}$ by taking f_1, f_2 as a basis and say they are in the anticlockwise order. Then we can determine the clockwise and the anticlockwise orientations of $\partial D(B)$. Denote the six points in Corollary 3.3 by

$$Y_1, \quad A_1, \quad Y_2, \quad A_2, \quad Y_3, \quad A_3$$

in the clockwise order. We assume the points in the first row in Corollary 3.3 are Y_i and the others are A_i (in appropriate orders).

REMARK 3.5. The letter Y stands for the way in which the three domains contact at a point. (To see the correct picture, we should assume that the vector f_1 faces to the right.) A denotes the way upside down, though this letter does not show a correct picture.

LEMMA 3.6. *Consider the arrows of the universal covering of the McKay quiver such that they intersect $\partial D(B)$. Among them, assume $\tilde{a}$ is the nearest to Y_i in the clockwise direction of $\partial D(B)$ and $\tilde{a}'$ in the anti-clockwise direction. Then $\tilde{a}$ is going out of $D(B)$ and $\tilde{a}'$ is going into $D(B)$. For the points nearest to A_i, the same assertion holds with directions of the arrows reversed.*

PROOF. The three arrows of the universal covering of the McKay quiver nearest to Y_i go around Y_i in the anti-clockwise direction. $\tilde{a}$ and $\tilde{a}'$ are two of these three arrows and they have one point P in common, which must be contained in $D(B)$. Since we determined the direction of $\partial D(B)$ to be clockwise, $\tilde{a}'$ goes into P and $\tilde{a}$ goes out of P. □

3.4. Local coordinates. In this section, we introduce a local coordinate of $\mathcal{M}_\theta$. Throughout this section, $B \in \mathcal{N}_\theta$ denotes a representation of the McKay quiver such that $[B] \in \mathcal{M}_\theta$ is a T-fixed point.

Fix isomorphisms $R_\rho \cong \mathbb{C}$. Then we can regard the linear map $b_i^\rho : R_\rho \to R_{\rho\rho_i}$ as a complex number. We call $b_i^\rho \in \mathbb{C}$ the value of B at the arrow a_i^ρ.

LEMMA 3.7. *There exists a unique element $\lambda \in \mathrm{PAut}'_{\mathbb{C}[G]}(R)$ such that if we put $\lambda \cdot B = ({b'}_i^\rho)_{i,\rho}$ then ${b'}_i^\rho$ is either 0 or 1.*

PROOF. Fix $\rho \in G^*$. Since $\Gamma(B)$ is connected, we can take a sequence of arrows

$$\gamma = \{ \alpha_1^{\varepsilon_1}, \alpha_2^{\varepsilon_2}, \dots, \alpha_p^{\varepsilon_p} \}$$

such that α_k is an arrow of the McKay quiver contained in $\Gamma(B)$, $\varepsilon_i = \pm 1$ determines the direction of arrows and such that γ gives a connected path from $h(\rho_0)$ to $h(\rho)$. Put

$$\lambda_\rho = \prod_{k=1}^{p} \beta_k^{-\varepsilon_k}$$

where if $\alpha_k = a_i^\rho$ then $\beta_k = b_i^\rho$ is the value of B at this arrow. The last statement of Lemma 3.2 and the commutativity condition imply that λ_ρ is independent of the choice of the path γ. Put

$$\lambda = (\lambda_\rho)_\rho \in \operatorname{Aut}'_{\mathbb{C}[G]}(R).$$

Then $\lambda \cdot B$ has the desired property. The uniqueness follows from the connectedness of $\Gamma(B)$. □

So we can assume b_i^ρ is 0 or 1 in the sequel. Put

$$\widetilde{U}(B) = \{\, C = (c_i^\rho)_{i,\rho} \in \mathcal{N} \mid c_i^\rho \neq 0 \text{ if } b_i^\rho \neq 0 \,\}$$

and

$$U(B) = \{\, C = (c_i^\rho)_{i,\rho} \in \mathcal{N} \mid c_i^\rho = 1 \text{ if } b_i^\rho \neq 0 \,\}.$$

COROLLARY 3.8. *The action of* $\operatorname{Aut}'_{\mathbb{C}[G]}(R)$ *induces an isomorphism*

$$\operatorname{Aut}'_{\mathbb{C}[G]}(R) \times U(B) \cong \widetilde{U}(B)$$

and $U(B) \cong \widetilde{U}(B)/\operatorname{Aut}'_{\mathbb{C}[G]}(R)$ *becomes an open neighborhood of* $[B]$ *in* $\mathcal{M}_\theta$.

PROOF. The first part is proved in the same way as in the above lemma. If $C \in \widetilde{U}(B)$, then every C-invariant subrepresentation of R is also B-invariant. Therefore, C is also θ-stable and we obtain the second assertion. □

Thus we can restrict the action of the torus $\widetilde{T}$ on $\mathcal{M}_\theta$ to the open set $U(B)$.

COROLLARY 3.9. *For* $t = (t_1, t_2, t_3) \in \widetilde{T}$ *and* $(c_i^\rho) \in U(B)$, *put* $t \cdot (c_i^\rho) = (d_i^\rho)$ *with respect to the above action. We put*

$$(\alpha_1, \alpha_2, \alpha_3) = (k_\rho + \delta_{i1}, l_\rho + \delta_{i2}, m_\rho + \delta_{i3}) - (k_{\rho\rho_i}, l_{\rho\rho_i}, m_{\rho\rho_i}),$$

where k_ρ, l_ρ, m_ρ *are the weights for* $\rho \in G^*$ *given in* (3.2) *and where* δ_{ij} *denotes Kronecker's delta. Then* $d_i^\rho = t_1^{-\alpha_1} t_2^{-\alpha_2} t_3^{-\alpha_3} c_i^\rho$

PROOF. Do the same argument for $B' = (t_i b_i^\rho)_i^\rho$ as in the lemma, assuming b_i^ρ are either 0 or 1. By the definition of λ_ρ in the proof of the lemma, we have $t_i = \lambda_{\rho\rho_i}^{-1} \lambda_\rho$ if $b_i^\rho \neq 0$. Then the definition of the weights k_ρ, l_ρ, m_ρ implies that $\lambda_\rho = t_1^{-k_\rho} t_2^{-l_\rho} t_3^{-m_\rho}$. Since d_i^ρ is given by $d_i^\rho = \lambda_{\rho\rho_i} \lambda_\rho^{-1} t_i c_i^\rho$, the assertion follows. □

Now determine a coordinate on $U(B)$. Divide $\partial D(B)$ into 6 parts by the points $\{Y_i\}$ and $\{A_i\}$. We denote them by $\widehat{Y_1A_1}, \widehat{A_1Y_2}, \dots, \widehat{A_3Y_1}$. For an arrow $a = a_i^\rho$ of the McKay quiver, we denote by $\tilde{a} = \tilde{a}_i^\rho$ the arrow of the universal covering which goes from $h(\rho)$ to $h(\rho) + f_i$. Note that if $\tilde{a}_i^\rho$ meets $\partial D(B)$, then it is always going out of $D(B)$. Let $\mathbb{C}[\xi_1, \xi_2, \xi_3]$ be the polynomial ring in three variables ξ_1, ξ_2, ξ_3. We define a representation (u_i^ρ) of the McKay quiver, with values $u_i^\rho \in \mathbb{C}[\xi_1, \xi_2, \xi_3]$ as follows.

(1) If $\tilde{a}_i^\rho$ is an edge of $\Gamma(B)$, then $u_i^\rho = 1$.
(2) If $\tilde{a}_i^\rho$ meets $\widehat{Y_pA_p}$, then $u_i^\rho = \xi_p$.

(3) If $\tilde{a}_i^\rho$ meets $\widehat{A_pY_{p+1}}$, then $u_i^\rho = \xi_p\xi_{p+1}$ (Set $Y_4 = Y_1$ etc.)
(4) If $C(\rho\rho_i, j, k) \in \mathrm{Cycle}(B)$ with $\{i, j, k\} = \{1, 2, 3\}$ (i.e., $\tilde{a}_i^\rho$ is a diagonal of the diamond), then $u_i^\rho = \xi_1\xi_2\xi_3$.

By virtue of Lemma 3.4, there are no other cases (note that edges of hexagons are in one-to-one correspondence with arrows of the universal covering of the McKay quiver).

THEOREM 3.10. *The above representation* (u_i^ρ) *of the McKay quiver satisfies the commutativity condition and gives rise to an isomorphism*

$$\operatorname{Spec}\mathbb{C}[\xi_1, \xi_2, \xi_3] \cong U(B).$$

PROOF. We first prove that (u_i^ρ) satisfies the commutativity condition. It follows from the equality $u_1^\rho u_2^{\rho\rho_1} u_3^{\rho\rho_1\rho_2} = u_1^\rho u_3^{\rho\rho_1} u_1^{\rho\rho_1\rho_3} = \xi_1\xi_2\xi_3$, because $\mathbb{C}[\xi_1, \xi_2, \xi_3]$ is an integral domain and each arrow gives a non-zero value in this ring. We show $u_1^\rho u_2^{\rho\rho_1} u_3^{\rho\rho_1\rho_2} = \xi_1\xi_2\xi_3$, the other one is similar. First, if two of the arrows $a_1^\rho, a_2^{\rho\rho_1}, a_3^{\rho\rho_1\rho_2}$ are edges of $\Gamma(B)$, then they form half of a cycle in $\mathrm{Cycle}(B)$ by virtue of the commutativity condition for B and the third arrow satisfies the assumption of (4) above. Therefore, $u_1^\rho u_2^{\rho\rho_1} u_3^{\rho\rho_1\rho_2} = 1 \cdot 1 \cdot \xi_1\xi_2\xi_3$. Next assume exactly one of the three arrows is in $\Gamma(B)$. If we denote by a, a' the other two arrows, then Lemma 3.4 implies that the both of $\tilde{a}$ and $\tilde{a}'$ meet $\partial D(B)$. Moreover, we can see that if one meets $\widehat{Y_pA_p}$, then the other meets $\widehat{A_{p+1}Y_{p+2}}$. Therefore, we obtain $u_1^\rho u_2^{\rho\rho_1} u_3^{\rho\rho_1\rho_2} = 1 \cdot \xi_p \cdot \xi_{p+1}\xi_{p+2} = \xi_1\xi_2\xi_3$. Finally, assume none of the three arrows are in $\Gamma(B)$. Then by Lemma 3.2, the arrows are a_1^σ, $a_2^{\sigma\rho_1}$ and $a_3^{\sigma\rho_1\rho_2}$. Assume that

$$Y_p = h(\sigma) + (2/3)f_1 + (1/3)f_2.$$

Since $\tilde{a}_1^\sigma$ is going out of $D(B)$, Lemma 3.6 implies that $\tilde{a}_i^\sigma$ meets $\widehat{Y_pA_p}$ and hence that $u_1^\sigma = \xi_p$. Similar arguments show that $u_2^{\sigma\rho_1}$, $u_3^{\sigma\rho_1\rho_2}$ are the other variables. Thus we have proved that the commutativity relation holds for $(u_i^\rho)_{i,\rho}$ and we obtain an injection $\operatorname{Spec}\mathbb{C}[\xi_1, \xi_2, \xi_3] \to U$.

Conversely, let $\mathfrak{R}$ be a $\mathbb{C}$-algebra and consider any $\mathfrak{R}$-valued point $C = (c_i^\rho)_{i,\rho}$ of U. Let a and a' be two arrows such that $\tilde{a}$ and $\tilde{a}'$ meet $\widehat{Y_pA_p}$. We show that the values of C for a and a' are the same. $\widehat{Y_pA_p}$ is a connected path formed by edges of hexagons and the arrows of the universal covering meeting $\widehat{Y_pA_p}$ are in one-to-one correspondence with these edges. Two adjacent edges belong to the same hexagon and therefore one of two adjacent arrows(of the universal covering) meeting $\widehat{Y_pA_p}$ is going out of $D(B)$ and the other is going into $D(B)$. Let a be an arrow of the McKay quiver such that the lift $\tilde{a}$ meets Y_pA_p, which is going out of $D(B)$ by our choice of the lift. Assume there is $\tilde{a}'$ next to $\tilde{a}$ with the same properties. Then, there exists an arrow β of the universal covering going into $D(B)$ which meets both $\tilde{a}$ and $\tilde{a}'$. Take arrows γ, γ' of the universal covering such that $\tilde{a}, \beta, \gamma$ and $\tilde{a}', \beta, \gamma'$ form regular triangles respectively. Since both $\tilde{a}$ and $\tilde{a}'$ meet $\widehat{Y_pA_p}$, these triangles surround neither Y_p nor A_p. Therefore γ and γ' are mapped to arrows in the McKay quiver where B and hence C have value 1. Thus the commutativity relation implies that the values of C for a and a' are the same. Let $\zeta_p \in \mathfrak{R}$ be the common value. We show that c_i^ρ are determined by ζ_1, ζ_2 and ζ_3 as in (1), (2), (3), (4) above. (1) and (2) are obvious by the definitions of U and ζ_p.

Consider (3). The same argument as above shows that all the arrows a such that $\tilde{a}$ meet $\widehat{A_pY_{P+1}}$ give the same value for C. Take such $\tilde{a}$ the nearest to A_p (if it exists). Then do the similar arguments by using the triangle surrounding A_p. The commutativity relation implies that these values are $\zeta_p\zeta_{p+1}$.

Finally we show (4). The diagonals of adjacent diamonds give the same value by the commutativity relation. Assume $\tilde{a}$ is the diagonal of a diamond, which has an edge $\tilde{b}$ such that $\tilde{b}$ is not on any other diamond. Then the commutativity relation for B implies that $\tilde{b}$ is the unique arrow in the triangle containing $\tilde{b}$ but not containing $\tilde{a}$ such that it gives a non-zero value for B. Then (2) and (3) determine the values of C at the other edges of the triangle. Again the commutativity relation implies that C takes the value $\zeta_1\zeta_2\zeta_3$ at a. □

3.5. Nowhere vanishing 3-form on $\mathcal{M}_\theta$. We now show that $\mathcal{M}_\theta$ is a crepant resolution by constructing a nowhere vanishing 3-form on $\mathcal{M}_\theta$. The following construction is essentially due to Mukai [**8**]. Let E be a G-constellation which gives rise to a point $[E] \in \mathcal{M}_\theta$. Then the Zariski tangent space of $\mathcal{M}_\theta$ at $[E]$ is $G\text{-}\mathrm{Ext}^1_{\mathcal{O}_{\mathbb{C}^3}}(E,E)$. We denote by ω the product (or composite) map

$$G\text{-}\mathrm{Ext}^1_{\mathcal{O}_{\mathbb{C}^3}}(E,E) \times G\text{-}\mathrm{Ext}^1_{\mathcal{O}_{\mathbb{C}^3}}(E,E) \times G\text{-}\mathrm{Ext}^1_{\mathcal{O}_{\mathbb{C}^3}}(E,E) \longrightarrow G\text{-}\mathrm{Ext}^3_{\mathcal{O}_{\mathbb{C}^3}}(E,E).$$

Serre duality and the stability of E imply

$$G\text{-}\mathrm{Ext}^3_{\mathcal{O}_{\mathbb{C}^3}}(E,E) \cong G\text{-}\mathrm{Ext}^0_{\mathcal{O}_{\mathbb{C}^3}}(E, E \otimes K_{\mathbb{C}^3})^\vee \cong \mathbb{C}.$$

We show that ω determines a nowhere vanishing 3-form on $\mathcal{M}_\theta$. Let $B \in \mathcal{N}$ correspond to E. Then the Zariski tangent space is

$$G\text{-}\mathrm{Ext}^1_{\mathcal{O}_{\mathbb{C}^3}}(E,E) \cong \frac{\{\, C \in \mathrm{Hom}_{\mathbb{C}[G]}(R, V \otimes R) \mid C \wedge B + B \wedge C = 0 \,\}}{\{\, BA - (1_V \otimes A)B \mid A \in \mathrm{Hom}_{\mathbb{C}[G]}(R,R) \,\}}$$

and the above product is expressed as

$$\omega(C_1, C_2, C_3) = \mathrm{trace}(C_1 \wedge C_1 \wedge C_3).$$

Therefore, we have

$$\omega(C_1, C_2, C_3) = \omega(C_2, C_3, C_1).$$

On the other hand, the smoothness of $\mathcal{M}_\theta$ implies that the obstruction class in $G\text{-}\mathrm{Ext}^2_{\mathcal{O}_{\mathbb{C}^3}}(E,E)$ vanishes. That is, we have $\omega(C, C, C_3) = 0$. Thus ω is skew-symmetric.

We use the local coordinate ξ_1, ξ_2, ξ_3 and the universal representation on $U(B)$ in Theorem 3.10. Then we obtain

$$\omega\left(\frac{\partial}{\partial \xi_1}, \frac{\partial}{\partial \xi_2}, \frac{\partial}{\partial \xi_3}\right) = \pm 1$$

where the signature depends on the choice of a basis of $\wedge^3 V$. Thus ω does not vanish on any affine open $U(B)$ and is therefore nowhere vanishing on $\mathcal{M}_\theta$.

Remark 3.11. In contrast to the two-dimensional case [**8**], Serre duality doesn't imply that this is nowhere vanishing. But once we see that $\mathcal{M}_\theta$ is smooth of dimension three and that the (Fourier-Mukai) functor Φ_θ is fully faithful, then it automatically follows that the above 3-form is nowhere vanishing.

References

1. T. Bridgeland, A. King, and M. Reid, *The McKay correspondence as an equivalence of derived categories*, J. Amer. Math. Soc. **14** (2001), 535–554.
2. A. Bondal and D. Orlov, *Semiorthogonal decomposition for algebraic varieties*, Preprint math.AG/9506012, (1995).
3. A. Craw and A. Ishii, *Flops of G*-Hilb *and equvalences of derived categories by variaiton of GIT quotient*, To appear in Duke Math J., 43 pp.
4. R. P. Horja, *Derived category automorphisms from mirror symmetry*, Preprint math.AG/0103231, (2001).
5. Y. Ito and H. Nakajima, *The McKay correspondence and Hilbert schemes in dimension three*, Topology **39** (2000), 1155–1191.
6. A. King, *Moduli of representations of finite dimensional algebras*, Quart. J. Math. Oxford **45** (1994), 515–530.
7. P. Kronheimer, *The construction of ALE spaces as hyper-Kähler quotients*, J. Diff. Geom **29** (1989), 665–683.
8. S. Mukai, *Symplectic structure of the moduli space of sheaves on an Abelian or $k3$ surface*, Invent. Math., **77** (1984), 101–116.
9. I. Nakamura, *Hilbert schemes of Abelian group orbits*, J. Alg. Geom. **10** (2000), 757–779.
10. M. Reid, *McKay correspondence*, Proc. of algebraic geometry symposium (Kinosaki, Nov 1996), T. Katsura (ed.), (1997), pp. 14–41.
11. A. Sardo-Infirri, *Resolutions of orbifold singularities and the transportation problem on the McKay quiver*, Preprint math.AG/-9610005, (1996).
12. P. Seidel and R. Thomas, *Braid group actions on derived categories of coherent sheaves*, Duke Math. J. **108** (2001), 37–108.
13. B. Szendrői, *Diffeomorphisms and families of fourier–mukai transforms in mirror symmetry*, Applications of algebraic geometry to coding theory, physics and computation (Eilat, 2001), NATO Sci. Ser. II Math. Phys. Chem., 36, Kluwer Acad. Publ., Dordrecht., (2001), pp. 317–337.

Department of Urban and Environmental Engineering, Kyoto University, Kyoto 606-8501, Japan

E-mail address: `akira@kusm.kyoto-u.ac.jp`

Clay Mathematics Proceedings
Volume 3, 2004

Moduli Spaces of Bundles over Riemann Surfaces and the Yang–Mills Stratification Revisited

Frances Kirwan

Abstract. Refinements of the Yang–Mills stratifications of spaces of connections over a compact Riemann surface Σ are investigated. The motivation for this study is the search for a complete set of relations between the standard generators for the cohomology of the moduli spaces $\mathcal{M}(n,d)$ of stable holomorphic bundles of rank n and degree d when n and d are coprime and $n > 2$.

The moduli space $\mathcal{M}(n,d)$ of semistable holomorphic bundles of rank n and degree d over a Riemann surface Σ of genus $g \geq 2$ can be constructed as a quotient of an infinite dimensional affine space of connections $\mathcal{C}(n,d)$ by a complexified gauge group $\mathcal{G}_c(n,d)$, in an infinite-dimensional version of the construction of quotients in Mumford's geometric invariant theory [**30**]. When n and d are coprime, $\mathcal{M}(n,d)$ is the topological quotient of the semistable subset $\mathcal{C}(n,d)^{ss}$ of $\mathcal{C}(n,d)$ by the action of $\mathcal{G}_c(n,d)$. Any nonsingular complex projective variety on which a complex reductive group G acts linearly has a G-equivariantly perfect stratification by locally closed nonsingular G-invariant subvarieties with its set of semistable points X^{ss} as an open stratum. This stratification can be obtained as the Morse stratification for the normsquare of a moment map on X [**22**]; in the case of the moduli space $\mathcal{M}(n,d)$ the rôle of the normsquare of the moment map is played by the Yang-Mills functional. In [**28**] this Morse stratification of X is refined to obtain a stratification of X by locally closed nonsingular G-invariant subvarieties with the set X^s of stable points of X as an open stratum. The other strata can be defined inductively in terms of the sets of stable points of closed nonsingular subvarieties of X, acted on by reductive subgroups of G, and their projectivised normal bundles.

In their fundamental paper [**1**], Atiyah and Bott studied a stratification of $\mathcal{C}(n,d)$ defined using the Harder-Narasimhan type of a holomorphic bundle over Σ, which they expected to be the Morse stratification of the Yang-Mills functional (this was later shown to be the case [**5**]). The aim of this paper is to apply the

2000 *Mathematics Subject Classification.* Primary 14D20, 32G13, 14D21.

Key words and phrases. Moduli spaces of vector bundles, Yang-Mills stratification.

The author is a member of VBAC (Vector Bundles on Algebraic Curves), which is partially supported by EAGER (EC FP5 Contract no. HPRN-CT-2000-00099), and acknowledges with gratitude the hospitality of the University of Melbourne and the Australian National University during the writing of part of this paper.

methods of [**28**] to the Yang-Mills stratification to obtain refined stratifications of $\mathcal{C}(n,d)$, and to relate these stratifications to natural refinements of the notion of the Harder-Narasimhan type of a holomorphic bundle over Σ.

The motivation for this study was the search for a complete set of relations among the standard generators for the cohomology of the moduli spaces $\mathcal{M}(n,d)$ when the rank n and degree d are coprime and $n > 2$ [**11**]. The cohomology rings of the moduli spaces $\mathcal{M}(n,d)$ have been the subject of much interest over many years; see for example [**1, 3, 7, 8, 14, 16, 17, 20, 21, 29, 31, 32, 38, 40, 41**] among many other pieces of work. In the case when $n = 2$ we now have a very thorough understanding of the structure of the cohomology ring [**2, 21, 37, 41**]. For arbitrary n it is known that the cohomology has no torsion and formulas for computing the Betti numbers have been obtained, as well as a set of generators for the cohomology ring [**1, 6, 7, 16, 42**]. When $n = 2$ the relations between these generators can be explicitly described and in particular a conjecture of Mumford, that a certain set of relations is a complete set, was proved some years ago [**2, 21, 27, 37, 41**]. However less is known about the relations between the generators when $n > 2$, and the most obvious generalisation of Mumford's conjecture to the cases when $n > 2$ is false, although a modified version of the conjecture (using 'dual Mumford relations' together with the original Mumford relations) is true for $n = 3$ [**9**]. There is, however, a further generalisation of Mumford's relations, and the attempt by Richard Earl and the author to prove that this set is indeed complete was the original stimulus for studying refinements of the Yang-Mills stratification, although a different application has now appeared [**11, 19**].

The layout of this paper is as follows. §1 recalls background material on moduli spaces of bundles and different versions of Mumford's conjecture, while §2 reviews the Morse stratification of the normsquare of a moment map and some refinements of this stratification. §3 studies the structure of subbundles of semistable bundles over Σ which are direct sums of stable bundles all of the same slope, and this is used in §4 to define two canonical refinements of the Harder-Narasimhan filtration of a holomorphic bundle over Σ, and thus to construct two refinements of the Yang-Mills stratification. In the next two sections the stratification defined in §2 is applied to holomorphic bundles over Σ; its indexing set is studied in §5 and the associated strata are investigated in §6. This stratification corresponds to a third refinement of the Harder-Narasimhan filtration whose subquotients are all direct sums of stable bundles of the same slope. The relationship between these three filtrations is considered in §7, and §8 provides a brief conclusion.

1. Background material on moduli spaces of bundles

When n and d are coprime, the generators for the rational cohomology[1] of the moduli space $\mathcal{M}(n,d)$ given by Atiyah and Bott in [**1**] are obtained from a (normalised) universal bundle V over $\mathcal{M}(n,d) \times \Sigma$. With respect to the Künneth decomposition of $H^*(\mathcal{M}(n,d) \times \Sigma)$ the rth Chern class $c_r(V)$ of V can be written as

$$c_r(V) = a_r \otimes 1 + \sum_{j=1}^{2g} b_r^j \otimes \alpha_j + f_r \otimes \omega$$

[1]In this paper all cohomology will have rational coefficients.

where $\{1\}, \{\alpha_j : 1 \leq j \leq 2g\}$, and $\{\omega\}$ are standard bases for $H^0(\Sigma), H^1(\Sigma)$ and $H^2(\Sigma)$, and

$$a_r \in H^{2r}(\mathcal{M}(n,d)), \qquad b_r^j \in H^{2r-1}(\mathcal{M}(n,d)), \qquad f_r \in H^{2r-2}(\mathcal{M}(n,d)), \tag{1.1}$$

for $1 \leq r \leq n$ and $1 \leq j \leq 2g$. It was shown by Atiyah and Bott [**1**, Prop. 2.20 and p.580] that the classes a_r and f_r (for $2 \leq r \leq n$) and b_j^r (for $1 \leq r \leq n$ and $1 \leq j \leq 2g$) generate the rational cohomology ring of $\mathcal{M}(n,d)$.

Since tensoring by a fixed holomorphic line bundle of degree e gives an isomorphism between the moduli spaces $\mathcal{M}(n,d)$ and $\mathcal{M}(n,d+ne)$, we may assume without loss of generality that

$$(2g-2)n < d < (2g-1)n.$$

This implies that $H^1(\Sigma, E) = 0$ for any stable bundle of rank n and degree d [**34**, Lemma 5.2], and hence that $\pi_! V$ is a bundle of rank $d - n(g-1)$ over $\mathcal{M}(n,d)$, where

$$\pi : \mathcal{M}(n,d) \times \Sigma \to \mathcal{M}(n,d)$$

is the projection onto the first component and $\pi_!$ is the K-theoretic direct image map. It follows that

$$c_r(\pi_! V) = 0$$

for $r > d - n(g-1)$. Via the Grothendieck-Riemann-Roch theorem we can express the Chern classes of $\pi_! V$ as polynomials in the generators a_r, b_r^j, f_r described above, and hence their vanishing gives us relations between these generators. These are Mumford's relations, and they give us a complete set of relations when $n = 2$. We can generalise them for $n > 2$ as follows.

Suppose that $0 < \hat{n} < n$, and that $\hat{d}$ is coprime to $\hat{n}$. Then we have a universal bundle $\hat{V}$ over $\mathcal{M}(\hat{n}, \hat{d}) \times \Sigma$, and both V and $\hat{V}$ can be pulled back to $\mathcal{M}(\hat{n}, \hat{d}) \times \mathcal{M}(n,d) \times \Sigma$. If

$$\frac{\hat{d}}{\hat{n}} > \frac{d}{n}$$

then there are no non-zero holomorphic bundle maps from a stable bundle of rank $\hat{n}$ and degree $\hat{d}$ to a stable bundle of rank n and degree d, and hence, if

$$\pi : \mathcal{M}(\hat{n}, \hat{d}) \times \mathcal{M}(n,d) \times \Sigma \to \mathcal{M}(\hat{n}, \hat{d}) \times \mathcal{M}(n,d),$$

is the projection onto the first two components, it follows that $-\pi_!(\hat{V}^* \otimes V)$ is a bundle of rank $n\hat{n}(g-1) - d\hat{n} + \hat{d}n$ over $\mathcal{M}(\hat{n}, \hat{d}) \times \mathcal{M}(n,d)$. Thus

$$0 = c_r(-\pi_!(\hat{V}^* \otimes V)) \in H^*(\mathcal{M}(\hat{n}, \hat{d}) \times \mathcal{M}(n,d))$$

if $r > n\hat{n}(g-1) - d\hat{n} + \hat{d}n$ and hence the slant product

$$c_r(-\pi_!(\hat{V}^* \otimes V)) \backslash \gamma \in H^*(\mathcal{M}(n,d))$$

of $c_r(-\pi_!(\hat{V}^* \otimes V))$ with any homology class $\gamma \in H_*(\mathcal{M}(\hat{n}, \hat{d}))$ vanishes when

$$r > n\hat{n}(g-1) - d\hat{n} + \hat{d}n.$$

The relations between the generators a_r, b_r^j, f_r obtained in this way for $0 < \hat{n} < n$ and

$$\frac{d}{n} + 1 > \frac{\hat{d}}{\hat{n}} > \frac{d}{n}$$

and

$$n\hat{n}(g-1) - d\hat{n} + \hat{d}n < r < n\hat{n}(g+1) - d\hat{n} + \hat{d}n$$

(with a little more care taken when $\hat{n}$ and $\hat{d}$ are not coprime) are the ones we consider. They are essentially Mumford's relations when $n = 2$. To show that these form a complete set of relations, a natural strategy is to consider the Yang-Mills stratification which was used by Atiyah and Bott to obtain their generators for the cohomology ring.

Recall that a holomorphic vector bundle E over Σ is called *semistable* (respectively *stable*) if every holomorphic subbundle D of E satisfies

$$\mu(D) \leq \mu(E), \qquad (\text{respectively } \mu(D) < \mu(E)),$$

where $\mu(D) = \text{degree}(D)/\text{rank}(D)$ is the *slope* of D. Bundles which are not semistable are said to be *unstable*. Note that semistable bundles of coprime rank and degree are stable.

Let $\mathcal{E}$ be a fixed C^∞ complex vector bundle of rank n and degree d over Σ. Let $\mathcal{C}$ be the space of all holomorphic structures on $\mathcal{E}$ and let $\mathcal{G}_c$ denote the group of all C^∞ complex automorphisms of $\mathcal{E}$. Atiyah and Bott [**1**] identify the moduli space $\mathcal{M}(n,d)$ with the quotient $\mathcal{C}^{ss}/\mathcal{G}_c$ where $\mathcal{C}^{ss}$ is the open subset of $\mathcal{C}$ consisting of all semistable holomorphic structures on $\mathcal{E}$. The group $\mathcal{G}_c$ is the complexification of the gauge group $\mathcal{G}$ which consists of all smooth automorphisms of $\mathcal{E}$ which are unitary with respect to a fixed Hermitian structure on $\mathcal{E}$. We shall write $\overline{\mathcal{G}}$ for the quotient of $\mathcal{G}$ by its $U(1)$-centre and $\overline{\mathcal{G}}_c$ for the quotient of $\mathcal{G}_c$ by its $\mathbf{C}^*$-centre. There are natural isomorphisms

$$H^*(\mathcal{M}(n,d)) \cong H^*(\mathcal{C}^{ss}/\mathcal{G}_c) = H^*(\mathcal{C}^{ss}/\overline{\mathcal{G}}_c) \cong H^*_{\overline{\mathcal{G}}_c}(\mathcal{C}^{ss}) \cong H^*_{\overline{\mathcal{G}}}(\mathcal{C}^{ss})$$

between the cohomology of the moduli space and the $\overline{\mathcal{G}}$-equivariant cohomology of $\mathcal{C}^{ss}$, since the $\mathbf{C}^*$-centre of $\mathcal{G}_c$ acts trivially on $\mathcal{C}^{ss}$, while $\overline{\mathcal{G}}_c$ acts freely on $\mathcal{C}^{ss}$ and $\overline{\mathcal{G}}_c$ is the complexification of $\overline{\mathcal{G}}$.

In order to show that the restriction map $H^*_{\overline{\mathcal{G}}}(\mathcal{C}) \to H^*_{\overline{\mathcal{G}}}(\mathcal{C}^{ss})$ is surjective, Atiyah and Bott consider the Yang–Mills (or Atiyah–Bott–Shatz) stratification of $\mathcal{C}$. This stratification $\{\mathcal{C}_\mu : \mu \in \mathcal{M}\}$ is the Morse stratification for the Yang–Mills functional on $\mathcal{C}$, but it also has a more explicit description. It is indexed by the partially ordered set $\mathcal{M}$ consisting of all the Harder–Narasimhan *types* of holomorphic bundles of rank n and degree d, defined as follows. Any holomorphic bundle E over M of rank n and degree d has a filtration

$$0 = E_0 \subset E_1 \subset \cdots \subset E_P = E$$

of subbundles such that the subquotients $Q_p = E_p/E_{p-1}$ are semi-stable for $1 \leq p \leq P$ and satisfy

$$\mu(Q_p) = \frac{d_p}{n_p} > \frac{d_{p+1}}{n_{p+1}} = \mu(Q_{p+1})$$

where d_p and n_p are the degree and rank of Q_p and $\mu(Q_p)$ is its slope. This filtration is canonically associated to E and is called the Harder-Narasimhan filtration of E. We define the *type* of E to be

$$\mu = (\mu(Q_1), ..., \mu(Q_P)) \in \mathbb{Q}^n$$

where the entry $\mu(Q_p)$ is repeated n_p times. The semistable bundles have type $\mu_0 = (d/n, ..., d/n)$ and form the unique open stratum. The set $\mathcal{M}$ of all possible types of holomorphic vector bundles over Σ provides our indexing set, and if $\mu \in \mathcal{M}$ the subset $\mathcal{C}_\mu \subseteq \mathcal{C}$ is defined to be the set of all holomorphic vector bundles over Σ of type μ. A partial order on $\mathcal{M}$ with the property that the closure of the stratum indexed by μ is contained in the union of all strata indexed by $\mu' \geq \mu$ is defined as follows. Let $\sigma = (\sigma_1, ..., \sigma_n)$ and $\tau = (\tau_1, ..., \tau_n)$ be two types; then

$$\sigma \geq \tau \text{ if and only if } \sum_{j \leq i} \sigma_j \geq \sum_{j \leq i} \tau_j \text{ for } 1 \leq i \leq n-1. \tag{1.2}$$

The gauge group $\mathcal{G}$ acts on $\mathcal{C}$ preserving the stratification which is equivariantly perfect with respect to this action, which means that its equivariant Thom-Gysin sequences

$$\cdots \to H_{\mathcal{G}}^{j-2d_\mu}(\mathcal{C}_\mu) \to H_{\mathcal{G}}^j(U_\mu) \to H_{\mathcal{G}}^j(U_\mu - \mathcal{C}_\mu) \to \cdots$$

break up into short exact sequences

$$0 \to H_{\mathcal{G}}^{j-2d_\mu}(\mathcal{C}_\mu) \to H_{\mathcal{G}}^j(U_\mu) \to H_{\mathcal{G}}^j(U_\mu - \mathcal{C}_\mu) \to 0.$$

Here

$$d_\mu = \sum_{i>j} (n_i d_j - n_j d_i + n_i n_j (g-1)), \tag{1.3}$$

is the complex codimension of $\mathcal{C}_\mu$ in $\mathcal{C}$ and U_μ is the open subset of $\mathcal{C}$ which is the union of all those strata labelled by $\mu' \leq \mu$; we also have

$$H_{\mathcal{G}}^*(\mathcal{C}_\mu) \cong \bigotimes_{j=1}^{P} H^*_{\mathcal{G}(n_j,d_j)}(\mathcal{C}(n_j,d_j)^{ss}).$$

Atiyah and Bott show that the stratification is equivariantly perfect by considering the composition of the Thom-Gysin map $H_{\mathcal{G}}^{j-2d_\mu}(\mathcal{C}_\mu) \to H_{\mathcal{G}}^j(U_\mu)$ with restriction to $\mathcal{C}_\mu$, which is multiplication by the equivariant Euler class e_μ of the normal bundle to $\mathcal{C}_\mu$ in $\mathcal{C}$. They show that e_μ is not a zero-divisor in $H_{\mathcal{G}}^*(\mathcal{C}_\mu)$ and deduce that the Thom-Gysin maps $H_{\mathcal{G}}^{j-2d_\mu}(\mathcal{C}_\mu) \to H_{\mathcal{G}}^j(U_\mu)$ are all injective.

So putting this all together Atiyah and Bott obtain inductive formulas for the $\mathcal{G}$-equivariant Betti numbers of $\mathcal{C}^{ss}$, and they also conclude that there is a natural surjection

$$H^*(B\overline{\mathcal{G}}) \cong H^*_{\overline{\mathcal{G}}}(\mathcal{C}) \to H^*_{\overline{\mathcal{G}}}(\mathcal{C}^{ss}) \cong H^*(\mathcal{M}(n,d)). \tag{1.4}$$

Thus generators of the cohomology ring $H^*(B\overline{\mathcal{G}})$ give generators of the cohomology ring $\mathcal{M}(n,d)$.

The classifying space $B\mathcal{G}$ can be identified with the space $\mathrm{Map}_d(\Sigma, BU(n))$ of all smooth maps $f : \Sigma \to BU(n)$ such that the pullback to Σ of the universal vector bundle over $BU(n)$ has degree d. If we pull back this universal bundle using the evaluation map

$$\mathrm{Map}_d(\Sigma, BU(n)) \times \Sigma \to BU(n) : (f, m) \mapsto f(m)$$

then we obtain a rank n vector bundle $\mathcal{V}$ over $B\mathcal{G} \times \Sigma$. If further we restrict the pullback bundle induced by the maps

$$\mathcal{C}^{ss} \times E\mathcal{G} \times \Sigma \to \mathcal{C} \times E\mathcal{G} \times \Sigma \to \mathcal{C} \times_{\mathcal{G}} E\mathcal{G} \times \Sigma \xrightarrow{\simeq} B\mathcal{G} \times \Sigma$$

to $\mathcal{C}^{ss} \times \{e\} \times \Sigma$ for some $e \in E\mathcal{G}$ then we obtain a $\mathcal{G}$-equivariant holomorphic bundle on $\mathcal{C}^{ss} \times \Sigma$. The $\mathbf{C}^*$-centre of $\mathcal{G}$ acts as scalar multiplication on the fibres, and the associated projective bundle descends to a holomorphic projective bundle over $\mathcal{M}(n,d) \times \Sigma$. In fact this is the projective bundle of a holomorphic vector bundle V over $\mathcal{M}(n,d) \times \Sigma$ which has the universal property that, for any $[E] \in \mathcal{M}(n,d)$ representing a bundle E over Σ, the restriction of V to $\{[E]\} \times \Sigma$ is isomorphic to

E.

By a slight abuse of notation we define elements a_r, b_r^j, f_r in $H^*(B\mathcal{G};\mathbb{Q})$ by writing

$$c_r(\mathcal{V}) = a_r \otimes 1 + \sum_{j=1}^{2g} b_r^j \otimes \alpha_j + f_r \otimes \omega \qquad 1 \leq r \leq n.$$

where, as before, ω is the standard generator of $H^2(\Sigma)$ and $\alpha_1, ..., \alpha_{2g}$ form a fixed canonical cohomology basis for $H^1(\Sigma)$. In fact the ring $H^*(B\mathcal{G})$ is freely generated as a graded super-commutative algebra over $\mathbb{Q}$ by the elements

$$\{a_r : 1 \leq r \leq n\} \cup \{b_r^j : 1 \leq r \leq n, 1 \leq j \leq 2g\} \cup \{f_r : 2 \leq r \leq n\}$$

and if we omit a_1 we get $H^*(B\overline{\mathcal{G}})$. These generators restrict to the generators a_r, b_r^j, f_r given in (1.1) for $H^*(\mathcal{M}(n,d))$ under the surjection (1.4).

The relations among these generators for $H^*(\mathcal{M}(n,d);\mathbb{Q})$ are then given by the kernel of the restriction map (1.4) which is in turn determined by the map

$$H^*_{\mathcal{G}}(\mathcal{C}) \cong H^*_{\overline{\mathcal{G}}}(\mathcal{C}) \otimes H^*(BU(1)) \to H^*_{\overline{\mathcal{G}}}(\mathcal{C}^{ss}) \otimes H^*(BU(1)) \cong H^*_{\mathcal{G}}(\mathcal{C}^{ss}), \tag{1.5}$$

and the proof that the Yang–Mills stratification is equivariantly perfect leads to completeness criteria for a set of relations to be complete. Let $\mathcal{R}$ be a subset of the kernel of the restriction map $H^*_{\mathcal{G}}(\mathcal{C}) \to H^*_{\mathcal{G}}(\mathcal{C}^{ss})$. Suppose that for each unstable type $\mu \neq \mu_0$ there is a subset $\mathcal{R}_\mu$ of the ideal generated by $\mathcal{R}$ in $H^*_{\mathcal{G}}(\mathcal{C})$ such that the image of $\mathcal{R}_\mu$ under the restriction map

$$H^*_{\mathcal{G}}(\mathcal{C}) \to H^*_{\mathcal{G}}(\mathcal{C}_\nu)$$

is zero unless $\nu \geq \mu$ and when $\nu = \mu$ contains the ideal of $H^*_{\mathcal{G}}(\mathcal{C}_\mu)$ generated by the equivariant Euler class e_μ of the normal bundle to the stratum $\mathcal{C}_\mu$ in $\mathcal{C}$. Then $\mathcal{R}$ generates the kernel of the restriction map $H^*_{\mathcal{G}}(\mathcal{C}) \to H^*_{\mathcal{G}}(\mathcal{C}^{ss})$ as an ideal of $H^*_{\mathcal{G}}(\mathcal{C})$.

In fact Atiyah and Bott could have replaced the Yang-Mills stratification with a coarser stratification of $\mathcal{C}$ and obtained equivalent results. For any integers n_1 and d_1 let S_{n_1,d_1} be the subset of $\mathcal{C}$ consisting of all those holomorphic structures with Harder-Narasimhan filtration $0 = E_0 \subset E_1 \subset \cdots \subset E_s = E$ where E_1 has rank n_1 and degree d_1. We shall say that such a holomorphic structure has *coarse type* (n_1, d_1). Then S_{n_1,d_1} is locally a submanifold of finite codimension

$$\delta_{n_1,d_1} \quad = \quad nd_1 - n_1 d + n_1(n-n_1)(g-1)$$

in $\mathcal{C}$ and

$$H^*_{\mathcal{G}}(S_{n_1,d_1}) \cong H^*_{\mathcal{G}(n_1,d_1)}(\mathcal{C}(n_1,d_1)^{ss}) \otimes H^*_{\mathcal{G}(n-n_1,d-d_1)}(U(n_1,d_1)) \tag{1.6}$$

where

$$U(n_1,d_1) = \bigcup_{\frac{d_2}{n_2} < \frac{d_1}{n_1}} S(n-n_1, d-d_1)_{n_2,d_2}$$

is an open subset of $\mathcal{C}(n-n_1, d-d_1)$. Moreover the equivariant Euler class e_{n_1,d_1} of the normal to S_{n_1,d_1} in $\mathcal{C}$ is not a zero divisor in $H^*_{\mathcal{G}}(S_{n_1,d_1})$, so the stratification of $\mathcal{C}$ by coarse type

$$\left\{ S_{n_1,d_1} : 0 < n_1 < n, \frac{d_1}{n_1} > \frac{d}{n} \right\} \bigcup \{S_{n,d}\}$$

is equivariantly perfect. This means that we can modify our completeness criteria, so that for each pair of positive integers $(\hat{n}, \hat{d})$ with $0 < \hat{n} < n$ and $\frac{\hat{d}}{\hat{n}} > \frac{d}{n}$ we require a set of relations whose restriction to $H^*_{\mathcal{G}}(S_{n_1,d_1})$ is zero when $d_1/n_1 < \hat{d}/\hat{n}$ or $d_1/n_1 = \hat{d}/\hat{n}$ and $n_1 < \hat{n}$, and when $(n_1, d_1) = (\hat{n}, \hat{d})$ equals the ideal of $H^*_{\mathcal{G}}(S_{\hat{n},\hat{d}})$ generated by the equivariant Euler class $e_{\hat{n},\hat{d}}$ of the normal bundle to $S_{\hat{n},\hat{d}}$ in $\mathcal{C}$.

It is easy enough to prove that if $\gamma \in H_*^{\overline{\mathcal{G}}(\hat{n},\hat{d})}(\mathcal{C}(\hat{n},\hat{d})^s)$ where $\hat{d}/\hat{n} > d/n$ and if $r > n\hat{n}(g-1) + \hat{n}d - n\hat{d}$ then the image of the slant product $c_r(-\pi_!(\hat{\mathcal{V}}^* \otimes \mathcal{V}))\backslash\gamma$ under the restriction map

$$H^*_{\mathcal{G}}(\mathcal{C}) \to H^*_{\mathcal{G}}(S_{n_1,d_1}) \cong H^*_{\mathcal{G}(n_1,d_1)}(\mathcal{C}(n_1,d_1)^{ss}) \otimes H^*_{\mathcal{G}(n-n_1,d-d_1)}(U(n_1,d_1))$$

is zero when $d_1/n_1 < \hat{d}/\hat{n}$, and also when $d_1/n_1 = \hat{d}/\hat{n}$ and $n_1 < \hat{n}$.

By Lefschetz duality, since $\mathcal{C}(\hat{n},\hat{d})^s/\overline{\mathcal{G}}(\hat{n},\hat{d}) = \mathcal{M}^s(\hat{n},\hat{d})$ is a manifold of dimension

$$D(\hat{n},\hat{d}) = 2[(\hat{n}^2-1)(g-1)+g] = 2(\hat{n}^2 g - \hat{n}^2 + 1)$$

we have a natural map

$$LD : H_*^{\overline{\mathcal{G}}(\hat{n},\hat{d})}(\mathcal{C}(\hat{n},\hat{d})^s) \cong H_*(\mathcal{M}^s(\hat{n},\hat{d})) \to H^{D(\hat{n},\hat{d})-*}_{\overline{\mathcal{G}}(\hat{n},\hat{d})}(\mathcal{C}(\hat{n},\hat{d})^{ss})$$

such that if $\gamma \in H_*^{\overline{\mathcal{G}}(\hat{n},\hat{d})}(\mathcal{C}(\hat{n},\hat{d})^s)$ then $LD(\gamma)$ lies in the dual of $H^{\overline{\mathcal{G}}(\hat{n},\hat{d})}_{D(\hat{n},\hat{d})-*}(\mathcal{C}(\hat{n},\hat{d})^{ss})$ and takes a $\overline{\mathcal{G}}(\hat{n},\hat{d})$-equivariant cycle on $\mathcal{C}(\hat{n},\hat{d})^{ss}$ to its intersection, modulo $\overline{\mathcal{G}}(\hat{n},\hat{d})$, with γ. When $\hat{n}$ and $\hat{d}$ are coprime then $\mathcal{C}(\hat{n},\hat{d})^{ss}$ equals $\mathcal{C}(\hat{n},\hat{d})^s$ and its quotient by $\overline{\mathcal{G}}(\hat{n},\hat{d})$, namely $\mathcal{M}(\hat{n},\hat{d})$, is a compact manifold. In this case Lefschetz duality is essentially Poincaré duality and the map LD is an isomorphism.

We need to consider the restriction of a relation of the form $c_r(-\pi_!(\hat{\mathcal{V}}^* \otimes \mathcal{V}))\backslash\gamma$ to $H^*_{\mathcal{G}}(S_{\hat{n},\hat{d}})$. If $\gamma \in H_*^{\overline{\mathcal{G}}(\hat{n},\hat{d})}(\mathcal{C}(\hat{n},\hat{d})^s)$ where $\hat{d}/\hat{n} > d/n$ and if $r = n\hat{n}(g-1) + \hat{n}d - n\hat{d} + 1 + j$, then it turns out that the image of the slant product $c_r(-\pi_!(\hat{\mathcal{V}}^* \otimes \mathcal{V}))\backslash\gamma$ under the restriction map

$$H^*_{\mathcal{G}}(\mathcal{C}) \to H^*_{\mathcal{G}}(S_{\hat{n},\hat{d}}) \cong H^*_{\mathcal{G}(\hat{n},\hat{d})}(\mathcal{C}(\hat{n},\hat{d})^{ss}) \otimes H^*_{\mathcal{G}(n-\hat{n},d-\hat{d})}(U(\hat{n},\hat{d}))$$

is the product $(-a_1^{(1)})^j LD(\gamma) e_{\hat{n},\hat{d}}$ of the equivariant Euler class $e_{\hat{n},\hat{d}}$ of the normal bundle to $S_{\hat{n},\hat{d}}$ in $\mathcal{C}$ with the image of γ under the Lefschetz duality map LD and j copies of minus the generator $a_1^{(1)} \in H^*_{\mathcal{G}(\hat{n},\hat{d})}(\mathcal{C}(\hat{n},\hat{d})^{ss}) \cong H^*(BU(1)) \otimes H^*_{\overline{\mathcal{G}}(\hat{n},\hat{d})}(\mathcal{C}(\hat{n},\hat{d})^s)$ which comes from the copy of the polynomial ring $H^*(BU(1))$. The proof of this is based on Porteous's Formula (as in Beauville's alternative proof [**2**] of the theorem of Atiyah and Bott that the classes a_r, b_r^j, f_r generate $H^*(\mathcal{M}(n,d))$; cf. [**39**] and [**12**]), which allows us to deduce that the Poincaré dual of $\Delta^s \times U(\hat{n},\hat{d})$ in

$$H^*_{\mathcal{G}(\hat{n},\hat{d})}(\mathcal{C}(\hat{n},\hat{d})^s) \otimes H^*_{\mathcal{G}(\hat{n},\hat{d})}(\mathcal{C}(\hat{n},\hat{d})^{ss}) \otimes H^*_{\mathcal{G}(n-\hat{n},d-\hat{d})}(U(\hat{n},\hat{d}))$$

is $c_{\hat{n}^2(g-1)+1}(-\pi_!(\hat{\mathcal{V}}^* \otimes \mathcal{V}_1)$. In other words the restriction of $c_{\hat{n}^2(g-1)+1}(-\pi_!(\hat{\mathcal{V}}^* \otimes \mathcal{V}_1))$ to $\mathcal{C}(\hat{n},\hat{d})^s \times \mathcal{C}(\hat{n},\hat{d})^{ss} \times U(\hat{n},\hat{d})$ is the image of 1 under the $\mathcal{G}(\hat{n},\hat{d}) \times \mathcal{G}(\hat{n},\hat{d}) \times \mathcal{G}(n-\hat{n},d-\hat{d})$-equivariant Thom-Gysin map associated to the inclusion of $\Delta^s \times U(\hat{n},\hat{d}))$ in $\mathcal{C}(\hat{n},\hat{d})^s \times \mathcal{C}(\hat{n},\hat{d})^{ss} \times U(\hat{n},\hat{d}))$. We can express the higher Chern classes of $-\pi_!(\hat{\mathcal{V}}^* \otimes \mathcal{V}_1)$ in a similar way [**11**] by using Fulton's Excess Porteous Formula [**13**].

Recall that given $\eta \in H^*_{\mathcal{G}(\hat{n},\hat{d})}(\mathcal{C}(\hat{n},\hat{d})^{ss}) \otimes H^*_{\mathcal{G}(n-\hat{n},d-\hat{d})}(\mathcal{C}(n-\hat{n},d-\hat{d}))$ we wish to find a relation whose restriction to $\mathcal{C}(n_1,d_1)^{ss} \times U(n_1,d_1)$ is zero when

$d_1/n_1 < \hat{d}/\hat{n}$ or $d_1/n_1 = \hat{d}/\hat{n}$ and $n_1 < n$, and when $(n_1, d_1) = (\hat{n}, \hat{d})$ equals $\eta e_{\hat{n},\hat{d}}$. We have found such a relation when η lies in the image of the Lefschetz duality map LD which maps $H_*^{\overline{\mathcal{G}}(\hat{n},\hat{d})}(\mathcal{C}(\hat{n},\hat{d})^s)$ to $H^*_{\overline{\mathcal{G}}(\hat{n},\hat{d})}(\mathcal{C}(\hat{n},\hat{d})^{ss})$ and thus into

$$H^*_{\mathcal{G}(\hat{n},\hat{d})}(\mathcal{C}(\hat{n},\hat{d})^{ss}) = H^*_{\overline{\mathcal{G}}(\hat{n},\hat{d})}(\mathcal{C}(\hat{n},\hat{d})^{ss}) \otimes H^*(BU(1)),$$

and more generally when η has the form $\eta = (-a_1^{(1)})^j LD(\gamma)$, for some element γ of $H_*^{\overline{\mathcal{G}}(\hat{n},\hat{d})}(\mathcal{C}(\hat{n},\hat{d})^s)$. When $\hat{n}$ and $\hat{d}$ are coprime this gives us all $\eta \in H^*_{\mathcal{G}(\hat{n},\hat{d})}(\mathcal{C}^{ss})$. Moreover

$$H^*_{\mathcal{G}(\hat{n},\hat{d})}(\mathcal{C}(\hat{n},\hat{d})^{ss}) \otimes H^*_{\mathcal{G}(n-\hat{n},d-\hat{d})}(\mathcal{C}(n-\hat{n},d-\hat{d}))$$

is generated as a module over $H^*_{\mathcal{G}}(\mathcal{C})$ by $H^*_{\mathcal{G}(\hat{n},\hat{d})}(\mathcal{C}(\hat{n},\hat{d})^{ss})$, so when $\hat{n}$ and $\hat{d}$ are coprime, we have now obtained the relations we need from the slant products

$$\{c_r(-\pi_!(\hat{\mathcal{V}}^* \otimes \mathcal{V}))\backslash\gamma : r \geq n\hat{n}(g-1) + \hat{n}d - n\hat{d} + 1, \gamma \in H_*^{\overline{\mathcal{G}}(\hat{n},\hat{d})}(\mathcal{C}(\hat{n},\hat{d})^s)\},$$

and a little more work reduces the range of r and $\hat{d}$ (see [**11**] for more details).

This deals with the case when $\hat{n}$ and $\hat{d}$ are coprime, but the completeness criteria have not yet been shown to hold for pairs $\hat{n}$ and $\hat{d}$ with common factors. This was the original motivation for considering further modifications to the Yang–Mills stratification. The difficulty with using the Yang–Mills stratification itself, or the stratification of $\mathcal{C}$ by coarse type, is that in each case, although n and d are chosen to be coprime so that semistability and stability coincide for n and d, in the construction of the stratification other $\hat{n}$ and $\hat{d}$ appear for which semistability and stability do not coincide.

2. Stratifying a set of semistable points

In this section we shall describe briefly how to stratify the set X^{ss} of semistable points of a complex projective variety X equipped with a linear action of a complex reductive group G, so that the set X^s of stable points of X is an open stratum (see [**15, 22, 30, 34**] for background and [**28**] for more details).

We assume that X has some stable points but also has semistable points which are not stable. In [**23, 25**] it is described how one can blow X up along a sequence of nonsingular G-invariant subvarieties to obtain a G-invariant morphism $\tilde{X} \to X$ where $\tilde{X}$ is a complex projective variety acted on linearly by G such that $\tilde{X}^{ss} = \tilde{X}^s$. The set $\tilde{X}^{ss}$ can be obtained from X^{ss} as follows. Let $r > 0$ be the maximal dimension of a reductive subgroup of G fixing a point of X^{ss}, and let $\mathcal{R}(r)$ be a set of representatives of conjugacy classes of all connected reductive subgroups R of dimension r in G such that

$$Z_R^{ss} = \{x \in X^{ss} : \ R \text{ fixes } x\ \}$$

is non-empty. Then $\bigcup_{R\in\mathcal{R}(r)} GZ_R^{ss}$ is a disjoint union of nonsingular closed subvarieties of X^{ss}. The action of G on X^{ss} lifts to an action on the blow-up of X^{ss} along $\bigcup_{R\in\mathcal{R}(r)} GZ_R^{ss}$ which can be linearised so that the complement of the set of semistable points in the blow-up is the proper transform of the subset $\phi^{-1}(\phi(GZ_R^{ss}))$ of X^{ss} where $\phi : X^{ss} \to X//G$ is the quotient map (see [**23**] 7.17). Moreover no semistable point in the blow-up is fixed by a reductive subgroup of G of dimension at least r, and a semistable point in the blow-up is fixed by a reductive subgroup

R of dimension less than r in G if and only if it belongs to the proper transform of the subvariety Z_R^{ss} of X^{ss}.

If we repeat this process enough times, we obtain $\pi : \tilde{X}^{ss} \to X^{ss}$ such that $\tilde{X}^{ss} = \tilde{X}^s$. Equivalently we can construct a sequence

$$X^{ss}_{(R_0)} = X^{ss}, X^{ss}_{(R_1)}, \ldots, X^{ss}_{(R_\tau)} = \tilde{X}^{ss}$$

where $R_1, \ldots, R_\tau$ are connected reductive subgroups of G with

$$r = \dim R_1 \geq \dim R_2 \geq \cdots \dim R_\tau \geq 1,$$

and if $1 \leq l \leq \tau$ then $X_{(R_l)}$ is the blow up of $X^{ss}_{(R_{l-1})}$ along its closed nonsingular subvariety $GZ^{ss}_{R_l} \cong G \times_{N_l} Z^{ss}_{R_l}$, where N_l is the normaliser of R_l in G. Similarly, $\tilde{X}//G = \tilde{X}^{ss}/G$ can be obtained from $X//G$ by blowing up along the proper transforms of the images $Z_R//N$ in $X//G$ of the subvarieties GZ_R^{ss} of X^{ss} in decreasing order of $\dim R$.

If $1 \leq l \leq \tau$ then we have a G-equivariant stratification

$$\{\mathcal{S}_{\beta,l} : (\beta, l) \in \mathcal{B}_l \times \{l\}\}$$

of $X_{(R_l)}$ by nonsingular G-invariant locally closed subvarieties such that one of the strata, indexed by $(0, l) \in \mathcal{B}_l \times \{l\}$, coincides with the open subset $X^{ss}_{(R_l)}$ of $X_{(R_l)}$. Here $\mathcal{B}_l$ is a finite subset of a fixed positive Weyl chamber $\mathbf{t}_+$ in the Lie algebra $\mathbf{t}$ of a maximal compact torus T of G. In fact $\beta \in \mathbf{t}_+$ lies in $\mathcal{B}_l$ if and only if β is the closest point to 0 in the convex hull in $\mathbf{t}$ of some nonempty subset of the set of weights $\{\alpha_0, \ldots, \alpha_n\}$ for the linear action of T on $X_{(R_l)}$.

There is a partial ordering on $\mathcal{B}_l$ given by $\gamma > \beta$ if $\|\gamma\| > \|\beta\|$, with 0 as its minimal element, such that if $\beta \in \mathcal{B}_l$ then the closure in X_l of the stratum $\mathcal{S}_{\beta,l}$ satisfies

$$\overline{\mathcal{S}_{\beta,l}} \subseteq \bigcup_{\gamma \in \mathcal{B}_l, \gamma \geq \beta} \mathcal{S}_{\gamma,l}. \tag{2.1}$$

If $\beta \in \mathcal{B}_l$ and $\beta \neq 0$ then the stratum $\mathcal{S}_{\beta,l}$ retracts G-equivariantly onto its (transverse) intersection with the exceptional divisor E_l for the blow-up $X_{(R_l)} \to X^{ss}_{(R_{l-1})}$. This exceptional divisor is isomorphic to the projective bundle $\mathbb{P}(\mathcal{N}_l)$ over $G\hat{Z}^{ss}_{R_l}$, where $\hat{Z}^{ss}_{R_l}$ is the proper transform of $Z^{ss}_{R_l}$ in $X^{ss}_{(R_{l-1})}$ and $\mathcal{N}_l$ is the normal bundle to $G\hat{Z}^{ss}_{R_l}$ in $X^{ss}_{R_{l-1}}$. The stratification $\{\mathcal{S}_{\beta,l} : \beta \in \mathcal{B}_l\}$ is determined by the action of R_l on the fibres of $\mathcal{N}_l$ over $Z^{ss}_{R_l}$ (see [**23**] §7).

There is thus a stratification $\{\Sigma_\gamma : \gamma \in \Gamma\}$ of X^{ss} indexed by

$$\Gamma = \{R_1\} \sqcup (\mathcal{B}_1 \backslash \{0\}) \times \{1\} \sqcup \cdots \sqcup \{R_\tau\} \sqcup (\mathcal{B}_\tau \backslash \{0\}) \times \{\tau\} \sqcup \{0\} \tag{2.2}$$

defined as follows. We take as the highest stratum Σ_{R_1} the nonsingular closed subvariety $GZ^{ss}_{R_1}$ whose complement in X^{ss} can be naturally identified with the complement $X_{(R_1)} \backslash E_1$ of the exceptional divisor E_1 in $X_{(R_1)}$. We have $GZ^{ss}_{R_1} \cong G \times_{N_1} Z^{ss}_{R_1}$ where N_1 is the normaliser of R_1 in G, and $Z^{ss}_{R_1}$ is equal to the set of semistable points for the action of N_1, or equivalently for the induced action of N_1/R_1, on Z_{R_1}, which is a union of connected components of the fixed point set of R_1 in X. Moreover since R_1 has maximal dimension among those reductive subgroups of G with fixed points in X^{ss}, we have $Z^{ss}_{R_1} = Z^s_{R_1}$ where $Z^s_{R_l}$ denotes the set of stable points for the action of N_l/R_l on Z_{R_l} for $1 \leq l \leq \tau$.

Next we take as strata the nonsingular locally closed subvarieties

$$\Sigma_{\beta,1} = \mathcal{S}_{\beta,1} \backslash E_1 \text{ for } \beta \in \mathcal{B}_1 \text{ with } \beta \neq 0$$

of $X_{(R_1)} \backslash E_1 = X^{ss} \backslash G Z_{R_1}^{ss}$, whose complement in $X_{(R_1)} \backslash E_1$ is just $X^{ss}_{(R_1)} \backslash E_1 = X^{ss}_{(R_1)} \backslash E_1^{ss}$ where $E_1^{ss} = X^{ss}_{(R_1)} \cap E_1$, and then we take the intersection of $X^{ss}_{(R_1)} \backslash E_1$ with $G Z^{ss}_{R_2}$. This intersection is $G Z^s_{R_2}$ where $Z^s_{R_2}$ is the set of stable points for the action of N_2/R_2 on Z_{R_2}, and its complement in $X^{ss}_{(R_1)} \backslash E_1$ can be naturally identified with the complement in $X_{(R_2)}$ of the union of E_2 and the proper transform $\hat{E}_1$ of E_1.

The next strata are the nonsingular locally closed subvarieties

$$\Sigma_{\beta,2} = \mathcal{S}_{\beta,2} \backslash (E_2 \cup \hat{E}_1) \text{ for } \beta \in \mathcal{B}_2 \text{ with } \beta \neq 0$$

of $X_{(R_2)} \backslash (E_2 \cup \hat{E}_1)$, whose complement in $X_{(R_2)} \backslash (E_2 \cup \hat{E}_1)$ is $X^{ss}_{(R_2)} \backslash (E_2 \cup \hat{E}_1)$, and the stratum after these is $G Z^s_{R_3}$. We repeat this process for $1 \leq l \leq \tau$ and take X^s as our final stratum indexed by 0.

The given partial orderings on $\mathcal{B}_1, \ldots, \mathcal{B}_\tau$ together with the ordering in the expression (2.2) above for Γ induce a partial ordering on Γ, with R_1 as the maximal element and 0 as the minimal element, such that the closure in X^{ss} of the stratum Σ_γ indexed by $\gamma \in \Gamma$ satisfies

$$\overline{\Sigma_\gamma} \subseteq \bigcup_{\tilde{\gamma} \geq \gamma} \Sigma_{\tilde{\gamma}}. \tag{2.3}$$

It is possible to describe the strata Σ_γ in more detail. Either Σ_γ is $G Z^s_{R_l}$, for some l, or else it is

$$\mathcal{S}_{\beta,l} \backslash (E_l \cup \hat{E}_{l-1} \cup \ldots \cup \hat{E}_1)$$

for some l and $\beta \in \mathcal{B}_l \backslash \{0\}$, in which case by [**22**] §5 we have

$$\mathcal{S}_{\beta,l} = G Y^{ss}_{\beta,l} \cong G \times_{P_\beta} Y^{ss}_{\beta,l} \tag{2.4}$$

where $Y^{ss}_{\beta,l}$ fibres over $Z^{ss}_{\beta,l}$ with fibre $\mathbb{C}^{m_{\beta,l}}$ for some $m_{\beta,l} > 0$, and P_β is a parabolic subgroup of G with the stabiliser $\mathrm{Stab}(\beta)$ of β under the adjoint action of G as its maximal reductive subgroup. Here the fibration $p_\beta : Y^{ss}_{\beta,l} \to Z^{ss}_{\beta,l}$ sends y to a limit point of its orbit under the complex one-parameter subgroup of R_l generated by β. Moreover

$$\mathcal{S}_{\beta,l} \cap E_l = G(Y^{ss}_{\beta,l} \cap E_l) \cong G \times_{P_\beta} (Y^{ss}_{\beta,l} \cap E_l) \tag{2.5}$$

where $Y^{ss}_{\beta,l} \cap E_l$ fibres over $Z^{ss}_{\beta,l}$ with fibre $\mathbb{C}^{m_{\beta,l}-1}$ (see [**23**] Lemmas 7.6 and 7.11). Thus

$$\mathcal{S}_{\beta,l} \backslash E_l \cong G \times_{P_\beta} (Y^{ss}_{\beta,l} \backslash E_l) \tag{2.6}$$

where $Y^{ss}_{\beta,l} \backslash E_l$ fibres over $Z^{ss}_{\beta,l}$ with fibre $\mathbb{C}^{m_{\beta,l}-1} \times (\mathbb{C} \backslash \{0\})$. Furthermore if $\pi_l : E_l \cong \mathbb{P}(\mathcal{N}_l) \to G \hat{Z}^{ss}_{R_l}$ is the projection then Lemma 7.9 of [**23**] tells us that when $x \in \hat{Z}^{ss}_{R_l}$ the intersection of $\mathcal{S}_{\beta,l}$ with the fibre $\pi_l^{-1}(x) = \mathbb{P}(\mathcal{N}_{l,x})$ of π_l at x is the union of those strata indexed by points in the adjoint orbit $\mathrm{Ad}(G)\beta$ in the stratification of $\mathbb{P}(\mathcal{N}_{l,x})$ induced by the representation ρ_l of R_l on the normal $\mathcal{N}_{l,x}$ to $G \hat{Z}^{ss}_{R_l}$ at x. Careful analysis [**28**] shows that we can, if we wish, replace the indexing set $\mathcal{B}_l \backslash \{0\}$, whose elements correspond to the G-adjoint orbits $\mathrm{Ad}(G)\beta$ of elements of the indexing set for the stratification of $\mathbb{P}(\mathcal{N}_{l,x})$ induced by the representation ρ_l, by the set of their N_l-adjoint orbits $\mathrm{Ad}(N_l)\beta$. Then we still have (2.4) – (2.6), but now if $q_\beta : P_\beta \to \mathrm{Stab}(\beta)$ is the projection we have

$$\Sigma_\gamma = \Sigma_{\beta,l} = \mathcal{S}_{\beta,l} \backslash (E_l \cup \hat{E}_{l-1} \cup \ldots \cup \hat{E}_1) \cong G \times_{Q_{\beta,l}} Y_{\beta,l}^{\backslash E} \tag{2.7}$$

where $Q_{\beta,l} = q_\beta^{-1}(N_l \cap \mathrm{Stab}(\beta))$ and

$$Y_{\beta,l}^{\backslash E} = Y_{\beta,l}^{ss} \backslash (E_l \cup \hat{E}_{l-1} \cup \ldots \cup \hat{E}_1) \cap p_\beta^{-1}\left(Z_{\beta,l}^{ss} \cap \pi_l^{-1}(\hat{Z}_{R_l}^{ss}) \right),$$

and $p_\beta : Y_{\beta,l}^{\backslash E} \to Z_{\beta,l}^{ss} \cap \pi_l^{-1}(\hat{Z}_{R_l}^{ss})$ is a fibration with fibre $\mathbb{C}^{m_{\beta,l}-1} \times (\mathbb{C}\backslash\{0\})$.

This process gives us a stratification $\{\Sigma_\gamma : \gamma \in \Gamma\}$ of X^{ss} such that the stratum indexed by the minimal element 0 of Γ coincides with the open subset X^s of X^{ss}. We shall apply this construction to obtain a stratification of $\mathcal{C}^{ss}$, and thus inductively to refine the Yang-Mills stratification $\{\mathcal{C}_\mu : \mu \in \mathcal{M}\}$ of $\mathcal{C}$ by Harder–Narasimhan type.

3. Direct sums of stable bundles of equal slope

In the good case when n and d are coprime, then $\mathcal{C}^{ss} = \mathcal{C}^s$ and $\mathcal{M}(n,d) = \mathcal{C}^{ss}/\mathcal{G}_c$ is a nonsingular projective variety. When n and d are not coprime, we can use the description of $\mathcal{M}(n,d)$ as the geometric invariant theoretic quotient $\mathcal{C}//\mathcal{G}_c$ to construct a partial desingularisation $\tilde{\mathcal{M}}(n,d)$ of $\mathcal{M}(n,d)$. From this construction we can use the methods described in §2 to obtain a stratification of $\mathcal{C}^{ss}$ with $\mathcal{C}^s$ as an open stratum, and thus (by considering the subquotients of the Harder–Narasimhan filtration) obtain a stratification of $\mathcal{C}$ refining the stratification $\{\mathcal{C}_\mu : \mu \in \mathcal{M}\}$. To understand this refined stratification we need to use the description in [**26**] of the partial desingularisation $\tilde{\mathcal{M}}(n,d)$.

In fact in [**26**] $\tilde{\mathcal{M}}(n,d)$ is not constructed using the representation of $\mathcal{M}(n,d)$ as the geometric invariant theoretic quotient of $\mathcal{C}$ by $\mathcal{G}_c$, although (as is noted at [**26**, p.246]) this representation of $\mathcal{M}(n,d)$ would lead to the same partial desingularisation. Instead in [**26**] $\mathcal{M}(n,d)$ is represented as a geometric invariant theoretic quotient of a finite-dimensional nonsingular quasi-projective variety $R(\hat{n},\hat{d})$ by a linear action of $SL(p;\mathbb{C})$ where $p = \hat{d} + \hat{n}(1-g)$ with $\hat{d} >> 0$. We may assume that $\hat{d} >> 0$, since tensoring by a line bundle of degree l gives an isomorphism of $\mathcal{M}(n,d)$ with $\mathcal{M}(\hat{n},\hat{d}+\hat{n}l)$ for any $l \in \mathbf{Z}$. By [**34**, Lemma 5.2] if E is a semistable bundle over Σ of rank $\hat{n}$ and degree $\hat{d} > \hat{n}(2g-1)$ where g is the genus of Σ, then E is generated by its sections and $H^1(\Sigma,E) = 0$. If $p = \hat{d} + \hat{n}(1-g)$, this implies that $\dim H^0(\Sigma,E) = p$ and that there is a holomorphic map h from Σ to the Grassmannian $G(\hat{n},p)$ of $\hat{n}$-dimensional quotients of $\mathbb{C}^p$ such that the pullback $E(h) = h^*T$ of the tautological bundle T on $G(\hat{n},p)$ is isomorphic to E.

Let $R(\hat{n},\hat{d})$ be the set of all holomorphic maps $h : \Sigma \to G(\hat{n},p)$ such that $E(h) = h^*T$ has degree d and the map on sections $\mathbb{C}^p \to H^0(\Sigma,E(h))$ induced from the quotient bundle map $\mathbb{C}^p \times \Sigma \to E(h)$ is an isomorphism. For $\hat{d} >> 0$ this set $R(\hat{n},\hat{d})$ has the structure of a nonsingular quasi-projective variety and there is a quotient $\mathcal{E}$ of the trivial bundle of rank p over $R(\hat{n},\hat{d}) \times \Sigma$ satisfying the following properties (see [**34**, §5]).

(i) If $h \in R(\hat{n},\hat{d})$ then the restriction of E to $\{h\} \times \Sigma$ is the pullback $E(h)$ of the tautological bundle T on $G(\hat{n},p)$ along the map $h : \Sigma \to G(\hat{n},p)$.

(ii) If h_1 and h_2 lie in $R(\hat{n},\hat{d})$ then $E(h_1)$ and $E(h_2)$ are isomorphic as bundles over Σ if and only if h_1 and h_2 lie in the same orbit of the natural action of $GL(p;\mathbb{C})$ on $R(\hat{n},\hat{d})$.

(iii) If $h \in R(\hat{n},\hat{d})$ then the stabiliser of h in $GL(p;\mathbb{C})$ is isomorphic to the group $\mathrm{Aut}(E(h))$ of complex analytic automorphisms of $E(h)$.

If $N >> 0$ then $R(\hat{n},\hat{d})$ can be embedded as a quasi-projective subvariety of the product $(G(\hat{n},p))^N$ by a map of the form $h \mapsto (h(x_1),\dots,h(x_N))$ where $x_1,\dots,x_N$ are points of Σ. This embedding gives a linearisation of the action of $GL(p;\mathbb{C})$ on $R(\hat{n},\hat{d})$. If $N >> 0$ and $d >> 0$ then we also have the following:

(iv) The point $h \in R(\hat{n},\hat{d})$ is semistable in the sense of geometric invariant theory for the linear action of $SL(p;\mathbb{C})$ on $R(\hat{n},\hat{d})$ if and only if $E(h)$ is a semistable bundle. If h_1 and h_2 lie in $R(\hat{n},\hat{d})^{ss}$ then they represent the same point of $R(\hat{n},\hat{d})//SL(p;\mathbb{C})$ if and only if $\mathrm{gr}(E(h_1)) \cong \mathrm{gr}(E(h_2))$, and thus there is a natural identification of $\mathcal{M}(n,d)$ with $R(\hat{n},\hat{d})//SL(p;\mathbb{C})$ (see for example [**34**, §5]).

It is shown in [**24**] that the Atiyah–Bott formulas for the equivariant Betti numbers of $\mathcal{C}^{ss}$ can be obtained by stratifying $R(n,d)$ instead of $\mathcal{C}(n,d)$, and in fact throughout this paper we could work with either $R(n,d)$ or $\mathcal{C}(n,d)$. In particular properties (i) to (iv) above imply that the analysis in [**26**] of the construction of the partial desingularisation of $\tilde{\mathcal{M}}(\hat{n},\hat{d})$ as $\tilde{R}(\hat{n},\hat{d})//SL(p;\mathbb{C})$ applies equally well if we work with $\tilde{\mathcal{C}}//\mathcal{G}_c$.

To describe the construction of $\tilde{\mathcal{M}}(n,d)$, first of all we need to find a set $\mathcal{R}$ of representatives of the conjugacy classes of reductive subgroups of $SL(p;\mathbb{C})$ which occur as the connected components of stabilisers of semistable points of $R(n,d)$. In fact it is slightly simpler to describe the corresponding subgroups of $G = GL(p;\mathbb{C})$, and since the central one-parameter subgroup of $GL(p;\mathbb{C})$ consisting of nonzero scalar multiples of the identity acts trivially on $R(n,d)$, finding stabilisers in $GL(p;\mathbb{C})$ is essentially equivalent to finding stabilisers in $SL(p;\mathbb{C})$. By [**26**, pp. 248-9] such conjugacy classes in $GL(p;\mathbb{C})$ correspond to unordered sequences $(m_1,n_1),...,(m_q,n_q)$ of pairs of positive integers such that $m_1n_1+...+m_qn_q = n$ and n divides n_id for each i. A representative R of the corresponding conjugacy class is given by

$$R = GL(m_1;\mathbb{C}) \times ... \times GL(m_q;\mathbb{C})$$

embedded in $GL(p;\mathbb{C})$ using a fixed isomorphism of $\mathbb{C}^p$ with $\bigoplus_{i=1}^q(\mathbb{C}^{m_i} \otimes \mathbb{C}^{p_i}))$ where $d_i = n_id/n$ and $p_i = d_i + n_i(1-g) = n_ip/n$. Then GZ_R^{ss} consists of all those holomorphic structures E with

$$E \cong (\mathbb{C}^{m_1} \otimes D_1) \oplus \cdots \oplus (\mathbb{C}^{m_p} \otimes D_q) \tag{3.1}$$

where $D_1,\dots,D_q$ are all semistable and D_i has rank n_i and degree d_i, while GZ_R^s consists of all those holomorphic structures E with

$$E \cong (\mathbb{C}^{m_1} \otimes D_1) \oplus \cdots \oplus (\mathbb{C}^{m_p} \otimes D_q) \tag{3.2}$$

where $D_1,\dots,D_q$ are all stable and not isomorphic to one another, and D_i has rank n_i and rank d_i. Moreover the normaliser N of R in $GL(p;\mathbb{C})$ has connected component

$$N_0 \cong \prod_{1\le i\le q} (GL(m_i;\mathbb{C}) \times GL(p_i;\mathbb{C}))/\mathbb{C}^* \tag{3.3}$$

where $\mathbb{C}^*$ is the diagonal central one-parameter subgroup of $GL(m_i;\mathbb{C})\times GL(p_i;\mathbb{C})$, and $\pi_0(N) = N/N_0$ is the product

$$\pi_0(N) = \prod_{j\geq 0, k\geq 0} \mathrm{Sym}(\#\{i : m_i = j \text{ and } n_i = k\}) \tag{3.4}$$

where $\mathrm{Sym}(b)$ denotes the symmetric group of permutations of a set with b elements. Furthermore in terms of the notation of §2, if $R = R_l$ then a holomorphic structure belongs to one of the strata $\Sigma_{\beta,l}$ with $\beta \in \mathcal{B}_l \backslash \{0\}$ if and only if it has a filtration $0 = E_0 \subset E_1 \subset ... \subset E_s = E$ such that E is not isomorphic to $\bigoplus_{1\leq k\leq s} E_k/E_{k-1}$ but

$$\bigoplus_{1\leq k\leq s} E_k/E_{k-1} \cong (\mathbb{C}^{m_1}\otimes D_1)\oplus\cdots\oplus(\mathbb{C}^{m_p}\otimes D_q)$$

where $D_1,\dots,D_q$ are all stable and not isomorphic to one another, and D_i has rank n_i and rank d_i [**26**, p. 248]. Thus to understand the refined Yang-Mills stratification of $\mathcal{C}$, we need to study refinements

$$0 = E_0 \subset E_1 \subset ... \subset E_s = E$$

of the Harder-Narasimhan filtration of a holomorphic bundle E, such that each subquotient E_j/E_{j-1} is a direct sum of stable bundles all of the same slope. For this recall the following standard result (cf. the proof of [**26**, Lemma 3.2] and [**35**]).

Proposition 3.1. *Any semistable bundle E has a canonical subbundle of the form*

$$(\mathbb{C}^{m_1}\otimes D_1)\oplus\cdots\oplus(\mathbb{C}^{m_q}\otimes D_q)$$

with $D_1,\dots,D_q$ not isomorphic to each other and all stable of the same slope as E, such that any other subbundle of the same form

$$(\mathbb{C}^{m_1'}\otimes D_1')\oplus\cdots\oplus(\mathbb{C}^{m_r'}\otimes D_r'),$$

with $D_1',...,D_r'$ not isomorphic to each other and all stable of the same slope as E, satisfies $r\leq q$ and (after permuting the order of $D_1,...,D_q$ suitably) $D_j\cong D_j'$ for all $1\leq j\leq r$ and the inclusion

$$(\mathbb{C}^{m_1'}\otimes D_1)\oplus\cdots\oplus(\mathbb{C}^{m_r'}\otimes D_r)\to E$$

factors through the inclusion $(\mathbb{C}^{m_1}\otimes D_1)\oplus\cdots\oplus(\mathbb{C}^{m_q}\otimes D_q)\to E$ via injections $\mathbb{C}^{m_j'}\to\mathbb{C}^{m_j}$ for $1\leq j\leq r$.

If we choose a subbundle of E of maximal rank among those of the required form, this follows immediately from

Lemma 3.2. *Let $E, D_1,\dots,D_q, D_1',\dots,D_r'$ be bundles over Σ all of the same slope, with E semistable, $D_1,...,D_q$ and $D_1',\dots,D_r'$ all stable and $D_j = D_j'$ for $1\leq j\leq k$ for some $0\leq k\leq \min\{q,r\}$, but with no other isomorphisms between the bundles $D_1,...,D_q$ and $D_1',...,D_r'$. If*

$$\alpha : (\mathbb{C}^{m_1}\otimes D_1)\oplus\cdots\oplus(\mathbb{C}^{m_q}\otimes D_q)\to E$$

and

$$\beta : (\mathbb{C}^{m_1'}\otimes D_1')\oplus\cdots\oplus(\mathbb{C}^{m_r'}\otimes D_r')\to E$$

are injective bundle homomorphisms, then there exist nonnegative integers $n_1,...,n_k$ and linear injections $i_j:\mathbb{C}^{m_j}\to\mathbb{C}^{n_j}$ and $i_j':\mathbb{C}^{m_j'}\to\mathbb{C}^{n_j}$ for $1\leq j\leq k$ and an injective bundle homomorphism γ from

$$(\mathbb{C}^{n_1}\otimes D_1)\oplus\cdots\oplus(\mathbb{C}^{n_k}\otimes D_k)\oplus(\mathbb{C}^{m_{k+1}}\otimes D_{k+1})\oplus\cdots\oplus(\mathbb{C}^{m_q}\otimes D_q)$$

$$\oplus(\mathbb{C}^{m'_{k+1}} \otimes D'_{k+1}) \oplus \cdots \oplus (\mathbb{C}^{m'_r} \otimes D'_r)$$

to E such that α and β both factorise through γ via the injections i_j and i'_j for $1 \le j \le k$ in the obvious way.

Proof: Consider the kernel of

$$\begin{array}{rcl} \alpha \oplus \beta : (\mathbb{C}^{m_1+m'_1} \otimes D_1) \oplus \cdots \oplus (\mathbb{C}^{m_k+m'_k} \otimes D_k) & \oplus & \\ (\mathbb{C}^{m_{k+1}} \otimes D_{k+1}) \oplus \cdots \oplus (\mathbb{C}^{m_q} \otimes D_q) & \oplus & \\ (\mathbb{C}^{m'_{k+1}} \otimes D'_{k+1}) \oplus \cdots \oplus (\mathbb{C}^{m'_r} \otimes D'_r) & \to & E. \end{array}$$

The proof of [**34**, Lemma 5.1 (iii)] shows that this kernel is a subsheaf of the domain of $\alpha \oplus \beta$ which has the same slope as $E, D_1, ..., D_q$, and $D'_1, ..., D'_r$. Induction on $m_1 + ... + m_q + m'_1 + ... + m'_r$ using [**34**, Lemma 5.1] and the obvious projection from $\mathbb{C}^{m_1}$ onto $\mathbb{C}^{m_1-1}$ shows that such a subsheaf is in fact of the form

$$(U_1 \otimes D_1) \oplus \cdots \oplus (U_q \otimes D_q) \oplus (U_{q+1} \otimes D'_{k+1}) \oplus \cdots \oplus (U_{q-k+r} \otimes D'_r)$$

for some linear subspaces U_j of $\mathbb{C}^{m_1+m'_1}, ..., \mathbb{C}^{m_{k+1}}, ..., \mathbb{C}^{m'_r}$. Then since α and β are injective, we must have $\ker(\alpha \oplus \beta) = (U_1 \otimes D_1) \oplus \cdots \oplus (U_k \otimes D_k)$ for linear subspaces U_j of $\mathbb{C}^{m_j} \oplus \mathbb{C}^{m'_j}$ satisfying $U_j \cap \mathbb{C}^{m_j} = \{0\} = U_j \cap \mathbb{C}^{m'_j}$ for $1 \le j \le k$. The result follows easily.

A similar argument gives us

Corollary 3.3. *Let E be semistable and $D'_1, \ldots, D'_r$ all stable of the same slope as E and not isomorphic to each other. Then*

$$H^0(\Sigma, ((\mathbb{C}^{m'_1} \otimes D'_1) \oplus \cdots \oplus (\mathbb{C}^{m'_r} \otimes D'_r))^* \otimes E) \cong \bigoplus_{j=1}^{k} ((\mathbb{C}^{m'_j})^* \otimes \mathbb{C}^{m_j})$$

where $(\mathbb{C}^{m_1} \otimes D_1) \oplus \cdots \oplus (\mathbb{C}^{m_q} \otimes D_q)$ is the canonical subbundle of this form associated to E as in Proposition 3.1, and without loss of generality we assume $D_j \cong D'_j$ for $1 \le j \le k$ for some $0 \le k \le \min\{q, r\}$ and that there are no other isomorphisms between the bundles $D_1, ..., D_q$ and $D'_1, ..., D'_r$.

Definition 3.4. With the notation above we set

$$\mathrm{gr}(E) = \bigoplus_{i=1}^{q} \bigoplus_{j=1}^{r} (\mathbb{C}^{m_{ij}} \otimes D_i)\,.$$

for any semistable bundle E.

Remark 3.5. Of course, by the Jordan–Hölder theorem, given *any* filtration $0 = D_0 \subset D_1 \subset ... \subset D_t = E$ of E such that D_j/D_{j-1} is a direct sum of stable bundles of the same slope as E, we have $\bigoplus_{j=1}^{t} D_j/D_{j-1} \cong \mathrm{gr}(E)$.

4. Maximal and minimal Jordan–Hölder filtrations

Recall that the Harder-Narasimhan filtration of a holomorphic bundle E over Σ is a canonical filtration

$$0 = F_0 \subset F_1 \subset ... \subset F_s = E$$

of E such that F_j/F_{j-1} is semistable and $\mathrm{slope}(F_j/F_{j-1}) > \mathrm{slope}(F_{j+1}/F_j)$ for $0 < j < s$. In the last section we saw that any semistable bundle E has a canonical maximal subbundle of the form

$$(\mathbb{C}^{m_1} \otimes D_1) \oplus \cdots \oplus (\mathbb{C}^{m_q} \otimes D_q)$$

where $D_1, \ldots, D_q$ are not isomorphic to each other and are all stable of the same slope as E. This subbundle is nonzero if $E \neq 0$, since any nonzero semistable bundle is either itself stable or it has a proper stable subbundle of the same slope.

Therefore any semistable bundle E has a canonical filtration

$$0 = E_0 \subset E_1 \subset E_2 \subset \cdots \subset E_r = E \tag{4.1}$$

whose subquotients are direct sums of stable bundles, which is defined inductively so that

$$E_j/E_{j-1} \cong (\mathbb{C}^{m_{1j}} \otimes D_1) \oplus \cdots \oplus (\mathbb{C}^{m_{qj}} \otimes D_q) \tag{4.2}$$

where $D_1, ..., D_q$ are stable nonisomorphic bundles all of the same slope as E with nonnegative integers m_{ij} for $1 \leq i \leq q$ and $1 \leq j \leq r$, and E_j/E_{j-1} is the maximal subbundle of E/E_{j-1} of this form. If, moreover, we assume that

$$\sum_{j=1}^{r} m_{ij} > 0 \text{ for all } 1 \leq i \leq q$$

then the filtration (4.1), the bundles D_i and integers m_{ij} (for $1 \leq i \leq q$ and $1 \leq j \leq r$) and the decompositions (4.2) are canonically associated to E up to isomorphism of the bundles D_i, the usual action of $GL(m_{ij};\mathbb{C})$ on $\mathbb{C}^{m_{ij}}$ and the obvious action of the permutation group $\mathrm{Sym}(q)$ on this data. We can generalise the definition to the case when E is not necessarily semistable, by applying this construction to the subquotients of the Harder-Narasimhan filtration of E. This gives us a canonical refinement

$$0 = E_0 \subset E_1 \subset ... \subset E_t = E$$

of the Harder-Narasimhan filtration such that each subquotient E_j/E_{j-1} is the maximal subbundle of E/E_{j-1} which is a direct sum of stable bundles all having maximal slope among the nonzero subbundles of E/E_{j-1}. We shall call this refinement of the Harder-Narasimhan filtration the maximal Jordan–Hölder filtration of E.

Definition 4.1. Let s, $q_1, ..., q_s$ and $r_1, ..., r_s$ be positive integers and let d_k, n_{ik} and m_{ijk} (for $1 \leq i \leq q_k$, $1 \leq j \leq r_k$ and $1 \leq k \leq s$) be integers satisfying

$$n_{ik} > 0, \quad m_{ijk} \geq 0, \quad n_k > 0$$

and

$$n = \sum_{k=1}^{s} n_k, \quad d = \sum_{k=1}^{s} d_k, \quad \frac{d_k n_{ik}}{n_k} = d_{ik} \in \mathbb{Z}, \quad \frac{d_1}{n_1} > \frac{d_2}{n_2} > ... > \frac{d_s}{n_s}$$

where

$$n_k = \sum_{i=1}^{q_k} \sum_{j=1}^{r_k} n_{ik} m_{ijk}.$$

Denote by $[\mathbf{d}, \mathbf{n}, \mathbf{m}] = [(d_k)_{k=1}^{s}, (n_{ik})_{i=1,k=1}^{q_k,s}, (m_{ijk})_{i=1,j=1,k=1}^{q_k,r_k,s}]$ the orbit of

$$(\mathbf{d}, \mathbf{n}, \mathbf{m}) = ((d_k)_{k=1}^{s}, (n_{ik})_{i=1,k=1}^{q_k,s}, (m_{ijk})_{i=1,j=1,k=1}^{q_k,r_k,s})$$

under the action of the product of symmetric groups $\Sigma_{q_1} \times ... \times \Sigma_{q_k}$ on the set of such sequences, and let $\mathcal{I} = \mathcal{I}(\hat{n}, \hat{d})$ denote the set of all such orbits, for fixed $\hat{n}$ and $\hat{d}$. Given $[\mathbf{d}, \mathbf{n}, \mathbf{m}] \in \mathcal{I}(\hat{n}, \hat{d})$ let $s([\mathbf{d}, \mathbf{n}, \mathbf{m}] = s$ and let $S^{\max JH}_{[\mathbf{d},\mathbf{n},\mathbf{m}]}$ denote the subset of $\mathcal{C}$ consisting of those holomorphic structures on our fixed smooth bundle of rank $\hat{n}$ and degree $\hat{d}$ whose maximal Jordan–Hölder filtration

$$0 = E_{0,1} \subset E_{1,1} \subset \cdots \subset E_{r_1,1} = E_{0,2} \subset E_{1,2} \subset \cdots$$
$$\cdots \subset E_{r_{s-1},s-1} = E_{r_s,0} \subset \cdots \subset E_{r_s,s} = E$$

satisfies

$$E_{j,k}/E_{j-1,k} \cong (\mathbb{C}^{m_{1jk}} \otimes D_{1k}) \oplus \cdots \oplus (\mathbb{C}^{m_{q_k jk}} \otimes D_{q_k k})$$

for $1 \leq k \leq s$ and $1 \leq j \leq q_k$, where $D_{1k}, \ldots, D_{q_k k}$ are nonisomorphic stable bundles with

$$\mathrm{rank}(D_{ik}) = n_{ik} \text{ and } \deg(D_{ik}) = d_{ik}$$

and $E_{j,k}/E_{j-1,k}$ is the maximal subbundle of $E/E_{j-1,k}$ isomorphic to a direct sum of stable bundles of slope d_k/n_k. To make the notation easier on the eye, $S^{\max JH}_{[\mathbf{d},\mathbf{n},\mathbf{m}]}$ will often be denoted by $S_{[\mathbf{d},\mathbf{n},\mathbf{m}]}$.

Let $\mathcal{I}^{ss}$ denote the subset of $\mathcal{I}$ consisting of all orbits $[\mathbf{d}, \mathbf{n}, \mathbf{m}]$ for which $s([\mathbf{d}, \mathbf{n}, \mathbf{m}] = 1$. For simplicity we shall write $[\mathbf{n}, \mathbf{m}]$ for $[\mathbf{d}, \mathbf{n}, \mathbf{m}]$ when $s([\mathbf{d}, \mathbf{n}, \mathbf{m}]) = 1$ (which means that $\mathbf{d} = (d)$).

We have now proved

LEMMA 4.2. *$\mathcal{C}$ is the disjoint union of the subsets*

$$\{S_{[\mathbf{d},\mathbf{n},\mathbf{m}]} : [\mathbf{d}, \mathbf{n}, \mathbf{m}] \in \mathcal{I}\}.$$

REMARK 4.3. Note that when $s([\mathbf{d}, \mathbf{n}, \mathbf{m}]) = 1$ and $q_1 = r_1 = 1$ and $n_{11} = \hat{n}$ and $m_{111} = 1$ then we get $S_{[(n),(1)]} = \mathcal{C}^s$, and moreover

$$\mathcal{C}(n, d)^{ss} = \bigcup_{[\mathbf{d},\mathbf{n},\mathbf{m}] \in \mathcal{I}, s([\mathbf{d},\mathbf{n},\mathbf{m}])=1} S_{[\mathbf{d},\mathbf{n},\mathbf{m}]} = \bigcup_{[\mathbf{n},\mathbf{m}] \in \mathcal{I}^{ss}} S_{[\mathbf{n},\mathbf{m}]}.$$

REMARK 4.4. Let E be a semistable holomorphic structure on $\mathcal{E}$. If E represents an element of the closure of $S_{[\mathbf{n},\mathbf{m}]}$ in $\mathcal{C}(n, d)^{ss}$ for some

$$[\mathbf{n}, \mathbf{m}] = [(n_i)_{i=1}^q, (m_{ij})_{i=1,j=1}^{q,r}] \in \mathcal{I}^{ss},$$

then E has a filtration $0 = E_0 \subset E_1 \subset E_2 \subset \cdots \subset E_r = E$ such that if $1 \leq j \leq r$ then

$$E_j/E_{j-1} \cong (\mathbb{C}^{m_{1j}} \otimes D_1) \oplus \cdots \oplus (\mathbb{C}^{m_{qj}} \otimes D_q)$$

where $D_1, ..., D_q$ are semistable bundles all having the same slope as E, but this filtration is not necessarily the maximal Jordan–Hölder filtration of E. If the bundles $D_1, ..., D_q$ are not all stable or if two of them are isomorphic to each other, then

$$\dim \mathrm{Aut}(\mathrm{gr}(E)) > \sum_{i=1}^{q} (\sum_{j=1}^{r} m_{ij})^2$$

and so E lies in $S_{[\mathbf{n}', \mathbf{m}']}$ where $\mathbf{m}' = (m'_{ij})_{i=1,j=1}^{q',r'}$ satisfies

$$\sum_{i=1}^{q'} (\sum_{j=1}^{r'} m'_{ij})^2 > \sum_{i=1}^{q} (\sum_{j=1}^{r} m_{ij})^2.$$

If, on the other hand, $D_1, ..., D_q$ are all stable and not isomorphic to each other, then the maximal Jordan–Hölder filtration of E is of the form

$$0 = E'_0 \subset E'_1 \subset E'_2 \subset \cdots \subset E'_r = E$$

with

$$E'_j/E'_{j-1} \cong (\mathbb{C}^{m'_{1j}} \otimes D_1) \oplus \cdots \oplus (\mathbb{C}^{m'_{qj}} \otimes D_q)$$

for $1 \leq j \leq r'$, where $1 \leq r' \leq r$ and

$$m'_{i1} + ... + m'_{ir'} = m_{i1} + ... + m_{ir}$$

and $m'_{i1} + ... + m'_{ij} \geq m_{i1} + ... + m_{ij}$ for $1 \leq i \leq q$ and $1 \leq j \leq r'$. Thus we can define a partial order $\geq$ on $\mathcal{I}^{ss}$ such that $[\mathbf{n}', \mathbf{m}'] \geq [\mathbf{n}, \mathbf{m}]$ if and only if either

$$\sum_{i=1}^{q'} (\sum_{j=1}^{r'} m'_{ij})^2 > \sum_{i=1}^{q} (\sum_{j=1}^{r} m_{ij})^2$$

or $\mathbf{n}' = \mathbf{n}$ and $1 \leq r' \leq r$ and

$$m'_{i1} + ... + m'_{ir'} = m_{i1} + ... + m_{ir}$$

and $m'_{i1} + ... + m'_{ij} \geq m_{i1} + ... + m_{ij}$ for $1 \leq i \leq q$ and $1 \leq j \leq r'$, and then the closure of $S_{[\mathbf{n},\mathbf{m}]}$ in $\mathcal{C}(n,d)^{ss}$ is contained in

$$\bigcup_{[\mathbf{n}',\mathbf{m}'] \geq [\mathbf{n},\mathbf{m}]} S_{[\mathbf{n}',\mathbf{m}']}.$$

Using (1.2), we can then extend this partial order to $\mathcal{I}$ so that

$$\overline{S_{[\mathbf{d},\mathbf{n},\mathbf{m}]}} \subseteq \bigcup_{[\mathbf{d}',\mathbf{n}',\mathbf{m}'] \geq [\mathbf{d},\mathbf{n},\mathbf{m}]} S_{[\mathbf{d}',\mathbf{n}',\mathbf{m}']}.$$

Proposition 4.5. *Let* $[\mathbf{n}, \mathbf{m}] = [(n_i)_{i=1}^q, (m_{ij})_{i=1,j=1}^{q,r}] \in \mathcal{I}^{ss}$. *Then* $S_{[\mathbf{n},\mathbf{m}]}$ *is a locally closed complex submanifold of* $\mathcal{C}^{ss}$ *of finite codimension*

$$\sum_{i=1}^{q} \sum_{j=1}^{r-1} m_{ij} m_{ij+1} + (g-1) \left(\sum_{i_1,i_2=1}^{q} \sum_{1 \leq j_1 \leq j_2 \leq r} m_{i_1 j_1} m_{i_2 j_2} n_{i_1} n_{i_2} - \sum_{i=1}^{q} \sum_{j=1}^{r} (m_{ij} n_i)^2 \right).$$

Proof: (cf. [**1**, §7]) The rank and degree are the only C^∞ invariants of a vector bundle over Σ. Thus we may choose a C^∞ isomorphism of our fixed C^∞ bundle $\mathcal{E}$ over Σ with a bundle of the form

$$\bigoplus_{i=1}^{q} \bigoplus_{j=1}^{r} (\mathbf{C}^{m_{ij}} \otimes \mathcal{D}_i)$$

where $\mathcal{D}_i$ is a fixed C^∞ bundle over Σ of rank n_i and degree $d_i = n_i \hat{d}/\hat{n}$ for $1 \leq i \leq q$.

Let $\mathcal{Y}_{[\mathbf{n},\mathbf{m}]}$ be the subset of $\mathcal{C}^{ss}$ consisting of all semistable holomorphic structures E on $\mathcal{E}$ for which the subbundles

$$E_j = \bigoplus_{i=1}^{q} \bigoplus_{k=1}^{j} (\mathbf{C}^{m_{ik}} \otimes \mathcal{D}_i)$$

are holomorphic for $1 \leq j \leq r$, and for which there are nonisomorphic stable holomorphic structures $D_1, \ldots, D_q$ on the C^∞ bundles $\mathcal{D}_1, \ldots, \mathcal{D}_q$ such that the natural identification of E_j/E_{j-1} with $\bigoplus_{i=1}^q \mathbf{C}^{m_{ij}} \otimes \mathcal{D}_i$ becomes an isomorphism of E_j/E_{j-1} with $\bigoplus_{i=1}^q \mathbf{C}^{m_{ij}} \otimes D_i$ for $1 \leq j \leq r$, and finally for each $1 \leq j \leq r$ the quotient E_j/E_{j-1} is the maximal subbundle of E/E_{j-1} isomorphic to a direct sum of stable bundles of the same slope as E/E_{j-1}. Let $\mathcal{G}_c[\mathbf{n},\mathbf{m}]$ be the subgroup of the complexified gauge group $\mathcal{G}_c$ consisting of all C^∞ complex automorphisms of $\mathcal{E}$ which preserve the filtration of $\mathcal{E}$ by the subbundles $\bigoplus_{i=1}^q \bigoplus_{k=1}^j (\mathbf{C}^{m_{ik}} \otimes \mathcal{D}_i)$ and the decomposition of E_j/E_{j-1} as $\bigoplus_{i=1}^q (\mathbf{C}^{m_{ij}} \otimes \mathcal{D}_i)$ up to the action of the general linear groups $GL(m_{ij};\mathbb{C})$ and the permutation groups

$$\mathrm{Sym}(\#\{i : n_i = k \text{ and } m_{ij} = l_j, j = 1, ..., r\})$$

for all nonnegative integers k and $l_1, ..., l_r$. Since the filtration (4.1) and decompositions of E_1/E_0,...,E_r/E_{r-1} are canonically associated to E up to the actions of these general linear and permutation groups, we have

$$S_{[\mathbf{n},\mathbf{m}]} = \mathcal{G}_c\mathcal{Y}_{[\mathbf{n},\mathbf{m}]} \cong \mathcal{G}_c \times_{\mathcal{G}_c[\mathbf{n},\mathbf{m}]} \mathcal{Y}_{[\mathbf{n},\mathbf{m}]}.$$

As in [**1**, §7] we have that $\mathcal{Y}_{[\mathbf{n},\mathbf{m}]}$ is an open subset of an affine subspace of the infinite-dimensional affine space $\mathcal{C}$ and the injection

$$\mathcal{G}_c \times_{\mathcal{G}_c[\mathbf{n},\mathbf{m}]} \mathcal{Y}_{[\mathbf{n},\mathbf{m}]} \to \mathcal{C} \tag{4.3}$$

is holomorphic with image $S_{[\mathbf{n},\mathbf{m}]}$.

If $E \in \mathcal{Y}_{[\mathbf{n},\mathbf{m}]}$, let $\mathrm{End}'E$ be the subbundle of $\mathrm{End}\,E$ consisting of holomorphic endomorphisms of E preserving the maximal Jordan–Hölder filtration (4.1) and decomposition (4.2) up to isomorphism of the bundles D_i and the vector spaces $\mathbb{C}^{m_{ij}}$. Let $\mathrm{End}''E$ be the quotient of $\mathrm{End}\,E$ by $\mathrm{End}'E$. The normal to the $\mathcal{G}_c$-orbit of E in $\mathcal{C}$ can be canonically identified with $H^1(\Sigma, \mathrm{End}\,E)$ (see [**1**, §7]) and the image of $T_E\mathcal{Y}_{[\mathbf{n},\mathbf{m}]}$ in this can be canonically identified with the image of the natural map

$$H^1(\Sigma, \mathrm{End}'E) \to H^1(\Sigma, \mathrm{End}\,E)$$

which fits into the long exact sequence of cohomology induced by the short exact sequence of bundles

$$0 \to \mathrm{End}'E \to \mathrm{End}\,E \to \mathrm{End}''E \to 0.$$

Thus we get an isomorphism

$$T_E\mathcal{C}/(T_E\mathcal{Y}_{[\mathbf{n},\mathbf{m}]} + T_E\mathcal{O}) \cong H^1(\Sigma, \mathrm{End}''E)$$

where $\mathcal{O}$ is the $\mathcal{G}_c$-orbit of E in $\mathcal{C}$.

We have short exact sequences

$$0 \to (E/E_1)^* \otimes E \to \mathrm{End}\,E \to E_1^* \otimes E \to 0$$

and

$$0 \to E_1^* \otimes E_1 \to E_1^* \otimes E \to E_1^* \otimes (E/E_1) \to 0.$$

Let K be the kernel of the composition of surjections

$$\mathrm{End}\,E \to E_1^* \otimes E \to E_1^* \otimes (E/E_1).$$

Then we have short exact sequences

$$0 \to K \to \mathrm{End}\,E \to E_1^* \otimes (E/E_1) \to 0$$

and

$$0 \to (E/E_1)^* \otimes E \to K \to E_1^* \otimes E_1 \to 0.$$

Now $\mathrm{End}'E \subseteq K$ and the image of $\mathrm{End}'E$ in $E_1^* \otimes E_1$ is

$$\bigoplus_{i=1}^{q} \mathrm{End}(\mathbb{C}^{m_{i1}}) \otimes \mathrm{End}(D_i),$$

so since $\mathrm{End}''E = \mathrm{End}\,E/\mathrm{End}'E$ we get short exact sequences

$$0 \to \frac{K}{\mathrm{End}'E} \to \mathrm{End}''E \to E_1^* \otimes \left(\frac{E}{E_1}\right) \to 0 \tag{4.4}$$

and

(4.5)

$$0 \to \frac{\mathrm{End}'E + ((E/E_1)^* \otimes E)}{\mathrm{End}'E} \to \frac{K}{\mathrm{End}'E} \to \frac{E_1^* \otimes E_1}{\bigoplus_{i=1}^{q} \mathrm{End}(\mathbb{C}^{m_{i1}}) \otimes \mathrm{End}(D_i)} \to 0.$$

Since

$$\begin{aligned} \frac{\mathrm{End}'E + ((E/E_1)^* \otimes E)}{\mathrm{End}'E} &\cong \frac{(E/E_1)^* \otimes E}{\mathrm{End}'E \cap ((E/E_1)^* \otimes E)} \\ &\cong \frac{((E/E_1)^* \otimes E)/((E/E_1)^* \otimes E_1)}{(\mathrm{End}'E \cap ((E/E_1)^* \otimes E))/((E/E_1)^* \otimes E_1)} \\ &\cong \frac{(E/E_1)^* \otimes (E/E_1)}{\mathrm{End}'(E/E_1)} = \mathrm{End}''(E/E_1), \end{aligned}$$

the short exact sequence (4.5) becomes

$$0 \to \mathrm{End}''\left(\frac{E}{E_1}\right) \to \frac{K}{\mathrm{End}'E} \to \frac{E_1^* \otimes E_1}{\bigoplus_{i=1}^{q} \mathbb{C}^{m_{i1}^2} \otimes D_i^* \otimes D_i} \to 0. \tag{4.6}$$

From the sequences (4.4) and (4.6) it follows that the rank of $\mathrm{End}''E$ is

$$\begin{aligned} \mathrm{rank}(\mathrm{End}''E) &= \mathrm{rank}(K/\mathrm{End}'E) + \mathrm{rank}(E_1^* \otimes (E/E_1)) \\ &= \mathrm{rank}(\mathrm{End}''(E/E_1)) + \mathrm{rank}(E_1^* \otimes E_1) - \sum_{i=1}^{q} (m_{i1})^2 \mathrm{rank}(D_i^* \otimes D_i) \\ &\quad + \mathrm{rank}(E_1^* \otimes (E/E_1)) \\ &= \mathrm{rank}(\mathrm{End}''(E/E_1)) + \mathrm{rank}(E_1^* \otimes E) - \sum_{i=1}^{q} (m_{i1})^2 (n_i)^2. \end{aligned}$$

Thus by induction on r we have

$$\mathrm{rank}(\mathrm{End}''E) = \sum_{i_1,i_2=1}^{q} \sum_{1 \le j_1 \le j_2 \le r} m_{i_1 j_1} m_{i_2 j_2} n_{i_1} n_{i_2} - \sum_{i=1}^{q} \sum_{j=1}^{r} (n_i)^2.$$

Since $D_1, \dots, D_q$ all have the same slope $\hat{d}/\hat{n}$ as E we have $\deg(\mathrm{End}''E) = 0$. Therefore by Riemann-Roch

$$\dim H^1(\Sigma, \mathrm{End}''E) = \dim H^0(\Sigma, \mathrm{End}''E)$$
$$+(g-1) \left(\sum_{i_1,i_2=1}^{q} \sum_{1 \le j_1 \le j_2 \le r} m_{i_1 j_1} m_{i_2 j_2} n_{i_1} n_{i_2} - \sum_{i=1}^{q} \sum_{j=1}^{r} (n_i)^2 \right).$$

Moreover the short exact sequences (4.4) and (4.6) give us long exact sequences of cohomology

$$0 \to H^0(\Sigma, K/\mathrm{End}'E) \to H^0(\Sigma, \mathrm{End}''E) \to H^0(\Sigma, E_1^* \otimes (E/E_1)) \to \cdots$$

and

$$0 \to H^0(\Sigma, \mathrm{End}''(E/E_1)) \to H^0(\Sigma, K/\mathrm{End}'E)$$
$$\to H^0(\Sigma, E_1^* \otimes E_1/\bigoplus_{i=1}^q \mathrm{End}(\mathbb{C}^{m_{ij}}) \otimes \mathrm{End}(D_i)) \to \cdots$$

Now $E_1 \cong \bigoplus_{i=1}^q \mathbb{C}^{m_{i1}} \otimes D_i$ where $D_1, ..., D_q$ are nonisomorphic stable bundles all of the same slope as E_1, and

$$E_2/E_1 \cong \bigoplus_{i=1}^q \mathbb{C}^{m_{i2}} \otimes D_i$$

is the maximal subbundle of E/E_1 which is a direct sum of stable bundles all of the same slope as E/E_1. Since E_1 and E/E_1 have the same slope, it follows from Corollary 3.3 that

$$H^0(\Sigma, E_1^* \otimes (E/E_1)) = H^0(\Sigma, E_1^* \otimes (E_2/E_1)) \cong \bigoplus_{i=1}^q (\mathbb{C}^{m_{i1}})^* \otimes \mathbb{C}^{m_{i2}}.$$

By choosing an open cover $\mathcal{U}$ of Σ such that the filtration $0 = E_0 \subset E_1 \subset E_2 \subset \cdots \subset E_r = E$ is trivial over each $U \in \mathcal{U}$, and describing E in terms of upper triangular transition functions on $E_1 \oplus (E_2/E_1) \oplus ... \oplus (E/E_{r-1})|_{U\cap V}$ for $U, V \in \mathcal{U}$ which induce the identity on E_j/E_{j-1} for $1 \le j \le r$, we see that the natural map

$$H^0(\Sigma, \mathrm{End}''E) \to H^0(\Sigma, E_1^* \otimes (E/E_1)) = H^0(\Sigma, E_1^* \otimes (E_2/E_1))$$

is surjective. Also

$$H^0(\Sigma, E_1^* \otimes E_1/\bigoplus_{i=1}^q \mathrm{End}(\mathbb{C}^{m_{i1}} \otimes \mathrm{End}(D_i)) = \bigoplus_{i\neq j} (\mathbb{C}^{m_{i1}})^* \otimes \mathbb{C}^{m_{i1}} \otimes H^0(\Sigma, D_i^* \otimes D_j) = 0,$$

so

$$\dim H^0(\Sigma, \mathrm{End}''E) = \sum_{i=1}^q m_{i1} m_{i2} + \dim H^0(\Sigma, \mathrm{End}''(E/E_1))$$

and thus by induction on r we have

$$\dim H^0(\Sigma, \mathrm{End}''E) = \sum_{i=1}^q \sum_{j=1}^{r-1} m_{ij} m_{ij+1}.$$

Therefore $\dim H^1(\Sigma, \mathrm{End}''E)$ is equal to

(4.7)
$$\sum_{i=1}^q \sum_{j=1}^{r-1} m_{ij} m_{ij+1} + (g-1) \left(\sum_{i_1, i_2 = 1}^q \sum_{1 \le j_1 \le j_2 \le r} m_{i_1 j_1} m_{i_2 j_2} n_{i_1} n_{i_2} - \sum_{i=1}^q \sum_{j=1}^r (m_{ij} n_i)^2 \right).$$

In particular this tells us that the image in $\mathcal{C}$ of the derivative of the injection (4.3) has constant codimension, and it follows as in [**1**, §7] (see also [**1**, §§14 and 15]) that the subset $S_{[\mathbf{n},\mathbf{m}]}$ is locally a complex submanifold of $\mathcal{C}$ of finite codimension given by (4.7).

Corollary 4.6. *If $[\mathbf{d}, \mathbf{n}, \mathbf{m}] \in \mathcal{I}$ then $S_{[\mathbf{d},\mathbf{n},\mathbf{m}]}$ is a locally closed complex submanifold of $\mathcal{C}$ of codimension*

$$\sum_{1 \le k_2 < k_1 \le s} (n_{k_1} d_{k_2} - n_{k_2} d_{k_1} + n_{k_1} n_{k_2}(g-1)) + \sum_{k=1}^s \sum_{i=1}^q \sum_{j=1}^{r-1} m_{ijk} m_{ij+1k}$$

$$+(g-1)\sum_{k=1}^{s}\left(\sum_{i_1,i_2=1}^{q}\sum_{1\leq j_1\leq j_2\leq r} m_{i_1j_1k}m_{i_2j_2k}n_{i_1k}n_{i_2k}-\sum_{i=1}^{q}\sum_{j=1}^{r}(m_{ijk}n_{ik})^2\right).$$

Proof: This follows immediately from (1.3), Lemma 4.5 and the definition of $S_{[\mathbf{d},\mathbf{n},\mathbf{m}]}$ (Definition 4.1).

Remark 4.7. Let $0=E_0\subset E_1\subset ...\subset E_t=E$ be the maximal Jordan–Hölder filtration of a bundle E. Then the kernels of the duals of the inclusions $E_j\to E$ give us a filtration

$$0=F_0\subset F_1\subset ...\subset F_t=E'$$

of the dual E' of E, such that if $1\leq j\leq t$ then

$$F_j/F_{j-1}\cong (E_{t-j+1}/E_{t-j})'.$$

Thus F_j/F_{j-1} is a direct sum of stable bundles all of the same slope, and moreover F_{j-1} is the minimal subbundle of F_j such that F_j/F_{j-1} is a direct sum of stable bundles all of which have minimal slope among quotients of F_j.

Applying this construction with E replaced by E', we find that every holomorphic bundle E over Σ has a canonical filtration

$$0=F_0\subset F_1\subset ...\subset F_t=E,$$

which we will call the *minimal Jordan–Hölder filtration* of E, with the property that if $1\leq j\leq t$ then F_{j-1} is the minimal subbundle of F_j such that F_j/F_{j-1} is a direct sum of stable bundles all of which have minimal slope among quotients of F_j.

The minimal and maximal Jordan–Hölder filtrations of a bundle do not necessarily coincide. For example, consider the direct sum $E\oplus F$ of two semistable bundles with maximal Jordan–Hölder filtrations $0=E_0\subset E_1\subset ...\subset E_t=E$ and $0=F_0\subset F_1\subset ...\subset F_s=F$ where without loss of generality we may assume that $s\leq t$. If E and F have the same slope, then it is easy to check that the maximal Jordan–Hölder filtration of $E\oplus F$ is

$$0=E_0\oplus F_0\subset E_1\oplus F_1\subset ...\subset E_s\oplus F_s\subset E_{s+1}\oplus F_s\subset ...\subset E_t\oplus F_s;$$

that is, it is the direct sum of the maximal Jordan–Hölder filtrations of E and F with the shorter one extended trivially at the top. Similarly the minimal Jordan–Hölder filtration of $E\oplus F$ is the direct sum of the minimal Jordan–Hölder filtrations of E and F with the shorter one extended trivially at the bottom. Thus if the minimal and maximal Jordan–Hölder filtrations of E and F coincide (which will be the case if, for example, each of the subquotients E_j/E_{j-1} and F_j/F_{j-1} are stable) but these filtrations are not of the same length, then the minimal Jordan–Hölder filtration

$$0=E_0\oplus F_0\subset E_1\oplus F_0\subset ...\subset E_{t-s}\oplus F_0\subset E_{t-s+1}\oplus F_1\subset ...\subset E_t\oplus F_s$$

of $E\oplus F$ will be different from its maximal Jordan–Hölder filtration.

Definition 4.8. Given $[\mathbf{d},\mathbf{n},\mathbf{m}]\in\mathcal{I}$, let $S^{\min JH}_{[\mathbf{d},\mathbf{n},\mathbf{m}]}$ denote the subset of $\mathcal{C}$ consisting of those holomorphic structures on our fixed smooth bundle of rank $\hat{n}$ and degree $\hat{d}$ whose minimal Jordan–Hölder filtration is of the form

$$0=E_{0,1}\subset E_{1,1}\subset\cdots\subset E_{r_1,1}=E_{0,2}\subset E_{1,2}\subset\cdots$$
$$\cdots\subset E_{r_{s-1},s-1}=E_{r_s,0}\subset\cdots\subset E_{r_s,s}=E$$

with

$$E_{j,k}/E_{j-1,k} \cong (\mathbb{C}^{m_{1jk}} \otimes D_{1k}) \oplus \cdots \oplus (\mathbb{C}^{m_{q_k jk}} \otimes D_{q_k k})$$

for $1 \le k \le s$ and $1 \le j \le q_k$, where $D_{1k}, \dots, D_{q_k k}$ are nonisomorphic stable bundles with

$$\mathrm{rank}(D_{ik}) = n_{ik} \text{ and } \deg(D_{ik}) = d_{ik}$$

and $E_{j-1,k}$ is the minimal subbundle of $E_{j,k}$ such that $E_{j,k}/E_{j-1,k}$ is a direct sum of stable bundles of slope d_k/n_k.

5. More indexing sets

In this section we will consider the indexing set Γ for the stratification $\{\Sigma_\gamma : \gamma \in \Gamma\}$ of $\mathcal{C}^{ss}$ defined as in §2.

If $\gamma \in \Gamma$ then by (2.2) either $\gamma = 0$ or $\gamma = R_l$ or $\gamma \in \mathcal{B}_l \backslash \{0\} \times \{l\}$ for some $1 \le l \le \tau$. If $\gamma = 0$ then $\Sigma_\gamma = \mathcal{C}^s$, while by [**26**, pp.248-9] if $\gamma = R_l$ then there exists $[\mathbf{n}, \mathbf{m}] = [(n_i)_{i=1}^q, (m_{ij})_{i=1,j=1}^{q,r}] \in \mathcal{I}^{ss}$ with $r = 1$ and $q = q_1$, such that

$$R_l = \prod_{i=1}^q GL(m_i; \mathbb{C}) \tag{5.1}$$

where $m_i = m_{i1}$, and Σ_{R_l} consists of all those holomorphic structures E with

$$E \cong (\mathbb{C}^{m_1} \otimes D_1) \oplus \cdots \oplus (\mathbb{C}^{m_q} \otimes D_q) \tag{5.2}$$

where $D_1, \dots, D_q$ are all stable of slope $\hat{d}/\hat{n}$ and not isomorphic to one another.

In order to describe the strata $\{\Sigma_{\beta,l} : \beta \in \mathcal{B}_l \backslash \{0\}\}$ more explicitly, we need to look at the action of R_l on the normal $\mathcal{N}_{R_l}$ to $GZ_{R_l}^{ss}$ at a point represented by a holomorphic structure E of the form (5.2), and to understand the stratification on $\mathbb{P}(\mathcal{N}_{R_l})$ induced by this action of R_l. If we choose a C^∞ isomorphism of our fixed C^∞ bundle $\mathcal{E}$ with $(\mathbb{C}^{m_1} \otimes D_1) \oplus \cdots \oplus (\mathbb{C}^{m_q} \otimes D_q)$, then we can identify $\mathcal{C}$ with the infinite-dimensional vector space

$$\Omega^{0,1}(\mathrm{End}((\mathbb{C}^{m_1} \otimes D_1) \oplus \cdots \oplus (\mathbb{C}^{m_q} \otimes D_q)))$$

and the normal to the $\mathcal{G}_c$-orbit at E can be identified with $H^1(\Sigma, \mathrm{End}\, E)$, where $\mathrm{End}\, E$ is the bundle of holomorphic endomorphisms of E [**1**, §7]. If δ_i^j denotes the Kronecker delta then the normal to GZ_R^{ss} can be identified with

$$H^1(\Sigma, \mathrm{End}'_\oplus E) \cong \bigoplus_{i_1, i_2 = 1}^q \mathbb{C}^{m_{i_1} m_{i_2} - \delta_{i_1}^{i_2}} \otimes H^1(\Sigma, D_{i_1}^* \otimes D_{i_2})$$

where $\mathrm{End}'_\oplus E$ is the quotient of the bundle $\mathrm{End}\, E$ of holomorphic endomorphisms of E by the subbundle $\mathrm{End}_\oplus E$ consisting of those endomorphisms which preserve the decomposition (5.2). The action of $R_l = \prod_{i=1}^q GL(m_i; \mathbb{C})$ on this is given by the natural action on $\mathbb{C}^{m_{i_1} m_{i_2} - \delta_{i_1}^{i_2}}$ identified with the set of $m_{i_1} \times m_{i_2}$ matrices if $i_1 \ne i_2$ and the set of trace-free $m_{i_1} \times m_{i_1}$ matrices if $i_1 = i_2$; its weights α are therefore of the form $\alpha = \xi - \xi'$ where ξ and ξ' are weights of the standard representation of R_l on $\oplus_{i=1}^q \mathbb{C}^{m_i}$ (see [**26**, pp.251-2] noting the error immediately before (3.18)).

Any element β of the indexing set $\mathcal{B}_l$ is represented by the closest point to 0 of the convex hull of some nonempty set of these weights, and two such closest points can be taken to represent the same element of $\mathcal{B}_l$ if and only if they lie in the same $\mathrm{Ad}(N_l)$-orbit, where N_l is the normaliser of R_l (see [**22**] or [**28**]). By (3.3) the orbit of β under the adjoint action of the connected component of N_l is just its $\mathrm{Ad}(R_l)$-orbit, and so by (3.4) the $\mathrm{Ad}(N_l)$-orbit of β is the union

$$\bigcup_{w \in \pi_0(N_l)} w.\mathrm{Ad}(N_l)(\beta)$$

where $\pi_0(N_l)$ is the product of permutation groups

$$\pi_0(N_l) = \prod_{j \geq 0, k \geq 0} \mathrm{Sym}(\#\{i : m_i = j \text{ and } n_i = k\}).$$

We can describe this indexing set $\mathcal{B}_l$ more explicitly as follows. Let us take our maximal compact torus T_l in R_l to be the product of the standard maximal tori of the unitary groups $U(m_1)$,..., $U(m_q)$ consisting of the diagonal matrices, and let $\mathbf{t}_l$ be its Lie algebra. Let

$$M = m_1 + ... + m_q$$

and let $e_1, ..., e_M$ be the weights of the standard representation of T_l on $\mathbb{C}^{m_1} \oplus ... \oplus \mathbb{C}^{m_q}$. We take the usual invariant inner product on the Lie algebra $\mathbf{u}(p)$ of $U(p)$ given by $\langle A, B \rangle = -\mathrm{tr}\, A\bar{B}^t$ and restrict it to T_l. Since R_l is embedded in $GL(p; \mathbb{C})$ by identifying $\oplus_{i=1}^{q}(\mathbb{C}^{m_i} \otimes \mathbb{C}^{p_i})$ with $\mathbb{C}^p$, it follows that $e_1, ..., e_M$ are mutually orthogonal and $\|e_j\|^2 = 1/p_i$ if $m_1 + ... + m_{i-1} < j \leq m_1 + ... + m_i$.

Proposition 5.1. *Let β be any nonzero element of the Lie algebra $\mathbf{t}_l$ of the maximal compact torus T_l of R_l. Then β represents an element of $\mathcal{B}_l \backslash \{0\}$ if and only if there is a partition*

$$\{\Delta_{h,m} : (h, m) \in J\}$$

of $\{1, ..., M\}$, indexed by a subset J of $\mathbb{Z} \times \mathbb{Z}$ of the form

$$J = \{(h, m) \in \mathbb{Z} \times \mathbb{Z} : 1 \leq h \leq L, l_1(h) \leq m \leq l_2(h)\}$$

for some positive integer L and functions l_1 and $l_2 : \{1, ..., L\} \to \mathbb{Z}$ such that $l_1(h) \leq l_2(h)$ for all $h \in \{1, ..., L\}$, with the following properties. If

$$r_{h,m} = \sum_{j \in \Delta_{h,m}} \|e_j\|^{-2}$$

then the function $\epsilon : \{1, ..., L\} \to \mathbb{Q}$ defined by

$$\epsilon(h) = \left(\sum_{m=l_1(h)}^{l_2(h)} m r_{h,m} \right) \left(\sum_{m=l_1(h)}^{l_2(h)} r_{h,m} \right)^{-1}$$

satisfies $-1/2 \leq \epsilon(h) < 1/2$ and $\epsilon(1) > \epsilon(2) > ... > \epsilon(L)$, and

$$\frac{\beta}{\|\beta\|^2} = \sum_{h=1}^{L} \sum_{m=l_1(h)}^{l_2(h)} \sum_{j \in \Delta_{h,m}} (\epsilon(h) - m) \frac{e_j}{\|e_j\|^2}.$$

Remark 5.2. Note that because of the conditions on the function ϵ, the partition $\{\Delta_{h,m} : (h, m) \in J\}$ and its indexing can be recovered from the coefficients of β with respect to the basis $e_1/\|e_1\|^2, ..., e_M/\|e_M\|^2$ of $\mathbf{t}_l$.

Proof of Proposition 5.1: $\beta \in \mathbf{t}_l \backslash \{0\}$ represents an element of $\mathcal{B}_l \backslash \{0\}$ if and only if it is the closest point to 0 of the convex hull of

$$\{e_i - e_j : (i, j) \in S\}$$

for some nonempty subset S of $\{(i,j) \in \mathbb{Z} \times \mathbb{Z} : 1 \leq i,j \leq M\}$. Then β can be expressed in the form

$$\beta = \sum_{(i,j)\in S} \lambda^\beta_{ij}(e_i - e_j)$$

for some $\lambda^\beta_{ij} \in \mathbb{R}$ for $(i,j) \in S$ such that $\lambda^\beta_{ij} \geq 0$ and $\sum_{(i,j)\in S} \lambda^\beta_{ij} = 1$. Replacing S with its subset $\{(i,j) \in S : \lambda^\beta_{ij} > 0\}$ we may assume that $\lambda^\beta_{ij} > 0$ for all $(i,j) \in S$. Moreover clearly if $S = \{e_i - e_j\}$ has just one element then $\beta = e_i - e_j$, and we then take $J = \{(1,0),(1,1)\}$ with $\Delta_{(1,0)} = \{j\}$ and $\Delta_{(1,1)} = \{i\}$, so we can assume without loss of generality that $\lambda^\beta_{ij} < 1$ for all $(i,j) \in S$. Since $\beta \neq 0$ we can also assume that the convex hull of $\{e_i - e_j : (i,j) \in S\}$ does not contain 0.

In order to find the closest point to 0 of the convex hull of $\{e_i - e_j : (i,j) \in S\}$ we minimise

$$\| \sum_{(i,j)\in S} \lambda_{ij}(e_i - e_j)\|^2$$

subject to the constraints that $\lambda_{ij} \geq 0$ for all $(i,j) \in S$ and $\sum_{(i,j)\in S} \lambda_{ij} = 1$. Since the weights $e_1,...,e_M$ are mutually orthogonal, we have

$$\begin{aligned}\| \sum_{(i,j)\in S} \lambda_{ij}(e_i - e_j)\|^2 &= \| \sum_{i=1}^M (\sum_{j:(i,j)\in S} \lambda_{ij} - \sum_{j:(j,i)\in S} \lambda_{ji})e_i\|^2 \\ &= \sum_{i=1}^M (\sum_{j:(i,j)\in S} \lambda_{ij} - \sum_{j:(j,i)\in S} \lambda_{ji})^2 \|e_i\|^2.\end{aligned}$$

Using the method of Lagrange multipliers, we consider

$$\sum_{i=1}^M (\sum_{j:(i,j)\in S} \lambda_{ij} - \sum_{j:(j,i)\in S} \lambda_{ji})^2 \|e_i\|^2 - \lambda(\sum_{(i,j)\in S} \lambda_{ij} - 1).$$

If $(i,j) \in S$ then $i \neq j$ and $(j,i) \notin S$ since the convex hull of $\{e_i - e_j : (i,j) \in S\}$ does not contain 0, so

$$\frac{\partial}{\partial \lambda_{ij}}(\sum_{i=1}^M (\sum_{j:(i,j)\in S} \lambda_{ij} - \sum_{j:(j,i)\in S} \lambda_{ij})^2 \|e_i\|^2 - \lambda(\sum_{(i,j)\in S} \lambda_{ij} - 1))$$

is equal to

$$2(\sum_{k:(i,k)\in S} \lambda_{ik} - \sum_{k:(k,i)\in S} \lambda_{ki})\|e_i\|^2 - 2(\sum_{k:(j,k)\in S} \lambda_{jk} - \sum_{k:(k,j)\in S} \lambda_{kj})\|e_j\|^2 - \lambda.$$

Thus $\beta = \sum_{(i,j)\in S} \lambda^\beta_{ij}(e_i - e_j)$ where for each $(i,j) \in S$ we have either $\lambda^\beta_{ij} = 0$ or $\lambda^\beta_{ij} = 1$ (both of which are ruled out by the assumptions on S) or

$$(5.3) \quad (\sum_{k:(i,k)\in S} \lambda_{ik} - \sum_{k:(k,i)\in S} \lambda_{ki})\|e_i\|^2 - (\sum_{k:(j,k)\in S} \lambda_{jk} - \sum_{k:(k,j)\in S} \lambda_{kj})\|e_j\|^2 = \lambda/2$$

where λ is independent of $(i,j) \in S$.

From S we can construct a directed graph $G(S)$ with vertices $1,...,M$ and directed edges from i to j whenever $(i,j) \in S$. Let Δ_1,..., Δ_N be the connected components of this graph. Then $\{e_i - e_j : (i,j) \in S\}$ is the disjoint union of its subsets $\{e_i - e_j : (i,j) \in S, i,j \in \Delta_h\}$ for $1 \leq h \leq N$, and $\{e_i - e_j : (i,j) \in S, i,j \in \Delta_h\}$ is contained in the vector subspace of $\mathbf{t}_l$ spanned by the basis vectors

$\{e_k : k \in \Delta_h\}$. Since these subspaces are mutually orthogonal for $1 \leq h \leq N$, it follows that

$$\beta = \left(\sum_{h=1}^{L} \frac{1}{\|\beta_h\|^2} \right)^{-1} \sum_{h=1}^{L} \frac{\beta_h}{\|\beta_h\|^2} \tag{5.4}$$

where

$$\beta_h = \sum_{(i,j)\in S, i,j\in\Delta_h} \lambda_{ij}^{\beta h}(e_i - e_j)$$

is the closest point to 0 of the convex hull of $\{e_i - e_j : (i,j) \in S, i,j \in \Delta_h\}$ for $1 \leq h \leq N$, and without loss of generality we assume that β_h is nonzero when $1 \leq h \leq L$ and zero when $L < h \leq N$. Note that then

$$\|\beta\|^2 = \left(\sum_{h=1}^{L} \frac{1}{\|\beta_h\|^2} \right)^{-2} \sum_{h=1}^{L} \frac{\|\beta_h\|^2}{\|\beta_h\|^4} = \left(\sum_{h=1}^{L} \frac{1}{\|\beta_h\|^2} \right)^{-1} \tag{5.5}$$

so that

$$\frac{\beta}{\|\beta\|^2} = \sum_{h=1}^{L} \frac{\beta_h}{\|\beta_h\|^2} \tag{5.6}$$

and

$$\lambda_{ij}^{\beta h} = (\|\beta_h\|^2/\|\beta\|^2)\lambda_{ij}^{\beta}$$

if $i,j \in \Delta_h$. For $1 \leq h \leq L$ let κ_h be defined by

$$\kappa_h = \max\{(\sum_{k:(i,k)\in S} \lambda_{ik}^{\beta} - \sum_{k:(k,i)\in S} \lambda_{ki}^{\beta})\|e_i\|^2 : i \in \Delta_h\}$$

where a sum over the empty set is interpreted as 0. Then by (5.3) for $1 \leq h \leq L$ we can express Δ_h as a disjoint union

$$\Delta_h = \hat{\Delta}_{h,0} \sqcup \hat{\Delta}_{h,1} \sqcup \ldots \sqcup \hat{\Delta}_{h,l_h}$$

where

$$\hat{\Delta}_{h,m} = \{i \in \Delta_h : (\sum_{j:(i,j)\in S} \lambda_{ij}^{\beta} - \sum_{j:(j,i)\in S} \lambda_{ji}^{\beta})\|e_i\|^2 = \kappa_h - m|\lambda|/2\}. \tag{5.7}$$

Let us assume that $\lambda \leq 0$; the argument is similar if $\lambda \geq 0$. Then (5.3) tells us that if $(i,j) \in S$ then there exist h and m such that $i \in \hat{\Delta}_{h,m}$ and $j \in \hat{\Delta}_{h,m+1}$. Note that if $\hat{\Delta}_{h,m_1}$ and $\hat{\Delta}_{h,m_2}$ are nonempty then so is $\hat{\Delta}_{h,m}$ whenever $m_1 < m < m_2$, so without loss of generality we may assume that $\hat{\Delta}_{h,m}$ is nonempty when $1 \leq h \leq L$ and $0 \leq m \leq l_h$. Thus if $1 \leq h \leq L$ we have

$$\begin{aligned}\beta_h &= (\|\beta_h\|^2/\|\beta\|^2) \sum_{i\in\Delta_h} (\sum_{j\in\Delta_h,(i,j)\in S} \lambda_{ij}^{\beta} - \sum_{j\in\Delta_h,(j,i)\in S} \lambda_{ji}^{\beta})e_i \\ &= (\|\beta_h\|^2/\|\beta\|^2) \sum_{m=1}^{l_h} \sum_{i\in\hat{\Delta}_{h,m}} (\kappa_h - m|\lambda|/2)\frac{e_i}{\|e_i\|^2}.\end{aligned}$$

For $1 \leq h \leq L$ and $0 \leq m \leq l_h$ let $r_{h,m} = \sum_{j\in\hat{\Delta}_{h,m}} \|e_j\|^{-2}$; then by (5.7) we have

$$\sum_{i\in\hat{\Delta}_{h,m}} (\sum_{j:(i,j)\in S} \lambda_{ij}^{\beta} - \sum_{j:(j,i)\in S} \lambda_{ji}^{\beta}) = r_{h,m}(\kappa_h - m|\lambda|/2\}).$$

Recall that if $(i,j) \in S$ then $i \in \hat{\Delta}_{h,m}$ if and only if $j \in \hat{\Delta}_{h,m+1}$, so we get

$$\sum_{i \in \hat{\Delta}_{h,0}} \sum_{j:(i,j) \in S} \lambda_{ij}^{\beta} = r_{h,0}\kappa_h,$$

and hence

$$\sum_{i \in \hat{\Delta}_{h,1}} \sum_{j:(i,j)\in S} \lambda_{ij}^{\beta} = \sum_{i \in \hat{\Delta}_{h,1}} (\sum_{j:(i,j)\in S} \lambda_{ij}^{\beta} - \sum_{j:(j,i)\in S} \lambda_{ji}^{\beta}) + \sum_{j \in \hat{\Delta}_{h,0}} \sum_{j:(j,i)\in S} \lambda_{ji}^{\beta}$$
$$= r_{h,1}(\kappa_h - |\lambda|/2\}) + r_{h,0}\kappa_h,$$

and similarly

$$\sum_{i \in \hat{\Delta}_{h,m}} \sum_{j:(i,j)\in S} \lambda_{ij}^{\beta} = (r_{h,0} + r_{h,1} + ... + r_{h,m})\kappa_h - (r_{h,1} + 2r_{h,2} + ... + mr_{h,m})\frac{|\lambda|}{2}$$

if $1 \leq m \leq l_h$. Since $\sum_{(i,j)\in S, i,j \in \Delta_h} \lambda_{ij}^{\beta h} = 1$, it follows that

$$\|\beta\|^2/\|\beta_h\|^2 = ((l_h+1)r_{h,0} + l_h r_{h,1} + ... + r_{h,l_h})\kappa_h$$
$$-(l_h r_{h,1} + 2(l_h-1)r_{h,2} + 3(l_h-2)r_{h,3} + ... + l_h r_{h,l_h})|\lambda|/2,$$

and since $\sum_{(i,j)\in S, i,j\in\Delta_h} \lambda_{ij}^{\beta} - \sum_{(j,i)\in S,i,j\in\Delta_h} \lambda_{ji}^{\beta} = 0$ we have

$$0 = (r_{h,0} + r_{h,1} + ... + r_{h,l_h})\kappa_h - (r_{h,1} + 2r_{h,2} + ... + l_h r_{h,l_h})|\lambda|/2.$$

Thus

$$|\lambda|/2 = (\|\beta\|^2/\|\beta_h\|^2)(r_{h,0} + r_{h,1} + ... + r_{h,l_h})/\mu_h$$

and

$$\kappa_h = (\|\beta\|^2/\|\beta_h\|^2)(r_{h,1} + 2r_{h,2} + ... + l_h r_{h,l_h})/\mu_h$$

where

$$\mu_h = \sum_{i,j=0}^{l_h} ((l_h - i + 1)j - j(l_h - j + 1))r_{h,i}r_{h,j}$$
$$= \sum_{0 \leq i < j \leq l_h} ((j-i)j + (i-j)i)r_{h,i}r_{h,j} = \sum_{0\leq i<j\leq l_h} (j-i)^2 r_{h,i}r_{h,j}.$$

Therefore

$$\beta_h = \sum_{m=0}^{l_h}\sum_{j=0}^{l_h} \frac{(j-m)r_{h,j}}{\mu_h} \sum_{i \in \hat{\Delta}_{h,m}} \frac{e_i}{\|e_i\|^2}$$

and

$$\|\beta_h\|^2 = (r_{h,0} + r_{h,1} + ... + r_{h,l_h}))/\mu_h.$$

By defining $\Delta_{h,m} = \hat{\Delta}_{h,m-l_1(h)}$ for an appropriate integer $l_1(h)$, we can arrange that the function ϵ defined in the statement of the proposition takes values in the interval $[-1/2, 1/2)$, and then by amalgamating those Δ_h for which $\epsilon(h)$ takes the same value and rearranging them so that ϵ is a strictly decreasing function, we can assume that the required conditions on ϵ are satisfied, and we have

$$\frac{\beta}{\|\beta\|^2} = \sum_{h=1}^{L} \sum_{m=l_1(h)}^{l_2(h)} \sum_{i \in \Delta_{h,m}} (\epsilon(h) - m)\frac{e_i}{\|e_i\|^2}.$$

This gives us all the required properties if $L = N$; that is, if $\bigcup_{(h,m)\in J} \Delta_{h,m}$ is equal to $\{1, ..., M\}$. Otherwise we amalgamate $\{1, ..., M\} \setminus \bigcup_{(h,m)\in J} \Delta_{h,m}$ with Δ_{h_0,m_0} where (h_0, m_0) is the unique element of J such that $\epsilon(h_0) = 0 = m_0$ if such an element exists, and if there is no such element of J then we adjoin $(L+1,0)$ to J and define

$$\Delta_{L+1,0} = \{1,...,M\}\setminus \bigcup_{h=1}^{L} \bigcup_{m=l_1(h)}^{l_2(h)} \Delta_{h,m}.$$

Conversely, suppose that we are given any partition $\{\Delta_{h,m} : (h,m) \in J\}$ of $\{1,...,M\}$ indexed by

$$J = \{(h,m) \in \mathbb{Z}\times\mathbb{Z} : 1 \le h \le L, l_1(h) \le m \le l_2(h)\}$$

for some positive integer L and functions l_1 and $l_2 : \{1,...,L\} \to \mathbb{Z}$ with $l_1 \le l_2$, satisfying $\epsilon(h) \in [-1/2, 1/2)$ and $\epsilon(1) > \epsilon(2) > ... > \epsilon(L)$ where

$$\epsilon(h) = \left(\sum_{m=l_1(h)}^{l_2(h)} m r_{h,m}\right)\left(\sum_{m=l_1(h)}^{l_2(h)} r_{h,m}\right)^{-1}$$

for $r_{h,m} = \sum_{i\in\Delta_{h,m}} \|e_i\|^{-2}$. Suppose also that

$$\beta = \hat\beta/\|\hat\beta\|^2$$

(or equivalently $\hat\beta = \beta/|\|\beta|\|^2$) where

$$\hat\beta = \sum_{h=1}^{L}\sum_{m=l_1(h)}^{l_2(h)}\sum_{j\in\Delta_{h,m}} (\epsilon(h)-m)\frac{e_j}{\|e_j\|^2}.$$

It suffices to show that β is the closest point to 0 of the convex hull of $\{e_i - e_j : (i,j) \in S\}$, where S is the set of ordered pairs (i,j) with $i,j \in \{1,...,M\}$ such that $e_i \in \Delta_{h,m}$ and $e_j \in \Delta_{h,m+1}$ for some $(h,m) \in J$ such that $(h,m+1) \in J$. For this, it is enough to prove firstly that β lies in the convex hull of $\{e_i - e_j : (i,j) \in S\}$ and secondly that $(e_i - e_j).\beta = \|\beta\|^2$ (or equivalently that $(e_i - e_j).\hat\beta = 1$) for all $(i,j) \in S$. The latter follows easily from the choice of S: if $(i,j) \in S$ then there exists $(h,m) \in J$ such that $(h,m+1) \in J$ and $e_i \in \Delta_{h,m}$ and $e_j \in \Delta_{h,m+1}$, so

$$(e_i - e_j).\hat\beta = \epsilon(h) - m - \epsilon(h) + m + 1 = 1.$$

To show that β lies in the convex hull of $\{e_i - e_j : (i,j) \in S\}$, we note that

$$\sum_{m=l_1(h)}^{l_2(h)-1}\sum_{i\in\Delta_{h,m}}\sum_{j\in\Delta_{h,m+1}}\sum_{k=l_1(h)}^{m} \frac{\epsilon(h)r_{h,k} - k r_{h,k}}{r_{h,m}r_{h,m+1}\|e_i\|^2\|e_j\|^2}(e_i - e_j)$$

$$= \sum_{m=l_1(h)}^{l_2(h)}\sum_{j\in\Delta_{h,m}} (\epsilon(h)-m)\frac{e_j}{\|e_j\|^2}.$$

This means that

$$\beta = \sum_{(i,j)\in S} \lambda_{ij}^{\beta}(e_i - e_j)$$

where

$$\frac{\lambda_{ij}^{\beta}}{\|\beta\|^2} = \sum_{k=l_1(h)}^{m} \frac{\epsilon(h)r_{h,k} - k r_{h,k}}{r_{h,m}r_{h,m+1}\|e_i\|^2\|e_j\|^2} \tag{5.8}$$

if $i \in \Delta_{h,m}$ and $j \in \Delta_{h,m+1}$ for some $(h,m) \in J$ such that $(h, m+1) \in J$. Then

$$\sum_{(i,j)\in S} \frac{\lambda_{ij}^{\beta}}{\|\beta\|^2} = \sum_{h=1}^{L} \sum_{m=l_1(h)}^{l_2(h)-1} \sum_{i\in\Delta_{h,m}} \sum_{j\in\Delta_{h,m+1}} \sum_{k=l_1(h)}^{m} \frac{\epsilon(h) r_{h,k} - k r_{h,k}}{r_{h,m} r_{h,m+1} \|e_i\|^2 \|e_j\|^2}$$

$$= \sum_{h=1}^{L} \sum_{k=l_1(h)}^{l_2(h)-1} \sum_{m=k}^{l_2(h)-1} (\epsilon(h) - k) r_{h,k} = \sum_{h=1}^{L} \sum_{k=l_1(h)}^{l_2(h)-1} (l_2(h) - k)(\epsilon(h) - k) r_{h,k}.$$

Expanding the brackets, using the definition of ϵ and replacing the index k by m shows that this equals

$$\sum_{h=1}^{L} \sum_{m=l_1(h)}^{l_2(h)} (m^2 - \epsilon(h)^2) r_{h,m} = \sum_{h=1}^{L} \sum_{m=l_1(h)}^{l_2(h)} (\epsilon(h) - m)^2 r_{h,m}$$

$$= \sum_{h=1}^{L} \sum_{m=l_1(h)}^{l_2(h)} \sum_{i\in\Delta_{h,m}} \frac{(\epsilon(h) - m)^2}{\|e_i\|^2} = \frac{1}{\|\beta\|^2},$$

and so $\sum_{(i,j)\in S} \lambda_{ij}^{\beta} = 1$. Finally note that

$$\Big(\sum_{k_1=l_1(h)}^{m} (\epsilon(h) - k_1) r_{h,k_1}\Big)\Big(\sum_{k_2=l_1(h)}^{l_2(h)} r_{h,k_2}\Big) = \sum_{k_1=l_1(h)}^{m} \sum_{k_2=l_1(h)}^{l_2(h)} (k_2 r_{h,k_2} - k_1 r_{h,k_2}) r_{h,k_1}.$$

This sum is positive because if $k_2 > m$ then the contribution of the pair (k_1, k_2) to the sum is $(k_2 - k_1) r_{h,k_1} r_{h,k_2} > 0$, whereas if $k_2 \leq m$ then the total contribution of the pairs (k_1, k_2) and (k_2, k_1) is zero. Thus by (5.8) we have $\lambda_{ij}^{\beta} \geq 0$ for all $(i,j) \in S$, and hence β lies in the convex hull of $\{e_i - e_j : (i,j) \in S\}$ as required.

LEMMA 5.3. *If β satisfies the conditions of Proposition 5.1 and if $i \in \Delta_{h,m}$ and $j \in \Delta_{h',m'}$, then*

$$\beta.(e_i - e_j) = \|\beta\|^2$$

if and only if $h' = h$ and $m' = m+1$, and

$$\beta.(e_i - e_j) \geq \|\beta\|^2$$

if and only if either $m' \geq m+2$ or $m' = m+1$ and $h' \geq h$.

Proof: This follows immediately from the formula for β and conditions on the function ϵ in the statement of Proposition 5.1.

REMARK 5.4. Let β be as in Proposition 5.1. Then there is a unique bijection $\phi : J \to \{1, ..., t\}$ from the indexing set J of the partition $\{\Delta_{h,m} : (h,m) \in J\}$ of $\{1, ..., M\}$ to the set of positive integers $\{1, ..., t\}$, where t is the size of J, which takes the Hebrew lexicographic ordering on J to the standard ordering on integers; that is, $\phi(h,m) \leq \phi(h',m')$ if and only if either $m < m'$ or $m = m'$ and $h \leq h'$. We can define an increasing function

$$\delta : \{1, ..., t\} \to \{1, ..., t\}$$

such that $\delta(\phi(h,m))$ is the number of elements $(h',m') \in J$ such that either $m' < m+1$ or $m' = m+1$ and $h' < h$. Then $\delta(k) \geq k$ for all $k \in \{1, ..., t\}$, and if (h,m) and $(h,m+1)$ both belong to J then $\delta(\phi(h,m)) = \phi(h,m+1) - 1$ and $\delta(\phi(h,m)) < \delta(\phi(h,m)+1)$. Conversely if

$$k_2 - 1 = \delta(k_1) < \delta(k_1 + 1)$$

then there exists $(h,m) \in J$ with $(h,m+1) \in J$ such that $k_1 = \phi(h,m)$ and $k_2 = \phi(h,m+1)$.

When it is helpful to make the dependence on β explicit, we shall write $\delta_\beta : \{1,...,t_\beta\} \to \{1,...,t_\beta\}$ and $\{\delta_{h,m}(\beta) : (h,m) \in J_\beta\}$.

Lemma 5.3 tells us that if $i \in \Delta_{h,m}$ and $j \in \Delta_{h',m'}$ then $\beta.(e_i - e_j) \geq \|\beta\|^2$ if and only if $\phi(h',m') > \delta(\phi(h,m))$.

Definition 5.5. Recall that if $1 \leq i \leq q$ then $e_{m_1+...+m_{i-1}+1}, ..., e_{m_1+...+m_i}$ are the weights of the standard representation on $\mathbb{C}^{m_i}$ of the component $GL(m_i;\mathbb{C})$ of $R_l = \prod_{i=1}^q GL(m_i;\mathbb{C})$. If β and $\phi : J \to \{1,...,t\}$ are as in Remark 5.4 and $1 \leq i \leq q$ and $1 \leq k \leq t$, then set

$$\Delta^k = \Delta_{\phi^{-1}(k)}, \quad \Delta_i^k = \Delta_{\phi^{-1}(k)} \cap \{m_1 + ... + m_{i-1} + 1, ..., m_1 + ... + m_i\}$$

and let m_i^k denote the size of Δ_i^k, so that $m_i^1 + ... + m_i^t = m_i$.

Remark 5.6. By Remark 5.2 the partition $\{\Delta^k(\beta) : 1 \leq k \leq t_\beta\}$ of $\{1,...,M\}$ and the function $\delta_\beta : \{1,...,t_\beta\} \to \{1,...,t_\beta\}$ are determined by β. Conversely, from the partition $\{\Delta^k(\beta) : 1 \leq k \leq t_\beta\}$ of $\{1,...,M\}$ and the function $\delta_\beta : \{1,...,t_\beta\} \to \{1,...,t_\beta\}$ we can recover β as the closest point to 0 of the convex hull of

$$\{e_i - e_j : i \in \Delta^{k_1}(\beta) \text{ and } j \in \Delta^{k_2}(\beta) \text{ where } k_2 > \delta_\beta(k_1)\}.$$

Note, however, that although given any partition $\{\Delta^k : 1 \leq k \leq t\}$ of $\{1,...,M\}$ and increasing function $\delta : \{1,...,t\} \to \{1,...,t\}$ satisfying $\delta(k) \geq k$ for $1 \leq k \leq t$, we can consider the closest point β to 0 of the convex hull of

$$\{e_i - e_j : i \in \Delta^{k_1} \text{ and } j \in \Delta^{k_2} \text{ where } k_2 > \delta(k_1)\},$$

it is not necessarily the case that the associated partition $\{\Delta^k(\beta) : 1 \leq k \leq t_\beta\}$ of $\{1,...,M\}$ and function $\delta_\beta : \{1,...,t_\beta\} \to \{1,...,t_\beta\}$ coincide with the given partition $\{\Delta^k : 1 \leq k \leq t\}$ of $\{1,...,M\}$ and function $\delta : \{1,...,t\} \to \{1,...,t\}$. For example, some amalgamation and rearrangement may be needed as in the proof of Proposition 9.1.

6. Balanced δ-filtrations

The last section studied the indexing set Γ for the stratification $\{\Sigma_\gamma : \gamma \in \Gamma\}$ of $\mathcal{C}^{ss}$ defined as in §2. In this section we will consider what it means for a semistable holomorphic bundle over the Riemann surface Σ to belong to a stratum $\Sigma_\gamma = \Sigma_{\beta,l}$, where β is as in Proposition 5.1.

Definition 6.1. We shall say that a semistable bundle E has a δ-filtration

$$0 = E_0 \subset E_1 \subset ... \subset E_t = E$$

with associated function $\delta : \{1,...,t\} \to \{1,...,t\}$ if δ is an increasing function such that if $1 \leq k \leq t$ then $\delta(k) \geq k$ and the induced filtration

$$0 \subset \frac{E_k}{E_{k-1}} \subset \frac{E_{k+1}}{E_{k-1}} \subset ... \subset \frac{E_{\delta(k)}}{E_{k-1}}$$

is trivial.

Let $G(\delta)$ be the graph with vertices $1,...,t$ and edges joining i to j if $j - 1 = \delta(i) < \delta(i+1)$. Then the connected components of $G(\delta)$ are of the form

$$\{i^h_{l_1(h)}, ..., i^h_{l_2(h)}\}$$

for $1 \leq h \leq L$, where $i_j^h - 1 = \delta(i_{j-1}^h) < \delta(i_{j-1}^h + 1)$ if $1 < j \leq s_h$, and $i_1^h - 1$ is not in the image of δ and either $\delta(i_{s_h}^h) = u$ or $\delta(i_{s_h}^h) = \delta(i_{s_h}^h + 1)$. Moreover $l_1(h) \leq l_2(h)$ can be chosen so that if

$$\epsilon(h) = \left(\sum_{m=l_1(h)}^{l_2(h)} m\tilde{r}_{h,m} \right) \left(\sum_{m=l_1(h)}^{l_2(h)} \tilde{r}_{h,m} \right)^{-1},$$

where $\tilde{r}_{h,m} = \operatorname{rank}(E_{i_m^h}/E_{i_m^h - 1})$, then $-1/2 \leq \epsilon(h) < 1/2$, and the ordering of the components of $G(\delta)$ can be chosen so that

$$\epsilon(1) \geq \epsilon(2) \geq ... \geq \epsilon(L). \tag{6.1}$$

We shall say that the δ-filtration is *balanced* if the inequalities in (6.1) are all strict and if

$$i_{m_1}^{h_1} \leq i_{m_2}^{h_2} \quad \text{if and only if} \quad m_1 < m_2 \text{ or } m_1 = m_2 \text{ and } h_1 \leq h_2; \tag{6.2}$$

that is, if the usual ordering on $\{1, ..., t\}$ is the same as the Hebrew lexicographic ordering via the pairs (h, m).

REMARK 6.2. If β is as in Proposition 5.1 then the proof of that proposition shows that

$$\frac{1}{\|\beta\|^2} = \sum_{h=1}^{L} \left(\frac{\sum_{l_1(h) \leq i < j \leq l_2(h)} (j-i)^2 r_{h,i} r_{h,j}}{\sum_{l_1(h) \leq i \leq l_2(h)} r_{h,i}} \right) = \sum_{h=1}^{L} \sum_{m=l_1(h)}^{l_2(h)} (m - \epsilon(h))^2 r_{h,m} \tag{6.3}$$

where

$$r_{h,m} = \sum_{j \in \Delta_{h,m}} \|e_j\|^{-2} = \sum_{i=1}^{q} \sum_{j \in \Delta_i^{\phi(h,m)}} \|e_j\|^{-2}$$

$$= \sum_{i=1}^{q} \sum_{j \in \Delta_i^{\phi(h,m)}} p_i = \sum_{i=1}^{q} m_i^{\phi(h,m)} p_i = (1 - g + d/n) \sum_{i=1}^{q} m_i^{\phi(h,m)} n_i.$$

Thus

$$\frac{1}{\|\beta\|^2} = (1 - g + d/n) \sum_{h=1}^{L} \left(\frac{\sum_{l_1(h) \leq i < j \leq l_2(h)} (j-i)^2 \tilde{r}_{h,i} \tilde{r}_{h,j}}{\sum_{l_1(h) \leq i \leq l_2(h)} \tilde{r}_{h,i}} \right) \tag{6.4}$$

$$= (1 - g + d/n) \sum_{h=1}^{L} \sum_{m=l_1(h)}^{l_2(h)} (m - \epsilon(h))^2 \tilde{r}_{h,m}$$

where

$$\tilde{r}_{h,m} = \sum_{i=1}^{q} m_i^{\phi(h,m)} n_i$$

and $\epsilon(h)$ is given by

$$\left(\sum_{m=l_1(h)}^{l_2(h)} m r_{h,m} \right) \left(\sum_{m=l_1(h)}^{l_2(h)} r_{h,m} \right)^{-1} = \left(\sum_{m=l_1(h)}^{l_2(h)} m\tilde{r}_{h,m} \right) \left(\sum_{m=l_1(h)}^{l_2(h)} \tilde{r}_{h,m} \right)^{-1}.$$

Note that if $0 = E_0 \subset E_1 \subset ... \subset E_t = E$ is a filtration such that

$$E_k/E_{k-1} \cong \bigoplus_{i=1}^{q} \mathbb{C}^{m_i^k} \otimes D_i$$

if $1 \leq k \leq t$, where t and m_i^k for $1 \leq i \leq q$ and $1 \leq k \leq t$ are as in Definition 5.5 and $D_1, ..., D_q$ are nonisomorphic stable bundles of ranks $n_1, ..., n_q$ and all of the same slope d/n, then

$$\tilde{r}_{h,m} = \mathrm{rank}(E_{\phi(h,m)}/E_{\phi(h,m)-1}).$$

Proposition 6.3. *Let β be as in Proposition 5.1, let δ be as in Remark 5.4 and let E be a semistable holomorphic structure on $\mathcal{E}$.*

(i) If E represents an element of the stratum $\Sigma_{\beta,l}$ then E has a unique balanced δ-filtration $0 = E_0 \subset E_1 \subset ... \subset E_t = E$ such that

$$E_k/E_{k-1} \cong \bigoplus_{i=1}^q \mathbb{C}^{m_i^k} \otimes D_i$$

and hence

$$E_{\delta(k)}/E_{k-1} \cong \bigoplus_{i=1}^q \mathbb{C}^{m_i^k + ... + m_i^{\delta(k)}} \otimes D_i$$

if $1 \leq k \leq t$, where t and m_i^k for $1 \leq i \leq q$ and $1 \leq k \leq t$ are as in Definition 5.5 and $D_1, ..., D_q$ are nonisomorphic stable bundles of ranks $n_1, ..., n_q$ and all of the same slope d/n.

(ii) Conversely, if E has a balanced δ-filtration $0 = E_0 \subset E_1 \subset ... \subset E_t = E$ as in (i) then E represents an element of the stratum $\Sigma_{\beta,l}$ if and only if E has no filtration with the corresponding properties for any β' satisfying $\|\beta'\| > \|\beta\|$.

Proof: Recall from (2.7) that

$$\Sigma_{\beta,l} = \mathcal{G}_c Y_{\beta,l}^{\setminus E} \cong \mathcal{G}_c \times_{Q_{\beta,l}} Y_{\beta,l}^{\setminus E} \tag{6.5}$$

and that if E represents an element of $Y_{\beta,l}^{\setminus E}$ then its orbit under the complex one-parameter subgroup of R_l generated by β has a limit point in $Z_{R_l}^s$. This limit point is represented by the bundle $\mathrm{gr}(E)$ which is of the form

$$\mathrm{gr}(E) \cong \bigoplus_{i=1}^q \mathbb{C}^{m_i} \otimes D_i$$

where $D_1, ..., D_q$ are nonisomorphic stable bundles of ranks $n_1, ..., n_q$ and all of the same slope d/n. Recall also from [**1**, § 7] that $\mathcal{C}$ is an infinite dimensional affine space, and if we fix a C^∞ identification of the fixed C^∞ Hermitian bundle $\mathcal{E}$ with $\bigoplus_{i=1}^q \mathbb{C}^{m_i} \otimes D_i$ then we can identify $\mathcal{C}$ with the infinite dimensional vector space

$$\Omega^{0,1}(\mathrm{End}(\bigoplus_{i=1}^q \mathbb{C}^{m_i} \otimes D_i))$$

in such a way that the zero element of $\Omega^{0,1}(\mathrm{End}(\bigoplus_{i=1}^q \mathbb{C}^{m_i} \otimes D_i))$ corresponds to the given holomorphic structure on $\bigoplus_{i=1}^q \mathbb{C}^{m_i} \otimes D_i$. With respect to this identification, the action of $R_l = \prod_{i=1}^q GL(m_i;\mathbb{C})$ on $\mathcal{C}$ is the action induced by the obvious action of R_l on $\bigoplus_{i=1}^q \mathbb{C}^{m_i} \otimes D_i$. The one-parameter subgroup of R_l generated by β acts diagonally on $\mathbb{C}^{m_1} \oplus ... \oplus \mathbb{C}^{m_q}$ with weights $\beta.e_j$ for $j \in \{1, ..., M\}$ where $M = m_1 + ... + m_q$, and so it acts on

$$\Omega^{0,1}(\mathrm{End}(\bigoplus_{i=1}^q \mathbb{C}^{m_i} \otimes D_i)) = \bigoplus_{i_1,i_2=1}^q \Omega^{0,1}(\mathbb{C}^{m_{i_1}} \otimes (\mathbb{C}^{m_{i_2}})^* \otimes D_{i_1} \otimes D_{i_2}^*)$$

with weights $\beta.(e_i - e_j)$ for $i, j \in \{1, ..., M\}$. If $E \in \Sigma_\beta$ then E lies in the $\mathcal{G}_c$-orbit of an element of the sum of those weight spaces for which the weight $\beta.(e_i - e_j)$ satisfies $\beta.(e_i - e_j) \geq \|\beta\|^2$. By Remark 5.4 and Definition 5.5 we have a partition $\{\Delta^1, ..., \Delta^t\}$ of $\{1, ..., M\}$ such that $\beta.(e_i - e_j) \geq \|\beta\|^2$ if and only if $i \in \Delta^{k_1}$ and $j \in \Delta^{k_2}$ where $k_2 > \delta(k_1)$. So if we make identifications

$$\mathbb{C}^M = \bigoplus_{i=1}^{q} \mathbb{C}^{m_i} = \bigoplus_{i=1}^{q} \bigoplus_{k=1}^{t} \mathbb{C}^{m_i^k}$$

using the induced partition $\{\Delta_i^k : 1 \leq i \leq q, 1 \leq k \leq t\}$ of $\{1, ..., M\}$ as in Definition 5.5, then any $E \in \Sigma_\beta$ lies in the $\mathcal{G}_c$-orbit of an element of

$$\bigoplus_{i_1, i_2 = 1}^{q} \bigoplus_{k_1 = 1}^{t} \bigoplus_{k_2 = \delta(k_1)+1}^{t} \Omega^{0,1}(\mathbb{C}^{m_{i_1}^{k_1}} \otimes (\mathbb{C}^{m_{i_2}^{k_2}})^* \otimes D_{i_1} \otimes D_{i_2}^*).$$

This completes the proof of (i), as such an element of $\Omega^{0,1}(\mathrm{End}(\bigoplus_{i=1}^q \mathbb{C}^{m_i} \otimes D_i))$ represents a holomorphic structure E on $\mathcal{E}$ with a filtration of the required form (uniqueness follows from (6.5) and the fact that $Q_{\beta,l}$ preserves the filtration), and (ii) is now a consequence of (2.1).

Corollary 6.4. *Let β be as in Proposition 5.1, let δ be as in Remark 5.4 and let E be a semistable holomorphic structure on $\mathcal{E}$ with a balanced δ-filtration*

$$0 = E_0 \subset E_1 \subset ... \subset E_t = E \tag{6.6}$$

whose subquotients satisfy

$$E_k / E_{k-1} \cong \bigoplus_{i=1}^{q} \mathbb{C}^{m_i^k} \otimes D_i,$$

where t and m_i^k for $1 \leq i \leq q$ and $1 \leq k \leq t$ are as in Definition 5.5 and $D_1, ..., D_q$ are nonisomorphic stable bundles of ranks $n_1, ..., n_q$ and all of the same slope d/n. Then E represents an element of the stratum $\Sigma_{\beta,l}$ if and only if, in the notation of Definition 6.1, there is no $h \in \{1, ..., L\}$ and refinement

$$0 = E_0 \subset ... \subset E_{i^h_{l_1(h)}-1} \subset F_{l_1(h)} \subset E_{i^h_{l_1(h)}} \subset ... \subset E_{i^h_m - 1} \subset F_m \subset E_{i^h_m} \subset ...$$
$$... \subset E_{i^h_{l_2(h)}-1} \subset F_{l_2(h)} \subset E_{i^h_{l_2(h)}} \subset ... \subset E_t = E$$

of (6.6) with

$$\frac{\sum_{m=l_1(h)}^{l_2(h)} m \ \mathrm{rank}(F_m / E_{i^h_m - 1})}{\sum_{m=l_1(h)}^{l_2(h)} \ \mathrm{rank}(F_m / E_{i^h_m - 1})} < \frac{\sum_{m=l_1(h)}^{l_2(h)} m \ \mathrm{rank}(E_{i^h_m} / E_{i^h_m - 1})}{\sum_{m=l_1(h)}^{l_2(h)} \ \mathrm{rank}(E_{i^h_m} / E_{i^h_m - 1})} = \epsilon(h),$$

such that the induced filtrations

$$0 \subset \frac{E_{i^h_m - 1}}{E_{i^{h_1}_{m_1} - 1}} \subseteq \frac{F_m}{E_{i^{h_1}_{m_1} - 1}}$$

with

$$m_1 - \epsilon(h_1) \leq \frac{\sum_{m=l_1(h)}^{l_2(h)} m \mathrm{rank}(F_m / E_{i^h_m - 1})}{\sum_{m=l_1(h)}^{l_2(h)} \mathrm{rank}(F_m / E_{i^h_m - 1})}$$

and

$$0 \subset \frac{E_{i^h_m}}{F_m} \subseteq \frac{E_{i^{h_2}_{m_2}}}{F_m}$$

with

$$m_2 - \epsilon(h_2) \geq m - \frac{\sum_{m=l_1(h)}^{l_2(h)} m \ \mathrm{rank}(F_m/E_{i_m^h - 1})}{\sum_{m=l_1(h)}^{l_2(h)} \ \mathrm{rank}(F_m/E_{i_m^h - 1})}$$

are all trivial.

Remark 6.5. If $0 = E_0 \subset E_1 \subset ... \subset E_{j-1} \subset F \subset E_j \subset ... \subset E_t = E$ is a refinement of the filtration (6.6) of E such that the induced filtration

$$0 \subseteq \frac{E_{j-1}}{E_i} \subset \frac{F}{E_i}$$

is trivial for some $i < j-1$, then F/E_i is isomorphic to

$$\frac{E_{j-1}}{E_i} \oplus \frac{F}{E_{j-1}}$$

and so E can be given a filtration of the form

$$0 = E_0 \subset E_1 \subset ... \subset E_i \subset E'_{i+1} \subset E'_{i+2} \subset ... \subset E'_{j-1} \subset F \subset E_j \subset ... \subset E_t = E$$

where $E'_{i+1}/E_i \cong F/E_{j-1}$, $E'_{k+1}/E'_k \cong E_k/E_{k-1}$ for $i < k < j$ and $F/E'_{j-1} \cong E_{j-1}/E_{j-2}$. A similar result is true if the induced filtration

$$0 \subset \frac{F}{E_{j-1}} \subset \frac{E_i}{E_{j-1}}$$

is trivial for some $i > j$.

Proof of Corollary 6.4: This follows from [**28**] and the proof of Proposition 5.1, which tells us that if $\beta' \neq \beta$ is the closest point to 0 of the convex hull of $\{e_i - e_j : (i,j) \in S'\}$ where S' is a subset of S, then S' can be chosen so that the connected components of the graph $G(S')$ give us a refinement of the partition $\{\Delta_{h,m} : (h,m) \in J\}$ of $\{1, ..., M\}$ associated to β, which in turn gives us a refinement of the filtration (6.6) with the required properties.

Remark 6.6. Recall from Proposition 6.3 that a semistable bundle E represents an element of the stratum $\Sigma_{\beta,l}$ if and only if if has a balanced δ-filtration

$$0 = E_0 \subset E_1 \subset ... \subset E_t = E \tag{6.7}$$

such that if $1 \leq k \leq t$ then E_k/E_{k-1} is of the form $\bigoplus_{i=1}^q \mathbb{C}^{m_i^k} \otimes D_i$ where $D_1, ..., D_q$ are nonisomorphic stable bundles of ranks $n_1, ..., n_q$ and all of the same slope d/n, and moreover E has no balanced δ-filtration with the corresponding properties when β is replaced with β' satisfying $\|\beta'\| > \|\beta\|$. From (6.4) we have

$$\frac{1}{\|\beta\|^2} = (1 - g + d/n) \sum_{h=1}^{L} \left(\frac{\sum_{l_1(h) \leq i < j \leq l_2(h)} (j-i)^2 \tilde{r}_{h,i} \tilde{r}_{h,j}}{\sum_{l_1(h) \leq i \leq l_2(h)} \tilde{r}_{h,i}} \right), \tag{6.8}$$

where

$$\tilde{r}_{h,m} = \mathrm{rank}(E_{\phi(h,m)}/E_{\phi(h,m)-1}).$$

This gives us some sort of measure of the triviality of the balanced δ-filtration (6.7); very roughly speaking, the more trivial this filtration, the smaller the size of $\|\beta\|^{-2}$ and hence the larger $\|\beta\|$ becomes.

Let us therefore define the *triviality* of the balanced δ-filtration (6.7) with associated function δ to be

$$\left(\sum_{h=1}^{L}\left(\frac{\sum_{l_1(h)\leq i<j\leq l_2(h)}(j-i)^2\tilde{r}_{h,i}\tilde{r}_{h,j}}{\sum_{l_1(h)\leq i\leq l_2(h)}\tilde{r}_{h,i}}\right)\right)^{-1/2}$$

$$= \left(\sum_{h=1}^{L}\sum_{m=l_1(h)}^{l_2(h)}(m-\epsilon(h))^2\tilde{r}_{h,m}\right)^{-1/2}; \tag{6.9}$$

Remarks 5.2 and 5.6 tell us that this is well defined. Thus if $E \in \Sigma_{\beta,l}$ the balanced δ-filtration (6.7) associated to E by Proposition 10.1 can be thought of as having maximal triviality (according to this measure) among the balanced δ-filtrations of E.

Remark 6.7. Let β be as in Proposition 5.1, let δ be as in Remark 5.4 and let E be a semistable holomorphic structure on $\mathcal{E}$ with a δ-filtration

$$0 = E_0 \subset E_1 \subset ... \subset E_t = E$$

such that if $1 \leq k \leq t$ then $E_k/E_{k-1} \cong \bigoplus_{i=1}^q \mathbb{C}^{m_i^k} \otimes D_i$ where $D_1, ..., D_q$ are nonisomorphic stable bundles of ranks $n_1, ..., n_q$ and all of the same slope d/n. Then the proof of Proposition 6.3 and (2.3) shows that E represents an element of $\Sigma_{\tilde{\gamma}}$ for some $\tilde{\gamma} \in \Gamma$ satisfying $\tilde{\gamma} \geq \gamma = (\beta, l)$ with respect to the partial ordering on Γ defined just before (2.3).

Remark 6.8. If $[\mathbf{n},\mathbf{m}] \in \mathcal{I}^{ss}$ then we can define $\beta[\mathbf{n},\mathbf{m}] \in \Gamma$ as follows. Let $m_i = \sum_{j=1}^r m_{ij}$ and suppose that R_l is the subgroup of $GL(p;\mathbb{C})$ given by the embedding of $\prod_{i=1}^q GL(m_i;\mathbb{C})$ via a fixed identification of $\bigoplus_{i=1}^q \mathbb{C}^{n_i(1-g+d/n)} \otimes \mathbb{C}^{m_i}$ with $\mathbb{C}^p$. If $q = 1$ let

$$\beta[\mathbf{n},\mathbf{m}] = R_l,$$

and if $q > 1$ let $\beta[\mathbf{n},\mathbf{m}] \in \mathcal{B}_l \backslash \{0\}$ be represented by the closest point to 0 of the convex hull of the weights of the representation of the maximal compact torus T_l of R_l on

$$\bigoplus_{i_1,i_2=1}^{q}\bigoplus_{1\leq j_1<j_2\leq r}\mathbb{C}^{m_{i_1j_1}} \otimes (\mathbb{C}^{m_{i_2j_2}})^*$$

given by identifying $\bigoplus_{j=1}^r \mathbb{C}^{m_{ij}}$ with $\mathbb{C}^{m_i}$ for $1 \leq i \leq q$. Then we have from Remark 6.7 that

$$S_{[\mathbf{n},\mathbf{m}]} \subseteq \bigcup_{\gamma \geq \beta[\mathbf{n},\mathbf{m}]} \Sigma_\gamma, \tag{6.10}$$

and from Remark 4.4 and Proposition 6.3 that

$$\Sigma_{\beta[\mathbf{n},\mathbf{m}]} \subseteq \bigcup_{[\mathbf{n}',\mathbf{m}'] \geq [\mathbf{n},\mathbf{m}]} S_{[\mathbf{n}',\mathbf{m}']} \tag{6.11}$$

where $\geq$ denotes in (6.10) the partial order on Γ used in Remark 6.7, whereas in (6.11) it denotes the partial order on $\mathcal{I}^{ss}$ described in Remark 4.4.

REMARK 6.9. It follows from Proposition 6.3 that if R_l is as at (5.1) then a holomorphic structure belongs to

$$\bigcup_{\beta \in \mathcal{B}_l \setminus \{0\}} \Sigma_{\beta,l}$$

if and only if $E \not\cong \mathrm{gr}(E) \cong (\mathbb{C}^{m_1} \otimes D_1) \oplus \cdots \oplus (\mathbb{C}^{m_q} \otimes D_q)$ where $D_1, \ldots, D_q$ are all stable of slope d/n and ranks $n_1, \ldots, n_q$ and are not isomorphic to one another.

7. Pivotal filtrations

Let us now consider the relationship between the balanced δ-filtration associated to a semistable bundle E as in Proposition 6.3 and the maximal and minimal Jordan–Hölder filtrations defined in §8.

Indeed, motivated by Proposition 6.3, we can try to carry our analysis of the maximal Jordan–Hölder filtration

$$0 = E_0 \subset E_1 \subset ... \subset E_t = E \tag{7.1}$$

of a bundle E a bit further. Recall that if $1 \leq j \leq t$ then the subquotient E_j/E_{j-1} is the maximal subbundle of E/E_{j-1} which is a direct sum of stable bundles all having maximal slope among the nonzero subbundles of E/E_{j-1}. We can ask whether it is true for every subbundle F of E satisfying $E_{j-1} \subset F \subset E_j$ and $\mathrm{slope}(E_j/F) = \mathrm{slope}(E_j/E_{j-1})$ that E_j/F is the maximal subbundle of E/F which is a direct sum of stable bundles all having maximal slope among the nonzero subbundles of E/F. Of course if E_j/E_{j-1} is itself stable there are no such intermediate subbundles F, so this is trivially true, but it is not always the case (as Example 7.1 below shows).

If there does exist such an intermediate subbundle F, then by Lemma 3.2 both F/E_{j-1} and E_j/F are of the form

$$(\mathbb{C}^{m_1} \otimes D_1) \oplus \cdots \oplus (\mathbb{C}^{m_q} \otimes D_q) \tag{7.2}$$

where $D_1, \ldots, D_q$ are all stable of slope d/n and ranks $n_1, \ldots, n_q$ and are not isomorphic to one another. So we can then ask whether it is possible to find a canonical refinement

$$0 = F_0 \subset F_1 \subset ... \subset F_u = E \tag{7.3}$$

of the maximal Jordan–Hölder filtration (7.1) of E with an increasing function $\delta : \{1, ..., u\} \to \{1, ..., u\}$ such that $\delta(j) \geq j$ if $1 \leq j \leq u$, each subquotient $F_{\delta(j)}/F_{j-1}$ is of the form (7.2) and moreover for every subbundle F of E satisfying $F_{j-1} \subseteq F \subset F_j$ and $\mathrm{slope}(F_j/F) = \mathrm{slope}(F/F_{j-1})$, the quotient $F_{\delta(j)}/F$ is the maximal subbundle of E/F which is a direct sum of stable bundles all having maximal slope among the nonzero subbundles of E/F. The following example shows that even this cannot be achieved in a canonical way.

EXAMPLE 7.1. Let E_1 and E_2 be semistable bundles over Σ such that $\mathrm{slope}(E_1)$ equals $\mathrm{slope}(E_2)$, with maximal Jordan–Hölder filtrations

$$0 \subset D_1 \subset E_1 \quad \text{and} \quad 0 \subset D_2 \subset E_2$$

where D_1, D_2, E_1/D_1 and E_2/D_2 are nonisomorphic stable bundles all of the same slope as E_1 and E_2, and let

$$E = E_1 \oplus E_2.$$

We observed in Remark 4.7 that the maximal Jordan–Hölder filtration of a direct sum of semistable bundles of the same slope is the direct sum of their maximal

Jordan–Hölder filtrations (with the shorter one extended trivially at the top if they are not of the same length). Thus the maximal Jordan–Hölder filtration of E is

$$0 \subset D_1 \oplus D_2 \subset E_1 \oplus E_2 = E. \tag{7.4}$$

By Lemma 3.2 there are precisely two proper subbundles F of $D_1 \oplus D_2$ with $\mathrm{slope}(F) = \mathrm{slope}(D_1 \oplus D_2/F) = \mathrm{slope}(D_1 \oplus D_2)$, namely D_1 and D_2. The maximal Jordan–Hölder filtration of $E/D_1 = (E_1/D_1) \oplus E_2$ is

$$0 \subset (E_1/D_1) \oplus D_2 \subset (E_1/D_1) \oplus E_2,$$

so we can refine the filtration (7.5) of E to get

$$0 = F_0 \subset F_1 \subset F_2 \subset F_3 \subset F_4 = E \tag{7.5}$$

where $F_1 = D_1$, $F_2 = D_1 \oplus D_2$ and $F_3 = E_1 \oplus D_2$, and we can define $\delta : \{1,2,3,4\} \to \{1,2,3,4\}$ by $\delta(1) = 2$, $\delta(2) = 3$ and $\delta(3) = \delta(4) = 4$. If $1 \le j \le 4$ then $F_{\delta(j)}/F_{j-1}$ is the maximal subbundle of E/F_{j-1} which is a direct sum of stable subbundles all having maximal slope among the nonzero subbundles of E/F_{j-1}. Moreover this is trivially still true if we replace F_{j-1} by any subbundle F of E satisfying $F_{j-1} \subseteq F \subset F_j$ and $\mathrm{slope}(F_j/F) = \mathrm{slope}(F_j/F_{j-1})$, since the only such F is F_{j-1} itself. We can of course reverse the rôles of E_1 and E_2 in this construction, to get another refinement

$$0 \subset D_2 \subset D_1 \oplus D_2 \subset D_1 \oplus E_2 \subset E_1 \oplus E_2 = E \tag{7.6}$$

of (7.4). Thus there are precisely two refinements of the maximal Jordan–Hölder filtration of $E_1 \oplus E_2$ with the required propertes, and if E_1 has the same rank as E_2 and D_1 has the same rank as D_2 then by symmetry there can be no canonical choice.

Notice that if $\mathrm{rank}(D_1)/\mathrm{rank}(E_1) = \mathrm{rank}(D_2)/\mathrm{rank}(E_2)$ then neither of the δ-filtrations (7.5) and (7.6) is balanced since the inequalities (6.1) are not strict; however the δ-filtration (7.4) is balanced and has maximal triviality, in the sense of (6.9), among balanced δ-filtrations of $E_1 \oplus E_2$. If on the other hand $\mathrm{rank}(D_1)/\mathrm{rank}(E_1) \neq \mathrm{rank}(D_2)/\mathrm{rank}(E_2)$ then precisely one of the δ-filtrations (7.5) and (7.6) is balanced and it has maximal triviality, in the sense of (6.9), among balanced δ-filtrations of $E_1 \oplus E_2$. This filtration then determines the stratum Σ_γ to which E belongs, and in this case E represents an element of the open subset $\Sigma^s_{\beta,l}$ of Σ_γ.

Lemma 7.2. *Let E be a bundle over Σ with a filtration*

$$0 = F_0 \subset F_1 \subset ... \subset F_u = E$$

and let $\{\Delta_1, ..., \Delta_L\}$ be a partition of $\{1, ..., u\}$ such that if $1 \le h \le L$ and $\Delta_h = \{i^h_1, ..., i^h_{s_h}\}$ where $i^h_1 < i^h_2 < ... < i^h_{s_h}$, then the induced extension

$$0 \to F_{i^h_j}/F_{i^h_j - 1} \to F_{i^h_{j+1}-1}/F_{i^h_j - 1} \to F_{i^h_{j+1}-1}/F_{i^h_j} \to 0 \tag{7.7}$$

is trivial. Then we can associate to this filtration of E, partition $\{\Delta_1, ..., \Delta_L\}$ of $\{1, ..., u\}$ and trivialisations of the induced extensions (7.7) a sequence of elements of

$$H^1(\Sigma, (F_{i^h_j}/F_{i^h_j - 1}) \otimes (F_{i^h_{j+1}}/F_{i^h_{j+1}-1})^*)$$

or equivalently of extensions

$$0 \to F_{i^h_j}/F_{i^h_j - 1} \to E^h_j \to F_{i^h_{j+1}}/F_{i^h_{j+1}-1} \to 0$$

for $1 \le h \le L$ and $1 \le j \le s_h - 1$.

Proof: This lemma follows immediately from the well known bijective correspondence between holomorphic extensions of a holomorphic bundle D_1 over Σ by another holomorphic bundle D_2 and elements of $H^1(\Sigma, D_1 \otimes D_2)$. The extension

$$0 \to F_{i^h_{j+1}-1}/F_{i^h_j-1} \to F_{i^h_{j+1}}/F_{i^h_j-1} \to F_{i^h_{j+1}}/F_{i^h_{j+1}-1} \to 0$$

induced by the given filtration gives us an element of

$$H^1(\Sigma, (F_{i^h_{j+1}-1}/F_{i^h_j-1}) \otimes (F_{i^h_{j+1}}/F_{i^h_{j+1}-1})^*)$$

and the given trivialisation of the extension (7.7) gives us a decomposition of this as

$$H^1(\Sigma, (F_{i^h_j}/F_{i^h_j-1}) \otimes (F_{i^h_{j+1}}/F_{i^h_{j+1}-1})^*) \oplus H^1(\Sigma, (F_{i^h_{j+1}-1}/F_{i^h_j}) \otimes (F_{i^h_{j+1}}/F_{i^h_{j+1}-1})^*).$$

Projection onto the first summand gives us an extension

$$0 \to F_{i^h_j}/F_{i^h_j-1} \to E^h_j \to F_{i^h_{j+1}}/F_{i^h_{j+1}-1} \to 0$$

as required.

Remark 7.3. Lemma 5.3 and Remark 5.4 tell us that if $i \in \Delta^{k_1}_{i_1}$ and $j \in \Delta^{k_2}_{i_2}$ where $k_2 > \delta(k_1)$ then $(e_i - e_j).\beta \geq \|\beta\|^2$, and equality occurs if and only if there exists $(h,m) \in J$ with $(h, m+1) \in J$ such that $k_1 = \phi(h,m)$ and $k_2 = \phi(h, m+1)$. Thus, retaining the notation of the proof of Proposition 6.3, we observe that if E is represented by an element of

$$\bigoplus_{i_1,i_2=1}^{q} \bigoplus_{k_1=1}^{t} \bigoplus_{k_2=\delta(k_1)+1}^{t} \Omega^{0,1}(\mathbb{C}^{m^{k_1}_{i_1}} \otimes (\mathbb{C}^{m^{k_2}_{i_2}})^* \otimes D_{i_1} \otimes D^*_{i_2})$$

then the limit in $\mathcal{C}$ as $t \to \infty$ of $\exp(-it\beta)E$ is the bundle

$$\mathrm{gr}(E) \cong \bigoplus_{i=1}^{q} \mathbb{C}^{m_i} \otimes D_i$$

which is represented by the zero vector in

$$\bigoplus_{i_1,i_2=1}^{q} \bigoplus_{k_1=1}^{t} \bigoplus_{k_2=\delta(k_1)+1}^{t} \Omega^{0,1}(\mathbb{C}^{m^{k_1}_{i_1}} \otimes (\mathbb{C}^{m^{k_2}_{i_2}})^* \otimes D_{i_1} \otimes D^*_{i_2}),$$

and so the limit $p_\beta(E)$ of $\exp(-it\beta)E$ in the blow-up of $\mathcal{C}$ along $\mathcal{G}_c Z^{ss}_{R_l}$ is an element of the fibre

$$\mathbb{P}(\mathcal{N}_{l,\mathrm{gr}E}) = \mathbb{P}(H^1(\Sigma, \bigoplus_{i_1,i_2=1}^{q} \mathbb{C}^{m_{i_1} m_{i_2} - \delta^{i_2}_{i_1}} \otimes D_{i_1} \otimes D^*_{i_2})$$

of the exceptional divisor over $\mathrm{gr}\, E$. Indeed $p_\beta(E)$ is the element of this fibre represented by the sum in

$$\bigoplus_{h=1}^{L} \bigoplus_{m=l_1(h)}^{l_2(h)-1} H^1(\Sigma, (\bigoplus_{i_1}^{q} \mathbb{C}^{m^{\phi(h,m)}_{i_1}} \otimes D_{i_1}) \otimes (\bigoplus_{i_2}^{q} \mathbb{C}^{m^{\phi(h,m+1)}_{i_2}} \otimes D_{i_2})^*)$$

of the elements of

$$H^1(\Sigma, (\bigoplus_{i_1}^{q} \mathbb{C}^{m^{\phi(h,m)}_{i_1}} \otimes D_{i_1}) \otimes (\bigoplus_{i_2}^{q} \mathbb{C}^{m^{\phi(h,m+1)}_{i_2}} \otimes D_{i_2})^*)$$

for $1 \leq h \leq L$ and $l_1(h) \leq m \leq l_2(h)$ corresponding to the extensions

$$0 \to \bigoplus_{i_1}^{q} \mathbb{C}^{m_{i_1}^{\phi(h,m)}} \otimes D_{i_1} \to E_m^h \to \bigoplus_{i_2}^{q} \mathbb{C}^{m_{i_2}^{\phi(h,m+1)}} \otimes D_{i_2} \to 0$$

associated as in Lemma 7.2 to the δ-filtration $0 = E_0 \subset E_1 \subset ... \subset E_t = E$ of Proposition 6.3.

Proposition 7.4. *Let β be as in Proposition 5.1 and let E be a semistable bundle representing an element of $\Sigma_{\beta,l}$ with a balanced δ-filtration*

$$0 = E_0 \subset E_1 \subset ... \subset E_t = E$$

such that if $1 \leq k \leq t$ then $E_k/E_{k-1} \cong \bigoplus_{i=1}^q \mathbb{C}^{m_i^k} \otimes D_i$ as in Proposition 6.3. Suppose also that $k_1 \in \{1, ..., t\}$ is such that

$$\beta.e_j < 0 \text{ whenever } j \in \Delta^{k_2} \text{ with } k_2 > \delta(k_1).$$

Then whenever F is a subbundle of E with slope(F) = slope(E) and such that $E_{k_1-1} \subseteq F \subset E_{k_1}$, the subquotient $E_{\delta(k_1)}/F$ is the maximal subbundle of E/F which is a direct sum of stable bundles all having the same slope as E/F.

Proof: Since F/E_{k_1-1} is a subbundle of $E_{k_1}/E_{k_1-1} \cong \bigoplus_{i=1}^q \mathbb{C}^{m_i^{k_1}} \otimes D_i$ having the same slope as $D_1, ..., D_q$, it follows from Lemma 3.2 that

$$F/E_{k_1-1} = \bigoplus_{i=1}^{q} U_i \otimes D_i$$

where U_i is a linear subspace of $\mathbb{C}^{m_i^{k_1}}$ for $1 \leq i \leq q$, and so

$$E_{\delta(k_1)}/F \cong \bigoplus_{i=1}^{q} ((\mathbb{C}^{m_i^{k_1}}/U_i) \oplus \mathbb{C}^{m_i^{k_1+1} + ... + m_i^{\delta(k_1)}}) \otimes D_i$$

is a sum of stable bundles all having the same slope (which is equal to slope(E) and slope(E/F)). Let us suppose for a contradiction that E/F has a subbundle E'/F which is not contained in $E_{\delta(k_1)}/F$ and which is of the required form. Then we can choose $k_2 > \delta(k_1)$ such that $E' \subseteq E_{k_2}$ but E' is not contained in E_{k_2-1}, and then the inclusion of E' in E_{k_2} induces a nonzero map

$$\theta : E'/F \to E_{k_2}/E_{k_2-1} \cong \bigoplus_{i=1}^{q} \mathbb{C}^{m_i^{k_2}} \otimes D_i.$$

Since nonzero bundle maps between stable bundles of the same slope are always isomorphisms, by replacing E' by a suitable subbundle we can assume that $E'/F \cong D_{i_0}$ for some $i_0 \in \{1, ..., q\}$, and that we can decompose $\mathbb{C}^{m_{i_0}^{k_2}}$ as $\mathbb{C} \oplus \mathbb{C}^{m_{i_0}^{k_2}-1}$ in such a way that the projection $\theta_0 : E'/F \to D_{i_0}$ of θ onto the corresponding component D_{i_0} of E_{k_2}/E_{k_2-1} is an isomorphism. Then

$$\theta_0^{-1} : D_{i_0} \to E'/F \subseteq E_{k_2}/F$$

gives us a trivialisation of the extension of E_{k_2-1}/E_{k_1} by this component D_{i_0} of E_{k_2}/E_{k_2-1}. By the definition of $\Sigma_{\beta,l}$ the limit $p_\beta(E) \in \mathbb{P}(\mathcal{N}_{l,\mathrm{gr}E})$ of $\exp(-it\beta)E$ as $t \to \infty$ is semistable for the induced action of $\mathrm{Stab}(\beta)/T_\beta^c$ where $\mathrm{Stab}(\beta)$ is the stabiliser of β under the coadjoint action of R_l and T_β^c is the complex subtorus generated by β (see [**28**]), and by Remark 7.3 $p_\beta(E)$ is represented by the sum over $h \in \{1, ..., L\}$ and $m \in \{l_1(h), ..., l_2(h)\}$ of the elements of

$$H^1(\Sigma, (\bigoplus_{i_1}^{q} \mathbb{C}^{m_{i_1}^{\phi(h,m)}} \otimes D_{i_1}) \otimes (\bigoplus_{i_2}^{q} \mathbb{C}^{m_{i_2}^{\phi(h,m+1)}} \otimes D_{i_2})^*)$$

corresponding to the extensions

$$0 \to \bigoplus_{i_1}^{q} \mathbb{C}^{m_{i_1}^{\phi(h,m)}} \otimes D_{i_1} \to E_m^h \to \bigoplus_{i_2}^{q} \mathbb{C}^{m_{i_2}^{\phi(h,m+1)}} \otimes D_{i_2} \to 0$$

associated by Lemma 7.2 to the filtration $0 = E_0 \subset E_1 \subset ... \subset E_t = E$.

Let S_0 be the set of ordered pairs (i, j) with $i, j \in \{1, ..., M\}$ such that the component of $p_\beta(E)$ in the weight space corresponding to the weight $e_i - e_j$ for the action of the maximal torus T_l of $R_l = \prod_{i=1}^q GL(m_i; \mathbb{C})$ on

$$\bigoplus_{h=1}^{L} \bigoplus_{m=l_1(h)}^{l_2(h)-1} H^1(\Sigma, (\bigoplus_{i_1}^{q} \mathbb{C}^{m_{i_1}^{\phi(h,m)}} \otimes D_{i_1}) \otimes (\bigoplus_{i_2}^{q} \mathbb{C}^{m_{i_2}^{\phi(h,m+1)}} \otimes D_{i_2})^*)$$

is nonzero. Since $p_\beta(E)$ is semistable for the action of $\mathrm{Stab}(\beta)/T_\beta^c$, it follows that β is the closest point to 0 in the convex hull of $\{e_i - e_j : (i, j) \in S_0\}$. We may assume that T_l acts diagonally with respect to the decomposition of $\mathbb{C}^{m_{i_0}^{k_2}}$ as $\mathbb{C} \oplus \mathbb{C}^{m_{i_0}^{k_2}-1}$; let e_{j_0} be the weight of the action of T_l on the component $\mathbb{C}$ of $\mathbb{C}^{m_{i_0}^{k_2}}$ with respect to this decomposition. Since $k_2 > \delta(k_1)$ and the extension of E_{k_2-1}/E_{k_1} by the component of $E_{k_2}/E_{k_2-1} \cong \bigoplus_{i=1}^q \mathbb{C}^{m_i^{k_2}} \otimes D_i$ corresponding to the weight e_{j_0} is trivial, it follows that if $(i, j) \in S_0$ then $j \neq j_0$. Since β lies in the convex hull of $\{e_i - e_j : (i, j) \in S_0\}$ and $e_1, ..., e_M$ are mutually orthogonal, this means that

$$\beta . e_{j_0} \geq 0,$$

and as $j_0 \in \Delta^{k_2}$ and $k_2 > \delta(k_1)$, this gives us the required contradiction.

Remark 7.5. Dual to the definition of δ in Remark 5.4, we can define an increasing function $\delta' : \{1, ..., t\} \to \{1, ..., t\}$ such that $\delta'(\phi(h, m)) - 1$ is the number of elements $(h', m') \in J$ such that either $m' < m - 1$ or $m' = m - 1$ and $h' \leq h$. Then $\delta'(k) \leq k$ for all $k \in \{1, ..., t\}$, and if (h, m) and $(h, m - 1)$ both belong to J then $\delta'(\phi(h, m)) = \phi(h, m - 1) + 1$. Also $k_1 < \delta'(k_2)$ if and only if $k_2 > \delta(k_1)$, and Lemma 5.3 tells us that if $i \in \Delta^{k_1}$ and $j \in \Delta^{k_2}$ then $\beta.(e_i - e_j) \geq \|\beta\|^2$ if and only if $k_1 < \delta'(k_2)$. The dual version of Proposition 6.3 tells us that if $1 \leq k \leq t$ then $E_k/E_{\delta'(k)-1}$ is a direct sum of stable bundles all of the same slope, and using Remark 4.7 we obtain the following dual version of Proposition 7.4.

Proposition 7.6. *Let β be as in Proposition 5.1 and let E be a semistable bundle representing an element of $\Sigma_{\beta,l}$ with δ-filtration*

$$0 = E_0 \subset E_1 \subset ... \subset E_t = E$$

such that if $1 \leq k \leq t$ then $E_k/E_{k-1} \cong \bigoplus_{i=1}^q \mathbb{C}^{m_i^k} \otimes D_i$ as in Proposition 6.3. Suppose also that $k_1 \in \{1, ..., t\}$ is such that

$$\beta . e_j > 0 \text{ whenever } j \in \Delta^{k_2} \text{ with } k_2 < \delta'(k_1).$$

Then whenever F is a subbundle of E with slope(F) = slope(E) and such that $E_{k_1-1} \subset F \subseteq E_{k_1}$, the subbundle $E_{\delta'(k_1)-1}$ is the minimal subbundle of F such that $F/E_{\delta'(k_1)-1}$ is a direct sum of stable bundles all with the same slope as F.

Remark 7.7. It follows from the definition of $\Delta^1, ..., \Delta^t$ (Definition 5.5) that if $j_1 \in \Delta^{k_1}$ and $j_2 \in \Delta^{k_2}$ then $\beta.e_{j_1} < \beta.e_{j_2}$ if and only if $k_1 > k_2$, so we can choose k_- and k_+ such that $\beta.e_j < 0$ (respectively $\beta.e_j > 0$) if and only if $j \in \Delta^k$ with $k > k_-$ (respectively $k < k_+$). Then $k_- = k_+$ or $k_- = k_+ - 1$, depending on whether there exists j with $\beta.e_j = 0$. Propositions 7.4 and 7.6 tell us that if E is a semistable bundle representing an element of Σ_β with filtration

$$0 = E_0 \subset E_1 \subset ... \subset E_t = E$$

as in Proposition 6.3, then

$$0 \subset E_{\delta(j)+1}/E_{j-1} \subset E_{\delta(\delta(j)+1)+1}/E_{j-1} \subset ... \subset E/E_{j-1}$$

is the maximal Jordan–Hölder filtration of E/E_{j-1} if $j \geq \delta'(k_-)$, and

$$0 \subset ... \subset E_{\delta'(\delta'(j)-1)-1} \subset E_{\delta'(j)-1} \subset E_j$$

is the minimal Jordan–Hölder filtration of E_j if $j \leq \delta(k_+)$. Note also that $\delta'(k_-) \leq k_- \leq k_+ \leq \delta(k_+)$, so there are values of j satisfying both $j \geq \delta'(k_-)$ and $j \leq \delta(k_+)$.

There is a converse to Propositions 7.4 and 7.6.

Proposition 7.8. *Let β be as in Proposition 5.1 and let E be a semistable bundle with δ-filtration*

$$0 = E_0 \subset E_1 \subset ... \subset E_t = E$$

such that if $1 \leq k \leq t$ then $E_k/E_{k-1} \cong \bigoplus_{i=1}^q \mathbb{C}^{m_i^k} \otimes D_i$ as in Proposition 6.3. Suppose that every subbundle F of E with the same slope as E satisfies the following two properties:

(i) if $E_{k_1-1} \subseteq F \subset E_{k_1}$ for some $k_1 \in \{1, ..., t\}$ such that $\beta.e_j < 0$ whenever $j \in \Delta^{k_2}$ with $k_2 > \delta(k_1)$, then the subquotient $E_{\delta(k_1)/F}$ is the maximal subbundle of E/F which is a direct sum of stable bundles all with the same slope as E/F;

(ii) if $E_{k_1-1} \subset F \subseteq E_{k_1}$ for some $k_1 \in \{1, ..., t\}$ such that $\beta.e_j > 0$ whenever $j \in \Delta^{k_2}$ with $k_2 < \delta'(k_1)$, then the subbundle $E_{\delta'(k_1)-1}$ is the minimal subbundle of F such that $F/E_{\delta'(k_1)-1}$ is a direct sum of stable bundles all with the same slope as F.

Then E represents an element of the stratum $\Sigma_{\beta,l}$.

Proof: Suppose for a contradiction that E does not represent an element of $\Sigma_{\beta,l}$. Then (cf. [**28**]) after applying a change of coordinates to $\mathbb{C}^{m_i^k}$ for $1 \leq i \leq q$ and $1 \leq k \leq t$, we can assume that β is not equal to the closest point to 0 in the convex hull of $\{e_i - e_j : (i,j) \in S_0\}$ where S_0 is as in the proof of Proposition 8.4. Moreover

$$S_0 \subseteq \{(i,j) : \beta.(e_i - e_j) \geq \|\beta\|^2\}.$$

Thus β does not lie in the convex hull of

$$\{e_i - e_j : (i,j) \in S_0 \cap S\}$$

where $S = \{(i,j) : \beta.(e_i - e_j) = \|\beta\|^2\}$. ¿From Lemma 5.3 and Remark 5.4 we know that if $i \in \Delta_{h,m}$ and $j \in \Delta_{h',m'}$ then $\beta.(e_i - e_j) \geq \|\beta\|^2$ if and only if $m' \geq m+2$ or $m' = m+1$ and $h' \geq h$, and this happens if and only if $\phi(h', m') > \delta(\phi(h,m))$, while $\beta.(e_i - e_j) = \|\beta\|^2$ if and only if $m' = m+1$ and $h' = h$. By Remark 7.7 the hypothesis (i) on subbundles F of E tells us that if $E_{k_1-1} \subseteq F \subset E_{k_1}$ where $k_1 \geq \delta(k_-)$, then the subquotient $E_{\delta(k_1)/F}$ is the maximal subbundle of E/F which is a direct sum of stable bundles all with the same slope as E/F. This implies that if (h,m) and $(h,m+1)$ both lie in J and $\phi(h,m) \geq k_-$ then every pair (i,j) with $i \in \Delta_{h,m}$ and $j \in \Delta_{h,m+1}$ lies in S_0. Similarly the hypothesis (ii) tells us that if (h,m) and $(h,m-1)$ both lie in J and $\phi(h,m) \leq k_+$ then every pair (i,j) with $i \in \Delta_{h,m}$ and $j \in \Delta_{h,m-1}$ lies in S_0. Since $k_- \leq k_+$ this means that

$$S_0 \cap S = S.$$

This contradicts the fact that β is the closest point to 0 in the convex hull of $\{e_i - e_j : (i,j) \in S\}$ but does not lie in the convex hull of $\{e_i - e_j : (i,j) \in S_0 \cap S\}$, and thus completes the proof.

REMARK 7.9. Suppose that β corresponds to a partition $\{\Delta_{h,m} : (h,m) \in J\}$ of $\{1,...,M\}$, indexed by

$$J = \{(h,m) \in \mathbb{Z} \times \mathbb{Z} : 1 \leq h \leq L, l_1(h) \leq m \leq l_2(h)\}$$

where l_1 and $l_2 : \{1,...,L\} \to \mathbb{Z}$ satisfy $l_1(h) \leq l_2(h)$ for all $h \in \{1,...,L\}$, as in Proposition 5.1. Let E be a semistable bundle representing an element of Σ_β with δ-filtration

$$0 = E_0 \subset E_1 \subset ... \subset E_t = E$$

as in Proposition 6.3. If

$$l_1(h_1) - \epsilon_1(h_1) < l_1(h_2) - \epsilon(h_2) + 1$$

for all $h_1, h_2 \in \{1,..,L\}$, or equivalently if

$$\phi(h, l_1(h)) \leq \delta(1) + 1$$

for all $h \in \{1,...,L\}$, then the proof of Proposition 7.8 shows that

$$0 \subset E_{\delta(1)+1} \subset E_{\delta(\delta(1)+1)+1} \subset ... \subset E \tag{7.8}$$

is the maximal Jordan–Hölder filtration of E. Thus for such a β the stratum $\Sigma_{\beta,l}$ is contained in the subset $S^{\max JH}_{[\mathbf{n}_\beta,\mathbf{m}_\beta]}$ of $\mathcal{C}^{ss}$ defined at Definition 4.1, where $\mathbf{n}_\beta$ and $\mathbf{m}_\beta$ are determined by the filtration (7.8). If, on the other hand, there exists $h_0 \in \{1,...,L\}$ with

$$\phi(h_0, l_1(h_0)) > \delta(1) + 1,$$

then $E_{\delta(1)+1}$ may not be the maximal subbundle of E which is a direct sum of stable bundles all with the same slope as E; there may be a subbundle of $E_{\phi(h_0,l_1(h_0))}/E_{\phi(h_0,l_1(h_0)-1)}$ which provides a trivial extension of $E_{\delta(1)+1}$ by a direct sum of stable bundles all with the same slope as E (see Example 8.1 below). However, even in this case a careful analysis of the proof of Proposition 4.5 reveals that it can be modified to show that the intersection $\Sigma_\beta \cap S^{\max JH}_{[\mathbf{n},\mathbf{m}]}$ is a locally closed complex submanifold of $\mathcal{C}^{ss}$ of finite codimension for each $\beta \in \Gamma$ and $[\mathbf{n},\mathbf{m}] \in \mathcal{I}^{ss}$.

DEFINITION 7.10. We shall call a filtration $0 \subset P_1 \subset ... \subset P_\tau \subset E$ of a semistable bundle E a *pivot* if each subbundle P_j has the same slope as E and P_1 is the minimal subbundle of P_τ such that P_τ/P_1 is a direct sum of stable bundles of the same slope, while P_τ/P_1 is the maximal subbundle of E/P_1 which is a direct sum of stable bundles of the same slope. Any pivot determines a filtration

$$0 \subseteq ... \subseteq P_\tau^{-2} = P_1^{-1} \subseteq ... \subseteq P_{\tau-1}^{-1} \subseteq P_\tau^{-1} = P_1^0 = P_1 \subseteq ...$$

$$... \subseteq P_\tau = P_\tau^0 = P_1^{+1} \subseteq P_2^{+1} \subseteq ... \subseteq P_\tau^{+1} = P_1^{+2} \subseteq ... \subseteq E$$

of E where

$$0 \subset P_j^{+1}/P_j \subset P_j^{+2}/P_j \subset ... \subset E/P_j$$

is the maximal Jordan–Hölder filtration of E/P_j, and

$$0 \subset ... \subset P_j^{-2} \subset P_j^{-1} \subset P_j$$

is the minimal Jordan–Hölder filtration of P_j. A filtration $0 = E_0 \subset E_1 \subset ... \subset E_t = E$ of this form (once repetitions have been omitted) for some pivot $0 \subset P_1 \subset ... \subset P_\tau \subset E$ will be called a *pivotal filtration*. It will be called a *strongly pivotal filtration* if every subbundle F of E such that

$$P_{j-1}^m \subseteq F \subset P_j^m$$

for some $m \geq 0$ has

$$0 \subset P_j^m/F \subset P_j^{m+1}/F \subset ... \subset E/F$$

as the maximal Jordan–Hölder filtration of E/F, and every subbundle F of E such that

$$P_j^m \subset F \subseteq P_{j+1}^m$$

for some $m \leq 0$ has

$$0 \subset ... \subset P_j^{m-1} \subset P_j^m \subset F$$

as its minimal Jordan–Hölder filtration.

Note that a pivotal filtration is a δ-filtration where $\delta(k_1)$ is the number of $k_2 \in \{1, ..., t\}$ for which it is not the case that $E_{k_1} = P_{j_1}^{m_1}$ and $E_{k_2} = P_{j_2}^{m_2}$ with $m_1 \geq m_2$ or $m_1 = m_2 - 1$ and $j_1 \geq j_2$; if the associated δ-filtration is balanced then we will call the pivotal filtration balanced.

Theorem 7.11. *Let β be as in Proposition 5.1 and let E be a semistable bundle representing an element of $\Sigma_{\beta,l}$ with δ-filtration*

$$0 = E_0 \subset E_1 \subset ... \subset E_t = E \tag{7.9}$$

such that if $1 \leq k \leq t$ then $E_k/E_{k-1} \cong \bigoplus_{i=1}^q \mathbb{C}^{m_i^k} \otimes D_i$ as in Proposition 6.3. Then (7.9) is a balanced strongly pivotal δ-filtration with pivot

$$0 \subset E_{k_1-1} \subset E_{k_1} \subset ... \subset E_{\delta(k_1)} \subset E$$

for any k_1 satisfying $\delta'(k_-) \leq k_1 \leq k_+$.

Proof: This is an immediate consequence of Propositions 7.4 and 7.6 as in Remark 7.7.

8. Refinements of the Yang-Mills stratification

We thus have three refinements of the Harder-Narasimhan filtration of a holomorphic bundle E over Σ: the maximal Jordan–Hölder filtration, the minimal Jordan–Hölder filtration and the balanced δ-filtration of maximal triviality obtained by applying Proposition 6.3 to the subquotients of the Harder-Narasimhan filtration. Associated to these we have refinements of the Yang-Mills stratification of $\mathcal{C}$, each of which is a stratification of $\mathcal{C}$ by locally closed complex submanifolds of finite codimension, and has the set $\mathcal{C}^s$ of stable holomorphic structures on $\mathcal{E}$ as its open stratum. The first of these refined stratifications is the stratification

$$\{S_{[\mathbf{d},\mathbf{n},\mathbf{m}]}^{\max JH} : [\mathbf{d},\mathbf{n},\mathbf{m}] \in \mathcal{I}\} \tag{8.1}$$

of $\mathcal{C}$ defined in Definition 8.1, and another is the stratification

$$\{S_{[\mathbf{d},\mathbf{n},\mathbf{m}]}^{\min JH} : [\mathbf{d},\mathbf{n},\mathbf{m}] \in \mathcal{I}\} \tag{8.2}$$

defined dually using minimal Jordan–Hölder filtrations as in Remark 4.7 and Definition 4.8. A third refinement is the stratification obtained by applying the stratification $\{\Sigma_\gamma : \gamma \in \Gamma\}$ of $\mathcal{C}^{ss}$, whose indexing set was determined in §5 and whose strata were described in terms of balanced δ-filtrations in §6, to the $\mathcal{C}(n', d')^{ss}$ which appear inductively in the description of the Yang–Mills stratification.

Example 8.1. Recall from Remark 4.7 that the maximal Jordan–Hölder filtration of the direct sum $E \oplus F$ of two semistable bundles of the same slope is the direct sum of the maximal Jordan–Hölder filtrations of E and F with the shorter one extended trivially at the top, while the minimal Jordan–Hölder filtration of $E \oplus F$ is the direct sum of their minimal Jordan–Hölder filtrations with the shorter one extended trivially at the bottom. Suppose now that E and F have balanced δ-filtrations of maximal triviality given by

$$0 \subset E_{l_1(1)} \subset E_{l_1(1)+1} \subset ... \subset E_{l_2(1)} = E$$

and

$$0 \subset F_{l_1(2)} \subset F_{l_1(2)+1} \subset ... \subset F_{l_2(2)} = F$$

where the indices $l_1(1), ..., l_2(1) \in \mathbb{Z}$ and $l_1(2), ..., l_2(2) \in \mathbb{Z}$ have been chosen so that

$$\epsilon(1) = \sum_{m=l_1(1)}^{l_2(1)} m \text{ rank}(E_m/E_{m-1})$$

and

$$\epsilon(2) = \sum_{m=l_1(2)}^{l_2(2)} m \text{ rank}(F_m/F_{m-1})$$

lie in the interval $[-1/2, 1/2)$. To simplify the notation let us assume that $l_1(1) \leq l_1(2) \leq l_2(2) \leq l_2(1)$. If $\epsilon(1) > \epsilon(2)$ then $E \oplus F$ has a balanced δ-filtration given by

$$0 \subset E_{l_1(1)} \oplus 0 \subset ... \subset E_{l_1(2)} \oplus 0 \subset E_{l_1(2)} \oplus F_{l_1(2)} \subset ...$$
$$... \subset E_{m-1} \oplus F_{m-1} \subset E_m \oplus F_{m-1} \subset E_m \oplus F_m \subset ...$$
$$... \subset E_{l_2(2)} \oplus F_{l_2(2)} \subset E_{l_2(2)+1} \oplus F_{l_2(2)} \subset ... \subset E_{l_2(1)} \oplus F_{l_2(2)} = E \oplus F.$$

If we assume that E_i/E_{i-1} and F_j/F_{j-1} are stable for $l_1(1) \leq i \leq l_2(1)$ and $l_1(2) \leq j \leq l_2(2)$, then this filtration has no proper refinements with subquotients of the same slope as $E \oplus F$, so by Corollary 6.4 it is a balanced δ-filtration of $E \oplus F$ with maximal triviality (and in fact it is not hard to check that this is still true without the simplifying assumption). If $\epsilon(2) > \epsilon(1)$ then we replace the filtration above with the balanced δ-filtration

$$0 \subset E_{l_1(1)} \oplus 0 \subset ... \subset E_{l_1(2)-1} \oplus 0 \subset E_{l_1(2)-1} \oplus F_{l_1(2)} \subset ...$$
$$... \subset E_{m-1} \oplus F_{m-1} \subset E_{m-1} \oplus F_m \subset E_m \oplus F_m \subset ...$$
$$... \subset E_{l_2(2)} \oplus F_{l_2(2)} \subset E_{l_2(2)+1} \oplus F_{l_2(2)} \subset ... \subset E_{l_2(1)} \oplus F_{l_2(2)} = E \oplus F.$$

Thus we see that the the maximal Jordan–Hölder filtration, the minimal Jordan–Hölder filtration and the balanced δ-filtration of maximal triviality of a bundle E can all be different from one another, and that none of them is necessarily a refinement of the other two. Nonetheless, the concepts of maximal Jordan–Hölder filtration, minimal Jordan–Hölder filtration and balanced δ-filtration of maximal triviality on a bundle E are related by Theorem 7.11 via the notion of a pivotal filtration (see also Remark 6.8, Propositions 7.4, 7.6 and 7.8 and Remark 7.9).

References

1. M.F. Atiyah and R. Bott, *The Yang-Mills equations over Riemann surfaces* Philos. Trans. Roy. Soc. London Ser. A **308** (1982), 523-615.
2. V. Baranovsky, *Cohomology ring of the moduli space of stable vector bundles with odd determinant*, Izv. Russ. Acad. Nauk. **58 n4** (1994), 204-210.
3. A. Beauville, *Sur la cohomologie de certains espaces de modules de fibrés vectoriels*, Geometry and Physics, (Bombay 1992), Tata Inst. Fund. Res. (1995), 37-40.
4. S. Cappell, R. Lee and E. Miller, *The action of the Torelli group on the homology of representation spaces is non-trivial*, Topology **39** (2000), 851-871.
5. G. Daskalopoulos, *The topology of the space of stable bundles on a Riemann surface*, J. Differential Geometry **36** (1992), 699-746.
6. S. del Baño, *On the Chow motive of some moduli spaces*, J. Reine Angew. Math. **532** (2001), 105-132.
7. U.V. Desale and S. Ramanan, *Poincaré polynomials of the variety of stable bundles*, Math. Ann. **216** (1975), 233-244.
8. S.K. Donaldson, *Gluing techniques in the cohomology of moduli spaces*, in *Topological methods in modern mathematics* (Proceedings of 1991 Stony Brook conference in honour of the sixtieth birthday of J. Milnor), Publish or Perish (1993), 137-170.
9. R. Earl, *The Mumford relations and the moduli of rank three stable bundles*, Compositio Math. **109** (1997), 13-48.
10. R. Earl and F. Kirwan, *The Pontryagin rings of moduli spaces of arbitrary rank holomorphic bundles over a Riemann surface*, J. London Math. Soc. (2) 60 (1999), 835-846.
11. ———, *Complete sets of relations in the cohomology rings of moduli spaces of arbitrary rank holomorphic bundles over a Riemann surface*, 2003 preprint.
12. G. Ellingsrud and S. Strømme, *On the Chow ring of a geometric quotient*, Ann. Math. **130** (1989), 130-159.
13. W. Fulton, *Intersection Theory* (Second Edition) Springer (1998).
14. D. Gieseker, *A Degeneration of the Moduli Spaces of Stable Bundles*, J. Differential Geom. **19** (1984), 173-206.
15. ———, *Geometric Invariant Theory and applications to moduli problems*, Proc. Int. Cong. Math. (Helsinki, 1978) Academiae Scientarium Fennica (Helsinki, 1980).
16. G. Harder and M.S. Narasimhan, *On the cohomology groups of moduli spaces of vector bundles over curves*, Math. Ann. **212** (1975) 215-248.
17. R. Herrera and S. Salamon, *Intersection numbers on moduli spaces and symmetries of a Verlinde formula*, Comm. Math. Phys. **188** (1997), 521-534.
18. L. Jeffrey, Y-H. Kiem, F. Kirwan and J. Woolf, *Cohomology pairings on singular quotients in geometric invariant theory*, Transformation Groups **8** 217-259.
19. ———, *Intersection pairings on singular moduli spaces of bundles over a Riemann surface*, in preparation.
20. L. Jeffrey and F. Kirwan, *Intersection theory on moduli spaces of holomorphic bundles of arbitrary rank on a Riemann surface*, Ann. of Math. **148** (1998), 109-196.
21. A. King and P.E. Newstead, *On the cohomology of the moduli space of rank 2 vector bundles on a curve*, Topology **37** (1998), 407-418.
22. F. Kirwan, *Cohomology of quotients in symplectic and algebraic geometry*, Math. Notes vol. 31 Princeton Univ. Press, Princeton, NJ 1985.
23. ———, *Partial desingularisations of quotients of non-singular varieties and their Betti numbers*, Ann. Math. **122** (1985), 41-85.
24. ———, *On spaces of maps from Riemann surfaces to Grassmannians and applications to the cohomology of moduli of vector bundles* Ark. Math. **24** (1986), 221-275.
25. ———, *Rational intersection cohomology of quotient varieties*, Invent. Math. **86** (1986), 471-505.
26. ———, *On the homology of compactifications of moduli spaces of vector bundles over a Riemann surface*, Proc. London Math. Soc. **53** (1986), 237-266.
27. ———, *Cohomology rings of moduli spaces of bundles over Riemann surfaces*, J. Amer. Math. Soc. **5** (1992), 853-906.
28. ———, *Refinements of the Morse stratification of the normsquare of the moment map*, 2003 preprint.

29. J. Le Potier, *Lectures on Vector Bundles*, Cambridge Studies in Advanced Mathematics 54 CUP (1997).
30. D. Mumford, J. Fogarty and F. Kirwan, *Geometric Invariant Theory*, Third Edition, Ser. Modern Surveys Math. 34, Springer (1994).
31. M.S. Narasimhan and S. Ramanan, *Deformations of the moduli space of vector bundles over an algebraic curve*, Ann. Math. **101** (1975), 391-417.
32. M.S. Narasimhan and C.S. Seshadri, *Stable and unitary vector bundles over an algebraic curve*, Ann. Math. **82** (1965), 540-567.
33. P.E. Newstead, *Characteristic classes of stable bundles of rank 2 over an algebraic curve*, Trans. Amer. Math. Soc. **169** (1972), 337-345.
34. ———, *Introduction to moduli problems and orbit spaces*, Tata Inst. Lect. 51 (1978).
35. C.S. Seshadri, *Space of unitary vector bundles on a compact Riemann surface*, Ann. Math. 85 (1967), 303-336.
36. S.S. Shatz, *The decomposition and specialization of algebraic families of vector bundles*, Compositio Math. **35** (1977), 163-187.
37. B. Siebert and G. Tian, *Recursive relations for the cohomology ring of moduli spaces of stable bundles*, Tr. J. of Math. **19** (1995), 131-144.
38. M. Thaddeus, *Conformal field theory and the cohomology of the moduli space of stable bundles*, J. Differential Geom. **35** (1992), 131-149.
39. ———, *An introduction to the topology of the moduli space of stable bundles on a Riemann surface*, in *Geometry and Physics* Lecture Notes in Pure and Applied Mathematics **184** Dekker (1997), 71-100.
40. E. Witten, *Two dimensional gauge theories revisited*, J. Geom. Phys. **9** (1992), 303-368.
41. D. Zagier, *On the cohomology of moduli spaces of rank 2 vector bundles over curves*, Progress in Mathematics **129** The Moduli space of Curves (1995), 533-563.
42. ———, *Elementary aspects of the Verlinde formula and of the Harder–Narasimhan–Atiyah–Bott formula*, Proceedings of the Hirzebruch 65 Conference on Algebraic Geometry (Ramat Gan, 1993), Israel Math. Conf. Proc. **9** (1996), 445-462.

Mathematical Institute, Oxford University, Oxford OX1 3LB, UK
E-mail address: kirwan@maths.ox.ac.uk

Clay Mathematics Proceedings
Volume 3, 2004

On a classical correspondence between K3 surfaces II

Carlo Madonna and Viacheslav V. Nikulin

To memory of Andrei Nikolaevich Tyurin

Abstract. Let X be a K3 surface and H a primitive polarization of degree $H^2 = 2a^2$, $a > 1$. The moduli space of sheaves over X with the isotropic Mukai vector (a, H, a) is again a K3 surface Y which is endowed by a natural *nef* element h with $h^2 = 2$. We give necessary and sufficient conditions in terms of Picard lattices $N(X)$ and $N(Y)$ when $Y \cong X$, generalising our results in [**4**] for $a = 2$.

In particular, we show that $Y \cong X$ if for at least one $\alpha = \pm 1$ there exists $h_1 \in N(X)$ such that $h_1^2 = 2\alpha a$, $H \cdot h_1 \equiv 0 \mod a$, and the primitive sublattice $[H, h_1]_{pr} \subset N(X)$ contains x such that $x \cdot H = 1$.

We also show that all divisorial conditions on moduli of (X, H) (i.e. for Picard number 2) which imply $Y \cong X$ and $H \cdot N(X) = \mathbb{Z}$ are labelled by pairs $(d, \pm\mu)$ where $d \in \mathbb{N}$, $\pm\mu \subset (\mathbb{Z}/2a^2)^*$ such that $d \equiv \mu^2 \mod 4a^2$ and at least for one of $\alpha = \pm 1$ the equation $p^2 - dq^2 = 4a/\alpha$ has an integral solution (p, q) with $p \equiv \mu q \mod 2a$. For each such $\pm\mu$ and α, the number of d and the corresponding divisorial conditions is infinite. Some of these conditions were found (in different form) by A.N. Tyurin in 1987.

Introduction

Let X be a K3 surface with a primitive polarization H of degree $H^2 = 2a^2$. Let Y be the moduli of sheaves over X with the isotropic Mukai vector $v = (a, H, a)$ (see [**6**], [**7**]). Then Y is a K3 surface which is endowed with a natural *nef* element h with $h^2 = 2$. It is isogenous to X in the sense of Mukai.

Question 1. When is Y isomorphic to X?

We want to answer this question in terms of the Picard lattices $N(X)$ and $N(Y)$ of X and Y. Then our question reads as follows:

Question 2. Assume that N is a hyperbolic lattice, $\widetilde{H} \in N$ a primitive element with square $2a^2$. What are conditions on N and $\widetilde{H}$ such that for any K3 surface X with Picard lattice $N(X)$ and a primitive polarization $H \in N(X)$ of degree $2a^2$ the corresponding K3 surface Y is isomorphic to X if the pairs of lattices $(N(X), H)$ and $(N, \widetilde{H})$ are isomorphic as abstract lattices with fixed elements?

1991 *Mathematics Subject Classification.* 14J28.

Supported by Russian Fund of Fundamental Research (grant N 00-01-00170).

In other words, what are conditions on $(N(X), H)$ as an abstract lattice with a primitive vector H with $H^2 = 2a^2$ which are sufficient for Y to be isomorphic to X and are necessary if X is a general K3 surface with Picard lattice $N(X)$?

In [**4**] we answered this question when $a = 2$. Here we give an answer for any $a \geq 2$. For odd a, we additionally assume that $H \cdot N(X) = \mathbb{Z}$. (For even a, this is valid if $Y \cong X$.) The answer is given in Theorem 2.2 and also Proposition 2.1.

In particular, if the Picard number $\rho(X) = \mathrm{rk}\ N(X) \geq 12$, the result is very simple: $Y \cong X$ if and only if either there exists $x \in N(X)$ such that $x \cdot H = 1$ or a is odd and there exists $x \in N(X)$ such that $x \cdot H = 2$. This follows from results of Mukai [**7**] and also [**9**], [**10**].

The polarized K3 surfaces (X, H) with $\rho(X) = 2$ are especially interesting. Indeed, it is well-known that the moduli space of polarized K3 surfaces of any even degree H^2 is 19-dimensional. If X is general, i.e. $\rho(X) = 1$, then the surface Y cannot be isomorphic to X because $N(X) = \mathbb{Z}H$ where $H^2 = 2a^2$, $a > 1$, and $N(X)$ does not have elements with square 2, which is necessary if $Y \cong X$. Thus, if $Y \cong X$, then $\rho(X) \geq 2$ and X belongs to a codimension $\rho(X) - 1$ submoduli space of K3 surfaces. When $\rho = 2$ to describe connected components of the divisor it is equivalent to describe Picard lattices $N(X)$ of the surfaces X with fixed $H \in N(X)$ such that $\rho(X) = \mathrm{rk}\ N(X) = 2$ and a general K3 surface X with Picard lattice $N(X)$ has $Y \cong X$.

The pair $(N(X), H)$ with $\rho = \mathrm{rk}\ N(X) = 2$ and $H \cdot N(X) = \mathbb{Z}$ is defined up to isomorphism by $d = -\det N(X) > 0$ (it defines the Picard lattice $N(X)$ up to isomorphism, if $Y \cong X$) and by the invariant $\pm\mu = \{\mu, -\mu\} \subset (\mathbb{Z}/2a^2)^*$ (this is the invariant of the primitive vector $H \in N(X)$) such that $\mu^2 \equiv d \mod 4a^2$. See Proposition 3.1 concerning the definition of μ. We show that for a general X with such $N(X)$ we have $Y \cong X$ and for odd a additionally $H \cdot N(X) = \mathbb{Z}$, if and only if at least for one $\alpha = \pm 1$ there exists integral (p, q) such that

$$p^2 - dq^2 = 4a/\alpha \text{ and } p \equiv \mu q \mod 2a. \tag{1}$$

For each such $\pm\mu$ and α the set $\mathcal{D}_\alpha^\mu$ of d having such a solution (p, q) and $d \equiv \mu^2 \mod 4a^2$ is infinite since it contains the infinite subset

$$\{(\mu + 2ta/\alpha)^2 - 4a/\alpha > 0 \mid t\mu \equiv 1 \mod a\} \tag{2}$$

(set $q = 1$ in (1)). Thus, the set of possible divisorial conditions on moduli of (X, H) which imply $Y \cong X$ and $H \cdot N(X) = \mathbb{Z}$ is labelled by the set of pairs $(d, \pm\mu)$ described above, and it is infinite.

Some infinite series of divisorial conditions on moduli of X which imply $Y \cong X$ were found by A.N. Tyurin in [**17**] — [**19**]. He found, in different form, infinite series (2) for $\alpha = -1$ and any $\pm\mu$.

Surprisingly, solutions (p, q) of (1) can be interpreted as elements of the Picard lattices of X and Y. We get the following simple sufficient condition on (X, H) which implies $Y \cong X$. It seems that many known examples when it happens that $Y \cong X$ (e.g. see [**2**], [**8**], [**17**]) follow from this condition. This is one of the main results of the paper, and we want to formulate it exactly (a similar statement can also be formulated in terms of Y).

THEOREM 0.1. *Let X be a K3 surface with a primitive polarization H of degree $2a^2$, $a \geq 2$. Let Y be the moduli space of sheaves on X with the Mukai vector $v = (a, H, a)$.*

Then $Y \cong X$ if at least for one $\alpha = \pm 1$ there exists $h_1 \in N(X)$ such that

$$(h_1)^2 = 2\alpha a, \quad h_1 \cdot H \equiv 0 \mod a, \tag{3}$$

and the primitive sublattice $[H, h_1]_{pr} \subset N(X)$ generated by H, h_1 contains x such that $x \cdot H = 1$.

These conditions are necessary to have $Y \cong X$ and $H \cdot N(X) = \mathbb{Z}$ for a odd, if either $\rho(X) = 1$, or $\rho(X) = 2$ and X is a general K3 surface with its Picard lattice.

From our point of view, this statement is also very interesting because some elements h_1 of the Picard lattice $N(X)$ with negative square $(h_1)^2$ receive a very clear geometrical meaning (when $\alpha < 0$). For K3 surfaces this is well-known only for elements δ of the Picard lattice $N(X)$ with negative square $\delta^2 = -2$: then δ or $-\delta$ is effective.

As for the case $a = 2$, the fundamental tool to get the results above is the Global Torelli Theorem for K3 surfaces [**12**] and results of Mukai [**6**], [**7**]. Using the results of Mukai, we can calculate periods of Y using periods of X; by the Global Torelli Theorem [**12**], we can find out if Y is isomorphic to X.

1. Preliminary notations and results about lattices and K3 surfaces

1.1. Some notations about lattices. We use notations and terminology from [**10**] about lattices, their discriminant groups and forms. A *lattice* L is a non-degenerate integral symmetric bilinear form, i.e. L is a free $\mathbb{Z}$-module equipped with a symmetric pairing $x \cdot y \in \mathbb{Z}$ for x, $y \in L$, and this pairing should be non-degenerate. We denote $x^2 = x \cdot x$. The *signature* of L is the signature of the corresponding real form $L \otimes \mathbb{R}$. The lattice L is called *even* if x^2 is even for any $x \in L$. Otherwise, L is called *odd*. The *determinant* of L is defined to be $\det L = \det(e_i \cdot e_j)$ where $\{e_i\}$ is some basis of L. The lattice L is *unimodular* if $\det L = \pm 1$. The *dual lattice* of L is $L^* = Hom(L, \mathbb{Z}) \subset L \otimes \mathbb{Q}$. The *discriminant group* of L is $A_L = L^*/L$. It has order $|\det L|$. The group A_L is equipped with the *discriminant bilinear form* $b_L : A_L \times A_L \to \mathbb{Q}/\mathbb{Z}$ and the *discriminant quadratic form* $q_L : A_L \to \mathbb{Q}/2\mathbb{Z}$ if L is even. To get these forms, one should extend the form of L to a form on the dual lattice L^* with values in $\mathbb{Q}$.

For $x \in L$, we shall consider the invariant $\gamma(x) \geq 0$, where

$$x \cdot L = \gamma(x)\mathbb{Z}. \tag{4}$$

Clearly, $\gamma(x) | x^2$ if $x \neq 0$.

We denote by $L(k)$ the lattice obtained from a lattice L by multiplication of the form of L by $k \in \mathbb{Q}$. The orthogonal sum of lattices L_1 and L_2 is denoted by $L_1 \oplus L_2$. For a symmetric integral matrix A, we denote by $\langle A \rangle$ a lattice which is given by the matrix A in some basis. We denote

$$U = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}. \tag{5}$$

Any even unimodular lattice of signature $(1, 1)$ is isomorphic to U.

An embedding $L_1 \subset L_2$ of lattices is called *primitive* if L_2/L_1 has no torsion. We denote by $O(L)$, $O(b_L)$ and $O(q_L)$ the automorphism groups of the corresponding forms. Any $\delta \in L$ with $\delta^2 = -2$ defines a reflection $s_\delta \in O(L)$ which is given by the formula

$$x \to x + (x \cdot \delta)\delta,$$

$x \in L$. All such reflections generate the *2-reflection group* $W^{(-2)}(L) \subset O(L)$.

1.2. Some notations about K3 surfaces. Here we recall some basic notions and results about K3 surfaces, e.g. see [**12**], [**13**], [**14**]. A K3 surface S is a non-singular projective algebraic surface over $\mathbb{C}$ such that its canonical class K_S is zero and the irregularity $q_S = 0$. We denote by $N(S)$ the *Picard lattice* of S, which is a hyperbolic lattice with the intersection pairing $x \cdot y$ for $x, y \in N(S)$. Since the canonical class $K_S = 0$, the space $H^{2,0}(S)$ of 2-dimensional holomorphic differential forms on S has dimension one over $\mathbb{C}$, and

$$N(S) = \{x \in H^2(S, \mathbb{Z}) \mid x \cdot H^{2,0}(S) = 0\} \tag{6}$$

where $H^2(S, \mathbb{Z})$ with the intersection pairing is a 22-dimensional even unimodular lattice of signature $(3, 19)$. The orthogonal lattice $T(S)$ to $N(S)$ in $H^2(S, \mathbb{Z})$ is called the *transcendental lattice of* S. We have $H^{2,0}(S) \subset T(S) \otimes \mathbb{C}$. The pair $(T(S), H^{2,0}(S))$ is called the *transcendental periods of* S. The *Picard number* of S is $\rho(S) = \text{rk } N(S)$. A non-zero element $x \in N(S) \otimes \mathbb{R}$ is called *nef* if $x \neq 0$ and $x \cdot C \geq 0$ for any effective curve $C \subset S$. It is known that an element $x \in N(S)$ is ample if $x^2 > 0$, x is nef, and the orthogonal complement $x^\perp$ to x in $N(S)$ has no elements with square -2. For any element $x \in N(S)$ with $x^2 \geq 0$, there exists a reflection $w \in W^{(-2)}(N(S))$ such that the element $\pm w(x)$ is *nef*; it then is ample if $x^2 > 0$ and $x^\perp$ has no elements with square -2 in $N(S)$.

We denote by $V^+(S)$ the light cone of S, which is the half-cone of

$$V(S) = \{x \in N(S) \otimes \mathbb{R} \mid x^2 > 0\ \} \tag{7}$$

containing a polarization of S. In particular, all nef elements x of S belong to $\overline{V^+(S)}$: one has $x \cdot V^+(S) > 0$ for them.

The reflection group $W^{(-2)}(N(S))$ acts in $V^+(S)$ discretely, and its fundamental chamber is the closure $\overline{\mathcal{K}(S)}$ of the Kähler cone $\mathcal{K}(S)$ of S. It is the same as the set of all nef elements of S. Its faces are orthogonal to the set $\text{Exc}(S)$ of all exceptional curves r on S, which are non-singular rational curves r on S with $r^2 = -2$. Thus, we have

$$\overline{\mathcal{K}(S)} = \{0 \neq x \in \overline{V^+(S)} \mid x \cdot \text{Exc}(S) \geq 0\,\}. \tag{8}$$

2. General results on the Mukai correspondence between K3 surfaces with primitive polarizations of degrees $2a^2$ and 2 which gives isomorphic K3's

2.1. The correspondence. Let X be a K3 surface with a primitive polarization H of degree $2a^2$, $a > 0$. Let Y be the moduli space of (coherent) sheaves $\mathcal{E}$ on X with the isotropic Mukai vector $v = (a, H, a)$. This means that $\text{rk } \mathcal{E} = a$, $\chi(\mathcal{E}) = 2a$ and $c_1(\mathcal{E}) = H$. Let

$$H^*(X, \mathbb{Z}) = H^0(X, \mathbb{Z}) \oplus H^2(X, \mathbb{Z}) \oplus H^4(X, \mathbb{Z}) \tag{9}$$

be the full cohomology lattice of X with the Mukai product

$$(u, v) = -(u_0 \cdot v_2 + u_2 \cdot v_0) + u_1 \cdot v_1 \tag{10}$$

for $u_0, v_0 \in H^0(X, \mathbb{Z})$, $u_1, v_1 \in H^2(X, \mathbb{Z})$, $u_2, v_2 \in H^4(X, \mathbb{Z})$. We naturally identify $H^0(X, \mathbb{Z})$ and $H^4(X, \mathbb{Z})$ with $\mathbb{Z}$. Then the Mukai product is

$$(u, v) = -(u_0 v_2 + u_2 v_0) + u_1 \cdot v_1. \tag{11}$$

The element

$$v = (a, H, a) = (a, H, \chi - a) \in H^*(X, \mathbb{Z}) \tag{12}$$

is isotropic, i.e. $v^2 = 0$. In this case, Mukai showed [**6**], [**7**] that Y is a K3 surface, and one has the natural identification

$$H^2(Y, \mathbb{Z}) = (v^\perp/\mathbb{Z}v) \tag{13}$$

which also gives the isomorphism of the Hodge structures of X and Y. The element $h = (-1, 0, 1) \in v^\perp$ has square $h^2 = 2$, $h \mod \mathbb{Z}v$ belongs to the Picard lattice $N(Y)$ of Y and is *nef*. See [**5**], [**13**] and [**15**] concerning the geometry of (Y, h). For a general X, the K3 surface Y is a double plane.

We want to answer Question 2 which we precisely formulated in the Introduction: Using $N(X)$, say when $Y \cong X$.

2.2. Formulation of general results. We use notations from Sect. 2.1. Thus, we assume that X is a K3 surface with a primitive polarization H with $H^2 = 2a^2$ where $a > 1$. The following statement follows from results of Mukai [**7**] and some results from [**10**]. It is standard and well-known.

PROPOSITION 2.1. *If Y is isomorphic to X, then either $\gamma(H) = 1$ or a is odd and $\gamma(H) = 2$ for $H \in N(X)$ (see (4)).*

Assume that either $\gamma(H) = 1$ or a is odd and $\gamma(H) = 2$ for $H \in N(X)$. Then the Mukai identification (13) canonically identifies the transcendental periods $(T(X), H^{2,0}(X))$ and $(T(Y), H^{2,0}(Y))$. It follows that the Picard lattices $N(Y)$ and $N(X)$ have the same genus. In particular, $N(Y)$ is isomorphic to $N(X)$ if the genus of $N(X)$ contains only one class. If the genus of $N(X)$ contains only one class, then Y is isomorphic to X, if additionally the canonical homomorphism $O(N(X)) \to O(q_{N(X)})$ is epimorphic. Both these conditions are valid (in particular, $Y \cong X$), if $\rho(X) \geq 12$.

From now on we assume that $\gamma(H) = 1$ in $N(X)$, which is automatically valid for even a if $Y \cong X$.

The calculations below are valid for an arbitrary K3 surface X and a primitive vector $H \in N(X)$ with $H^2 = 2a^2$, $a > 0$ and $\gamma(H) = 1$. Let $K(H) = H^\perp_{N(X)}$ be the orthogonal complement to H in $N(X)$. Set $H^* = H/2a^2$. Then any element $x \in N(X)$ can be written as

$$x = nH^* + k^* \tag{14}$$

where $n \in \mathbb{Z}$ and $k^* \in K(H)^*$, because

$$\mathbb{Z}H \oplus K(H) \subset N(X) \subset N(X)^* \subset \mathbb{Z}H^* \oplus K(H)^*.$$

Since $\gamma(H) = 1$, the map $nH^* + [H] \mapsto k^* + K(H)$ gives an isomorphism of the groups $\mathbb{Z}/2a^2 \cong [H^*]/[H] \cong [u^* + K(H)]/K(H)$ where $u^* + K(H)$ has order $2a^2$ in $A_{K(H)} = K(H)^*/K(H)$. It follows that

$$N(X) = [\mathbb{Z}H, K(H), H^* + u^*]. \tag{15}$$

The element u^* is defined canonically mod $K(H)$. Since $H^* + u^*$ belongs to the even lattice $N(X)$, it follows that

$$(H^* + u^*)^2 = \frac{1}{2a^2} + {u^*}^2 \equiv 0 \mod 2. \tag{16}$$

Let $\overline{H^*} = H^* \mod [H] \in [H^*]/[H] \cong \mathbb{Z}/2a^2$ and $\overline{k^*} = k^* \mod K(H) \in A_{K(H)} = K(H)^*/K(H)$. We then have

$$N(X)/[H,K(H)] = (\mathbb{Z}/2a^2)(\overline{H^*}+\overline{u^*}) \subset (\mathbb{Z}/2a^2)\overline{H^*} + K(H)^*/K(H). \tag{17}$$

Also $N(X)^* \subset \mathbb{Z}H^* + K(H)^*$ since $H + K(H) \subset N(X)$, and for $n \in \mathbb{Z}$, $k^* \in K(H)^*$ we have $x = nH^* + k^* \in N(X)^*$ if and only if

$$(nH^* + k^*)\cdot(H^* + u^*) = \frac{n}{2a^2} + k^*\cdot u^* \in \mathbb{Z}.$$

It follows that

$$\begin{aligned} N(X)^* &= \{nH^* + k^* \mid n \in \mathbb{Z},\ k^* \in K(H)^*, n \equiv -2a^2\ (k^*\cdot u^*) \mod 2a^2\} \\ &\subset \mathbb{Z}H^* + K(H)^*, \end{aligned} \tag{18}$$

and

$$\begin{aligned} N(X)^*/[H,K(H)] &= \{-2a^2(\overline{k^*}\cdot\overline{u^*})\,\overline{H^*} + \overline{k^*}\} \mid \overline{k^*} \in A_{K(H)}\} \\ &\subset (\mathbb{Z}/2a^2)\overline{H^*} + A_{K(H)}. \end{aligned} \tag{19}$$

We introduce *the characteristic map of the polarization* H

$$\kappa(H) : K(H)^* \to A_{K(H)}/(\mathbb{Z}/2a^2)(u^* + K(H)) \to A_{N(X)} \tag{20}$$

where for $k^* \in K(H)^*$ we have

$$\kappa(H)(k^*) = -2a^2(k^*\cdot u^*)H^* + k^* + N(X) \in A_{N(X)}. \tag{21}$$

It is epimorphic, its kernel is $(\mathbb{Z}/2a^2)(u^* + K(H))$, and it gives the canonical isomorphism

$$\overline{\kappa(H)} : A_{K(H)}/(\mathbb{Z}/2a^2)(u^* + K(H)) \cong A_{N(X)}. \tag{22}$$

For the corresponding discriminant forms we have

$$\kappa(k^*)^2 \mod 2 = (k^*)^2 + 2a^2(k^*\cdot u^*)^2 \mod 2. \tag{23}$$

Now we can formulate our main result:

THEOREM 2.2. *The surface Y is isomorphic to X if the following conditions (a), (b), (c) are valid:*

(a) $\gamma(H) = 1$ for $H \in N(X)$;

(b) there exists $\widetilde{h} \in N(X)$ with $\widetilde{h}^2 = 2$, $\gamma(\widetilde{h}) = 1$ and such that there exists an embedding

$$f : K(H) \to K(\widetilde{h})$$

of negative definite lattices such that

$$K(\widetilde{h}) = [f(K(H)), 2af(u^*)],\ w^* + K(\widetilde{h}) = af(u^*) + K(\widetilde{h});$$

(c) the dual to f embedding $f^* : K(\widetilde{h})^* \to K(H)^*$ commutes (up to multiplication by ± 1) with the characteristic maps $\kappa(H)$ and $\kappa(\widetilde{h})$, i.e.

$$\kappa(\widetilde{h})(k^*) = \pm\kappa(H)(f^*(k^*)) \tag{24}$$

for any $k^* \in K(\widetilde{h})^*$.

The conditions (a), (b) and (c) are necessary if for odd a additionally $\gamma(H) = 1$ for $H \in N(X)$, and rk $N(X) \le 19$, and X is a general K3 surface with Picard lattice $N(X)$ in the following sense: the automorphism group of the transcendental periods $(T(X), H^{2,0}(X))$ is ± 1. (Recall that $Y \cong X$ if rk $N(X) = 20$.)

2.3. Proofs. Let us denote by e_1 the canonical generator of $H^0(X,\mathbb{Z})$ and by e_2 the canonical generator of $H^4(X,\mathbb{Z})$. They generate the sublattice U in $H^*(X,\mathbb{Z})$. Consider the Mukai vector $v = ae_1 + H + ae_2 = (a, H, a)$. We have

$$N(Y) = v^{\perp}_{U\oplus N(X)}/\mathbb{Z}v. \tag{25}$$

Let us calculate $N(Y)$. Let $K(H) = H^{\perp}_{N(X)}$. Then we have the embedding of lattices of finite index

$$\mathbb{Z}H \oplus K(H) \subset N(X) \subset N(X)^* \subset \mathbb{Z}H^* \oplus K(H)^* \tag{26}$$

where $H^* = H/2a^2$. We have the orthogonal decomposition up to finite index

$$U \oplus \mathbb{Z}H \oplus K(H) \subset U \oplus N(X) \subset U \oplus \mathbb{Z}H^* \oplus K(H)^*. \tag{27}$$

Let $s = x_1e_1 + x_2e_2 + yH^* + z^* \in v^{\perp}_{U\oplus N(X)}$, $z^* \in K(H)^*$. Then $-ax_1 - ax_2 + y = 0$ since $s \in v^{\perp}$ and hence $(s, v) = 0$. Thus, $y = ax_1 + ax_2$ and

$$s = x_1e_1 + x_2e_2 + a(x_1 + x_2)H^* + z^*. \tag{28}$$

Here $s \in U \oplus N(X)$ if and only if $x_1, x_2 \in \mathbb{Z}$ and $a(x_1 + x_2)H^* + z^* \in N(X)$. This orthogonal complement contains

$$[\mathbb{Z}v, K(H), \mathbb{Z}h] \tag{29}$$

where $h = -e_1 + e_2$, and this is a sublattice of finite index in $(v^{\perp})_{U\oplus N(X)}$. The generators v, generators of $K(H)$ and h are free, and we can rewrite s above using these generators with rational coefficients as follows:

$$s = \frac{-x_1 + x_2}{2}h + \frac{x_1 + x_2}{2a}v + z^*, \tag{30}$$

where $a(x_1 + x_2)H^* + z^* \in N(X)$. Equivalently, for $h^* = h/2$,

$$s = x_1'h^* + x_2'\frac{v}{2a} + z^*, \tag{31}$$

where $x_1', x_2' \in \mathbb{Z}$, $z^* \in K(H)^*$, $x_1' \equiv x_2' \mod 2$, and $ax_2'H^* + z^* \in N(X)$.

From these calculations, we get

Claim. Assume that $\gamma(H) = 1$. Then

$$N(X) = [H, K(H), \frac{H}{2a^2} + u^*], \tag{32}$$

$$N(Y) = [h, K(h) = [K(H), 2au^*], \frac{h}{2} + au^*] = [h, K(h), \frac{h}{2} + w^*], \tag{33}$$

where $u^* + K(H)$ has order $2a^2$ in $A_{K(H)}$, $w^* = au^*$, $K(h) = [K(H), 2w^* = 2au^*]$. Here we match our notations with Sect. 2.2. We have $\det N(X) = \det K(H)/2a^2$ and $\det N(Y) = \det K(h)/2$ (in particular, $\gamma(h) = 1$). Thus, $\det N(X) = \det N(Y)$ for this case, since $\det K(h) = \det K(H)/a^2$. We can formally put here $h = \frac{H}{a}$ since $h^2 = \left(\frac{H}{a}\right)^2 = 2$.

From the claim, we get

LEMMA 2.3. *For the Mukai identification (13), the sublattice $T(X) \subset T(Y)$ has index* 1 *if $\gamma(H) = 1$ for $H \in N(X)$.*

PROOF. Indeed, since $H \in N(X)$, $T(X) \perp N(X)$ and $T(X) \cap \mathbb{Z}v = \{0\}$, the Mukai identification (13) gives an embedding $T(X) \subset T(Y)$. We then have $\det T(Y) = \det T(X)/[T(Y) : T(X)]^2$. Moreover, $|\det T(X)| = |\det N(X)|$ and $|\det T(Y)| = |\det N(Y)|$ because the transcendental and the Picard lattice are orthogonal complements to each other in the unimodular lattice $H^2(X,\mathbb{Z}) \cong H^2(Y,\mathbb{Z})$. By (32) and (33) we get the statement. $\square$

The statement of Lemma 2.3 is a particular case of the general statement by Mukai [**7**] that

$$[T(Y) : T(X)] = q = \min |v \cdot x| \tag{34}$$

for all $x \in H^0(X,\mathbb{Z}) \oplus N(X) \oplus H^4(X,\mathbb{Z})$ such that $v \cdot x \neq 0$. For our Mukai vector $v = (a, H, a)$ we obviously get that $q = 1$ if and only if for $H \in N(X)$ either $\gamma(H) = 1$ or $\gamma(H) = 2$ and a is odd. Thus, for a even we have the only case $\gamma(H) = 1$.

Proof of Proposition 2.1. By (34), if $Y \cong X$, we have either $\gamma(H) = 1$ or $\gamma(H) = 2$ and a is odd.

Assume that $\gamma(H) = 1$ or $\gamma(H) = 2$ and a is odd. Then $T(X) = T(Y)$ for the Mukai identification (13). By the discriminant forms technique (see [**10**]), the discriminant quadratic forms $q_{N(X)} = -q_{T(X)}$ and $q_{N(Y)} = -q_{T(Y)}$ are isomorphic. Thus, the lattices $N(X)$ and $N(Y)$ have the same signatures and discriminant quadratic forms. It follows (see [**10**]) that they have the same genus: $N(X) \otimes \mathbb{Z}_p \cong N(Y) \otimes \mathbb{Z}_p$ for any prime p and the ring of p-adic integers $\mathbb{Z}_p$. Additionally, assume that either the genus of $N(X)$ or the genus of $N(Y)$ contains only one class. Then $N(X)$ and $N(Y)$ are isomorphic.

If additionally the canonical homomorphism $O(N(X)) \to O(q_{N(X)})$ (equivalently, $O(N(Y)) \to O(q_{N(Y)})$) is epimorphic, then the Mukai identification $T(X) = T(Y)$ can be extended to give an isomorphism $\phi : H^2(X,\mathbb{Z}) \to H^2(Y,\mathbb{Z})$ of cohomology lattices. The Mukai identification is an identification on $H^{2,0}(X) = H^{2,0}(Y)$. Multiplying ϕ by ± 1 and by elements of the reflection group $W^{(-2)}(N(X))$, if necessary, we can assume that $\phi(H^{2,0}(X)) = H^{2,0}(Y)$ and ϕ maps the Kähler cone of X to the Kähler cone of Y. By the global Torelli Theorem for K3 surfaces [**12**], ϕ is then defined by an isomorphism of K3 surfaces X and Y.

If $\rho(X) \geq 12$, by [**10**] Theorem 1.14.4, the primitive embedding of $T(X) = T(Y)$ into the cohomology lattice $H^2(X,\mathbb{Z})$ of K3 surfaces is unique up to automorphisms of the lattice $H^2(X,\mathbb{Z})$. As above, it then follows that X is isomorphic to Y. The given proof of Proposition 2.1 is standard and well-known. $\square$

Proof of Theorem 2.2. Assume that $\gamma(H) = 1$. The Mukai identification then gives the canonical identification

$$T(X) = T(Y). \tag{35}$$

Thus, it gives the canonical identifications

$$\begin{aligned} A_{N(X)} &= N(X)^*/N(X) = (U \oplus N(X))^*/(U \oplus N(X)) = T(X)^*/T(X) \\ &= A_{T(X)} = A_{T(Y)} = T(Y)^*/T(Y) = N(Y)^*/N(Y) = A_{N(Y)}. \end{aligned} \tag{36}$$

Here $A_{N(X)} = N(X)^*/N(X) = (U \oplus N(X))^*/(U \oplus N(X))$ because U is unimodular, $(U \oplus N(X))^*/(U \oplus N(X)) = T(X)^*/T(X) = A_{T(X)}$ because $U \oplus N(X)$ and $T(X)$ are orthogonal complements to each other in the unimodular lattice $H^*(X,\mathbb{Z})$. Here $A_{T(Y)} = T(Y)^*/T(Y) = N(Y)^*/N(Y) = A_{N(Y)}$ because $T(Y)$ and $N(Y)$ are orthogonal complements to each other in the unimodular lattice $H^2(Y,\mathbb{Z})$. E. g. the identification $(U\oplus N(X))^*/(U\oplus N(X)) = T(X)^*/T(X) = A_{T(X)}$ is given by the canonical correspondence

$$x^* + (U\oplus N(X)) \to t^* + T(X) \tag{37}$$

if $x^*\in (U\oplus N(X))^*$, $t^*\in T(X)^*$ and $x^*+t^*\in H^*(X,\mathbb{Z})$.

By (33), we also have the canonical embedding of lattices

$$K(H)\subset K(h) = [K(H), 2au^*]. \tag{38}$$

We have the key statement:

Lemma 2.4. *Assume that $\gamma(H)=1$. The canonical embedding (38) (it is given by (33)) $K(H)\subset K(h)$ of lattices, and the canonical identification $A_{N(X)}=A_{N(Y)}$ (given by (36)) agree with the characteristic homomorphisms $\kappa(H): K(H)^*\to A_{N(X)}$ and $\kappa(h): K(h)^*\to A_{N(Y)}$, i.e. $\kappa(h)(k^*)=\kappa(H)(k^*)$ for any $\kappa^*\in K(h)^*\subset K(H)^*$ (this embedding is dual to (38)).*

Proof. As the proof of Lemma 2.3.2 in [**4**]. □

Let us finish the proof of Theorem 2.2. We have the Mukai identification (it is defined by (13)) of the transcendental periods

$$(T(X),\, H^{2,0}(X)) = (T(Y),\, H^{2,0}(Y)). \tag{39}$$

For general X with the Picard lattice $N(X)$, it is the unique isomorphism of the transcendental periods up to multiplication by ±1. If $X\cong Y$, this (up to ±1) isomorphism can be extended to $\phi: H^2(X,\mathbb{Z})\cong H^2(Y,\mathbb{Z})$. The restriction of ϕ on $N(X)$ gives then isomorphism $\phi_1: N(X)\cong N(Y)$ which is ±1 on $A_{N(X)}=A_{N(Y)}$ under the identification (36). The element $\widetilde{h}=(\phi_1)^{-1}(h)$ and $f=\phi^{-1}$ satisfy Theorem 2.2 by Lemma 2.4.

The other way around, under conditions of Theorem 2.2, by Lemma 2.4, one can construct an isomorphism $\phi_1: N(X)\cong N(Y)$ which is ±1 on $A_{N(X)}=A_{N(Y)}$. It can be extended to be ±1 on the transcendental periods under the Mukai identification (39). Then it is defined by the isomorphism $\phi: H^2(X,\mathbb{Z})\to H^2(Y,\mathbb{Z})$. Multiplying ϕ by ±1 and by reflections from $W^{(-2)}(N(X))$, if necessary (the group $W^{(-2)}(N(X))$ acts identically on the discriminant group $N(X)^*/N(X)$), we can assume that ϕ maps the Kähler cone of X to the Kähler cone of Y. By the global Torelli Theorem for K3 surfaces [**12**], it is then defined by an isomorphism of X and Y. □

3. The case of Picard number 2

3.1. The case when $\rho(X)=2$ and $\gamma(H)=1$ for a odd. Here we apply results of Sect. 2 to X and Y with Picard number 2.

We start with some preliminary considerations on K3 surfaces X with Picard number 2 and a primitive polarization H of degree $H^2=2a^2$, $a\ge1$. Thus, we assume that rk $N(X)=2$. Additionally, we assume that $\gamma(H)=1$ for $H\in N(X)$ (we have this condition if a is even and $Y\cong X$). Let

$$K(H)=H^{\perp}_{N(X)}=\mathbb{Z}\delta$$

and $\delta^2 = -t$ where $t > 0$ is even. Then $\delta \in N(X)$ is defined uniquely up to $\pm\delta$. It then follows that

$$N(X) = [\mathbb{Z}H,\, \mathbb{Z}\delta,\, \mu H^* + \frac{\delta}{2a^2}]$$

where $H^* = H/2a^2$ and $g.c.d.(\mu, 2a^2) = 1$. The element

$$\pm\mu \mod 2a^2 \in (\mathbb{Z}/2a^2)^*$$

is *the invariant of the pair* $(N(X), H)$ up to isomorphisms of lattices with the primitive vector H of $H^2 = 2a^2$. If δ changes to $-\delta$, then $\mu \mod 2a^2$ changes to $-\mu \mod 2a^2$. We have

$$(\mu H^* + \frac{\delta}{2a^2})^2 = \frac{1}{2a^2}(\mu^2 - \frac{t}{2a^2}) \equiv 0 \mod 2.$$

Then $t = 2a^2 d$, for some $d \in \mathbb{N}$ and $\mu^2 \equiv d \mod 4a^2$. Thus, $d \mod 4a^2 \in (\mathbb{Z}/4a^2)^{*\,2}$. Obviously, $-d = \det(N(X))$.

Any element $z \in N(X)$ can be written as $z = (xH + y\delta)/2a^2$ where $x \equiv \mu y \mod 2a^2$. In these considerations, one can replace H by any primitive element of $N(X)$ with square $2a^2$. Thus, we have:

PROPOSITION 3.1. *Let X be a K3 surface with Picard number $\rho = 2$ and a primitive polarization H of degree $H^2 = 2a^2$, $a > 0$, and $\gamma(H) = 1$ for $H \in N(X)$.*

The pair $(N(X), H)$ has the invariants $d \in \mathbb{N}$ and $\pm\mu \mod 2a^2 \in (\mathbb{Z}/2a^2)^*$ such that $\mu^2 \equiv d \mod 4a^2$. (It follows that $d \equiv 1 \mod 4$.)

For the invariants d, μ we have: $\det N(X) = -d$, and $K(H) = H^{\perp}_{N(X)} = \mathbb{Z}\delta$ where $\delta^2 = -2a^2 d$. Moreover,

$$N(X) = [H, \delta, (\mu H + \delta)/2a^2], \tag{40}$$

$$N(X) = \{z = (xH + y\delta)/2a^2 \mid x, y \in \mathbb{Z} \text{ and } x \equiv \mu y \mod 2a^2\}, \tag{41}$$

and $z^2 = (x^2 - dy^2)/2a^2$.

For any primitive element $H' \in N(X)$ with $(H')^2 = H^2 = 2a^2$ and the same invariant $\pm\mu$, there exists an automorphism $\phi \in O(N(X))$ such that $\phi(H) = H'$.

Applying Proposition 3.1 to $a = 1$, we get that the pair $(N(Y), h)$ with $h^2 = 2$ and $\gamma(h) = 1$ is defined by its determinant $\det N(Y) = -d$ where $d \equiv 1 \mod 4$. Thus, from Proposition 3.1, we get

PROPOSITION 3.2. *Under conditions and notations of Propositions 3.1, all elements $h' = (xH + y\delta)/(2a^2) \in N(X)$ with $(h')^2 = 2$ are in one to one correspondence with integral solutions (x, y) of the equation*

$$x^2 - dy^2 = 4a^2 \tag{42}$$

such that $x \equiv \mu y \mod 2a^2$.

The Picard lattices of X and Y are isomorphic, $N(X) \cong N(Y)$, if and only if there exists such a solution.

PROOF. It follows from the fact that $\gamma(h) = 1$ for $h \in N(Y)$ because of (33). □

The crucial statement is

THEOREM 3.3. *Let X be a K3 surface, $\rho(X)=2$ and H a primitive polarization of X of degree $H^2=2a^2$, $a>1$. Let Y be the moduli space of sheaves on X with the Mukai vector $v=(a,H,a)$ and the canonical nef element $h=(-1,0,1) \mod \mathbb{Z}v$. Assume that $\gamma(H)=1$ (for even a only for this case can we have $Y\cong X$). Then we can introduce the invariants $\pm\mu \mod 2a^2 \in (\mathbb{Z}/2a^2)^*$ and $d\in\mathbb{N}$ of $(N(X),H)$ as in Proposition 3.1. Thus, we have*

$$\gamma(H)=1,\ \det N(X)=-d \text{ where } \mu^2\equiv d \mod 4a^2. \tag{43}$$

With the notations of Propositions 3.1, all elements $\widetilde{h}=(xH+y\delta)/(2a^2)\in N(X)$ with square $\widetilde{h}^2=2$ satisfying Theorem 2.2 are in one to one correspondence with integral solutions (x,y) of the equation

$$x^2-dy^2=4a^2 \tag{44}$$

with $x\equiv\mu y \mod 2a^2$ and $x\equiv\pm 2a \mod d$.

In particular (by Theorem 2.2), for a general X with $\rho(X)=2$ and $\gamma(H)=1$ for odd a, we have $Y\cong X$ if and only if the equation $x^2-dy^2=4a^2$ has an integral solution (x,y) with $x\equiv\mu y \mod 2a^2$ and $x\equiv\pm 2a \mod d$. Moreover, a nef element $h=(xH+y\delta)/2a^2$ with $h^2=2$ defines the structure of a double plane on X which is isomorphic to the double plane Y if and only if $x\equiv\pm 2a \mod d$.

PROOF. Let $\widetilde{h}\in N(X)$ satisfy conditions of Theorem 2.2.

By Proposition 3.2, all primitive

$$\widetilde{h}=(xH+y\delta)/(2a^2)\in N(X) \tag{45}$$

with $\widetilde{h}^2=2$ are in one to one correspondence with integral (x,y) which satisfy the equation $x^2-dy^2=4a^2$ and $x\equiv\mu y \mod 2a^2$, and any integral solution of the equation $x^2-dy^2=4a^2$ with $x\equiv\mu y \mod 4a^2$ gives such an $\widetilde{h}$.

Let $k=aH+b\delta\in\widetilde{h}^{\perp}=\mathbb{Z}\alpha$. Then $(k,h)=ax-byd=0$ and $(a,b)=\lambda(yd,x)$. Hence, we have

$$(\lambda(ydH+x\delta))^2=\lambda^2(2a^2y^2d^2-2a^2dx^2)=2a^2\lambda^2d(y^2d-x^2)=-2(2a^2)^2\lambda^2d. \tag{46}$$

Since $\alpha^2=-2d$, we get $\lambda=1/2a^2$ and $\alpha=(ydH+x\delta)/2a^2$. There exists a unique (up to ±1) embedding $f:K(H)=\mathbb{Z}\delta\to K(h)=\mathbb{Z}\alpha$ of one-dimensional lattices. It is given by $f(\delta)=a\alpha$ up to ±1. Thus, its dual is defined by $f^*(\alpha^*)=a\delta^*$ where $\alpha^*=\alpha/2d$ and $\delta^*=\delta/2a^2d$. To satisfy the conditions of Theorem 2.2, we should have

$$\kappa(\widetilde{h})(\alpha^*)=\pm a\kappa(H)(\delta^*). \tag{47}$$

Further we denote $\nu=\mu^{-1} \mod 2a^2$. We have $u^*=\nu d\delta^*$, $w^*=af(u^*)=\nu\frac{\alpha}{2}$, and

$$\kappa(\widetilde{h})(\alpha^*)=(-2\alpha^*\cdot w^*)\widetilde{h}^*+\alpha^*+N(X) \tag{48}$$

by (21). Here $\widetilde{h}^*=\widetilde{h}/2$. We then have $\alpha^*\cdot w^*=-\frac{\nu}{2}$, and

$$\kappa(\widetilde{h})(\alpha^*)=\nu\widetilde{h}^*+\alpha^*+N(X)=(\frac{\nu x+y}{2})H^*+(\frac{\nu yd+x}{2})\delta^* \tag{49}$$

where $H^*=H/2a^2$.

We have $u^*=\nu d\delta^*=\nu\delta/2a^2$. By (21), $\kappa(H)(\delta^*)=(-2a^2\delta^*\cdot u^*)H^*+\delta^*+N(X)$. We have $\delta^*\cdot u^*=-\nu/2a^2$. It follows that

$$\kappa(H)(\delta^*)=\nu H^*+\delta^*. \tag{50}$$

By (49) and (50), we then obtain that $\kappa(H)(a\delta^*) = \pm\kappa(\widetilde{h})(\alpha^*)$ is equivalent to $(\nu yd + x)/2 \equiv \pm a \mod d$ and hence $x + \nu yd \equiv \pm 2a \mod d$ since the group $N(X)^*/N(X)$ is cyclic of order d and it is generated by $\nu H^* + \delta^* + N(X)$. Thus, finally we get $x \equiv \pm 2a \mod d$.

This finishes the proof. $\square$

By Theorem 3.3, for given $a > 1$, $\pm\mu \in (\mathbb{Z}/2a^2)^*$ and $d \in \mathbb{N}$ such that $d \equiv \mu^2 \mod 4a^2$, we should look for all integral (x, y) such that.

$$x^2 - dy^2 = 4a^2,\ x \equiv \mu y \mod 2a^2,\ x \equiv \pm 2a \mod d. \tag{51}$$

We first describe the set of integral (x, y) such that

$$x^2 - dy^2 = 4a^2,\ x \equiv \pm 2a \mod d. \tag{52}$$

Considering $\pm(x, y)$, we can assume that $x \equiv 2a \mod d$. Then $x = 2a - kd$ where $k \in \mathbb{Z}$. We have $4a^2 - 4akd + k^2d^2 - dy^2 = 4a^2$. Thus,

$$d = \frac{y^2 + 4ak}{k^2}.$$

Let l be prime. As in [**4**], it is easy to see that if $l^{2t+1}|k$ and l^{2t+2} does not divide k, then $l^{2t+2}|4ak$. It follows that $k = -\alpha q^2$ where $q \in \mathbb{Z}$, $\alpha \in \mathbb{Z}$ is square-free and $\alpha|2a$. Then

$$d = \frac{y^2 - 4a\alpha q^2}{\alpha^2 q^4}.$$

It follows that $\alpha q|y$ and $y = \alpha qp$ where $p \in \mathbb{Z}$. We then get

$$d = \frac{p^2 - 4a/\alpha}{q^2}. \tag{53}$$

Equivalently,

$$\alpha|2a \text{ is square-free},\ p^2 - dq^2 = \frac{4a}{\alpha}\ . \tag{54}$$

We recall that α can be negative.

Thus, solutions α, (p, q) of (54) give all solutions

$$(x, y) = \pm(2a + \alpha dq^2,\ \alpha pq) \tag{55}$$

of (52). We call them *associated solutions.* Thus, all solutions (x, y) of (52) are associated solutions (55) to all solutions α, (p, q) of (54). If one additionally assumes that $q > 0$, then (x, y) and α, (p, q) are in one to one correspondence (by our construction).

Now let us consider associated solutions (55) which satisfy the additional condition $x \equiv \mu y \mod 2a^2$. We have

$$2a + \alpha dq^2 \equiv \mu\alpha pq \mod 2a^2. \tag{56}$$

Using $d \equiv \mu^2 \mod 4a^2$, we get

$$\frac{2a}{\alpha} \equiv \mu q(p - \mu q) \mod 2a^2/\alpha. \tag{57}$$

Since $p^2 - dq^2 = 4a/\alpha$, we get from (56)

$$-\frac{2a}{\alpha} \equiv -p(p - \mu q) \mod 2a^2/\alpha. \tag{58}$$

Taking the sum of (57) and (58), we get

$$(p-\mu q)^2 \equiv 0 \mod 2a^2/\alpha. \tag{59}$$

Since $\alpha|2a^2$ and α is square-free, it follows easily that $a|(p-\mu q)$ and $(p-\mu q)/a$ is an integer. From (58), we then get

$$a|p-\mu q \ \text{ and } \ 2 \equiv \alpha p\left(\frac{p-\mu q}{a}\right) \mod 2a. \tag{60}$$

The condition $x \equiv \mu y \mod 2a^2$ is equivalent to (60).

From (60), we get $\alpha|2$. Thus, $\alpha = \pm 1$ or $\alpha = \pm 2$. Let us consider both cases.

Assume $\alpha = \pm 1$. By (59), then $2a|(p-\mu q)$, and we can rewrite (60) as

$$2a|p-\mu q \ \text{ and } \ \pm 1 \equiv p\left(\frac{p-\mu q}{2a}\right) \mod a. \tag{61}$$

For $\alpha = \pm 1$ we have $p^2 - dq^2 = \pm 4a$. It follows that $(p-\mu q)(p+\mu q) \equiv \pm 4a \mod 4a^2$. If $2a|(p-\mu q)$, then $((p-\mu q)/2a)(p+\mu q) \equiv ((p-\mu q)/2a)2p \equiv \pm 2 \mod 2a$ which is equivalent to (61).

Thus, for $\alpha = \pm 1$, associated solutions (x, y) to $\alpha = \pm 1$, (p, q) satisfy the additional condition $x \equiv \mu y \mod 2a^2$ if and only if $p \equiv \mu q \mod 2a$. Equivalently, $ap \equiv \mu aq \mod 2a^2$ which is equivalent to $h_1 = (apH + aq\delta)/2a^2 \in N(X)$. The equation $p^2 - dq^2 = \pm 4a$ is equivalent to $(h_1)^2 = \pm 2a$. We also have $h_1 \cdot H = ap \equiv 0 \mod a$. Conversely, assume $h_1 = (uH + v\delta)/2a^2 \in N(X)$, and $(h_1)^2 = \pm 2a$, $h_1 \cdot H \equiv 0 \mod a$. We then have $h_1 \cdot H = u \equiv 0 \mod a$. Thus, $u = ap$ and $h_1 = (ap + v\delta)/2a^2$. Since $h_1 \in N(X)$, then $ap \equiv \mu v \mod 2a^2$. Thus, $v = aq$ and $p \equiv \mu q \mod 2a$. Since $(h_1)^2 = \pm 2a$, we get $p^2 - dq^2 = \pm 4a$. We also remark that from the conditions (60) and $p^2 - dq^2 = \pm 4a$, it follows that

$$\text{g.c.d}(\alpha a, p) = \text{g.c.d}(\alpha a, q) = 1; \ \text{g.c.d}(p, q)|(2/\alpha). \tag{62}$$

Thus, (p, q) is an "almost primitive" solution of the equation $p^2 - dq^2 = \pm 4a$. It is primitive if a is even, and $\text{g.c.d}(p, q)|2$ if a is odd.

Now assume that $\alpha = \pm 2$. Then (60) is equivalent to

$$a|p-\mu q \ \text{ and } \ \pm 1 \equiv p\left(\frac{p-\mu q}{a}\right) \mod a. \tag{63}$$

Assume that $a|p-\mu q$ and $p^2 - dq^2 = \pm 2a$. Then $(p-\mu q)(p+\mu q) \equiv \pm 2a \mod 4a^2$ and $((p-\mu q)/a)(p+\mu q) \equiv ((p-\mu q)/a)2p \equiv \pm 2 \mod a$. If a is odd, this is equivalent to (63).

Assume that a is even. If $(p-\mu q)/a$ is even, then $p+\mu q$ is also even. From $p^2 - dq^2 = \pm 2a$, we then get $((p-\mu q)/2a)(p+\mu q) \equiv \pm 1 \mod 2a$ which is impossible for even $p + \mu q$. Thus, $p - \mu q \equiv a \mod 2a$ and $\mu q \equiv p + a \mod 2a$. From $((p-\mu q)/a)(p+\mu q) \equiv \pm 2 \mod 4a$, we then get $((p-\mu q)/a)(2p+a) \equiv \pm 2 \mod 2a$ and $((p-\mu q)/a)p + ((p-\mu q)/a)(a/2) \equiv \pm 1 \mod a$. Since $(p-\mu q)/a$ is odd, it follows that (63) is never satisfied for even a.

Thus, we get that for $\alpha = \pm 2$ the number a is odd and the condition $x \equiv \mu y \mod 2a^2$ is equivalent to $p \equiv \mu q \mod a$. Let us consider this case. From $p^2 - dq^2 = \pm 2a$ and d odd, we get that $p \equiv q \mod 2$. Since a is odd, we then get $p \equiv \mu q \mod 2a$ and $p^2 \equiv \mu^2 q^2 \mod 4a$. This contradicts $p^2 - dq^2 = \pm 2a$ because $d \equiv \mu^2 \mod 4a^2$. Thus, $\alpha = \pm 2$ is impossible for odd a too.

Finally we get the main results.

THEOREM 3.4. *Under the conditions of Theorem 3.3, for a general X with $\rho(X) = 2$ and $\gamma(H) = 1$ for odd a, we have $Y \cong X$ if and only if at least for one $\alpha = \pm 1$ there exists integral (p, q) such that*

$$p^2 - dq^2 = \frac{4a}{\alpha} \quad \text{and} \quad p \equiv \mu q \mod 2a. \tag{64}$$

Solutions (p, q) of (64) are "almost primitive"; they satisfy (62).

Solutions (p, q) of (64) give all solutions (44) of Theorem 3.3 as associated solutions

$$(x, y) = \pm(2a + \alpha dq^2, \alpha pq).$$

Interpreting, as above, solutions (p, q) of (64) as elements of $N(X)$, we also get

THEOREM 3.5. *Under the conditions of Theorem 3.3, for a general X with $\rho(X) = 2$ and $\gamma(H) = 1$ for odd a, we have $Y \cong X$ if and only if at least for one $\alpha = \pm 1$ there exists $h_1 \in N(X)$ such that*

$$h_1^2 = 2\alpha a \quad \text{and} \quad h_1 \cdot H \equiv 0 \mod a.$$

Applying additionally Theorem 2.2, we get the following simple sufficient condition when $Y \cong X$ which is valid for X with any $\rho(X)$. This is one of the main results of the paper.

THEOREM 3.6. *Let X be a K3 surface and H a primitive polarization of degree $2a^2$, $a \geq 2$. Let Y be the moduli space of sheaves on X with the Mukai vector $v = (a, H, a)$.*

Then $Y \cong X$ if at least for one $\alpha = \pm 1$ there exists $h_1 \in N(X)$ such that

$$(h_1)^2 = 2\alpha a, \quad h_1 \cdot H \equiv 0 \mod a, \tag{65}$$

and $\gamma(H) = 1$ for $H \in [H, h_1]_{pr}$ where $[H, h_1]_{pr}$ is the primitive sublattice of $N(X)$ generated by H, h_1.

This condition is necessary to have $Y \cong X$ and $\gamma(H) = 1$ for a odd if either $\rho(X) = 1$, or $\rho(X) = 2$, and X is a general K3 surface with its Picard lattice (i.e. the automorphism group of the transcendental periods $(T(X), H^{2,0}(X))$ is ± 1).

PROOF. The cases $\rho(X) \leq 2$ have been considered. We can assume that $\rho(X) > 2$. Let $N = [H, h_1]_{\mathrm{pr}}$. All considerations above for $N(X)$ of rk $N(X) = 2$ will be valid for N. We can construct an associated with h_1 solution $\widetilde{h} \in N$ with $\widetilde{h}^2 = 2$ such that H and $\widetilde{h}$ satisfy conditions of Theorem 2.2 for $N(X)$ replaced by N. It is easy to see that the conditions (b) and (c) will still be satisfied if we extend f in (b) by $\pm$ the identity on the orthogonal complement $N^{\perp}_{N(X)}$. $\square$

It seems that many known examples of $Y \cong X$ (e. g. see [**2**], [**8**], [**17**]) follow from Theorem 3.6. Theorems 3.5 and 3.6 are also interesting because they give a very clear geometric interpretation of some elements $h_1 \in N(X)$ with negative square h_1^2 (for negative α).

Below we consider an application of Theorem 3.4.

3.2. Divisorial conditions on moduli (X, H) when $\gamma(H) = 1$ for odd a. Further we use the following notations. We fix $a \in \mathbb{N}$, $\alpha \in \{1, -1\}$ and $\overline{\mu} = \{\mu, -\mu\} \subset (\mathbb{Z}/2a^2)^*$. We denote by $\mathcal{D}(a)^{\overline{\mu}}_{\alpha}$ the set of all $d \in \mathbb{N}$ such that $d \equiv \mu^2 \mod 4a^2$ and there exists an integral (p, q) such that $p^2 - dq^2 = 4a/\alpha$ and $p \equiv \mu q \mod 2a$. We denote by $\mathcal{D}(a)^{\overline{\mu}}$ the union of $\mathcal{D}(a)^{\overline{\mu}}_{\alpha}$ for all $\alpha \in \{1, -1\}$, by $\mathcal{D}(a)_{\alpha}$ the union of $\mathcal{D}(a)_\alpha^{\overline{\mu}}$ for all $\overline{\mu}=\{\mu,-\mu\}\subset(\mathbb{Z}/2a^2)^*$, and by $\mathcal{D}(a)$ the union of all $\mathcal{D}(a)_\alpha$ for all $\alpha\in\{1,-1\}$.

Assume that X is a K3 surface with a primitive polarization H of degree $H^2=2a^2$ where $a>1$. The moduli space Y of sheaves on X with Mukai vector $v=(a,H,a)$ has the canonical *nef* element $h=(-1,0,1) \mod \mathbb{Z}v$ with $h^2=2$. It follows that Y is never isomorphic to X if $\rho(X)=1$. Since the dimension of moduli of (X,H) is equal to $20-\rho(X)$, it follows that describing general (X,H) with $\rho(X)=2$ and $Y\cong X$, we at the same time describe all possible divisorial conditions on moduli of (X,H) when $Y\cong X$. See [**9**], [**10**] and also [**3**]. They are described by invariants of the pairs $(N(X),H)$ where rk $N(X)=2$. By Theorem 3.4, we get

Theorem 3.7. *All possible divisorial conditions on moduli of polarized K3 surfaces (X,H) with a primitive polarization H with $H^2=2a^2$, $a>1$, which imply $Y\cong X$ and $\gamma(H)=1$ for a odd, are labelled by the set $\mathcal{D}iv(a)$ of all pairs*

$$(d,\,\overline{\mu})$$

where $\overline{\mu}=\{\mu,\,-\mu\}\subset(\mathbb{Z}/2a^2)^*$, $d\in\mathcal{D}(a)^{\overline{\mu}}=\bigcup_\alpha\mathcal{D}(a)_\alpha^{\overline{\mu}}$. Here $\alpha\in\{1,-1\}$.

For any $\overline{\mu}=\{\mu,\,-\mu\}\subset(\mathbb{Z}/2a^2)^*$ and any $\alpha\in\{1,-1\}$ the set

$$\begin{aligned}\mathcal{D}(a)_\alpha^{\overline{\mu}} &= \{d=\frac{p^2-4a/\alpha}{q^2}\in\mathbb{N}\mid q\in\mathbb{N},\ p\equiv\mu q \mod 2a,\ d\equiv\mu^2 \mod 4a^2\}\\ &\supset \{(\mu+t(2a/\alpha))^2-4a/\alpha\in\mathbb{N}\mid t\mu\equiv 1 \mod a\}\end{aligned}$$

is infinite (put $q=1$ to get the last infinite subset).

In particular, for any $a>1$, the set of possible divisorial conditions on moduli of (X,H) which imply $Y\cong X$ is infinite.

To enumerate the sets $\mathcal{D}(a)_\alpha^{\overline{\mu}}$, $\alpha\in\{1,-1\}$, it is most important to enumerate the sets $\mathcal{D}(a)_\alpha$. This is almost equivalent to finding all possible $d\in\mathbb{N}$ such that $d \mod 4a^2\in(\mathbb{Z}/4a^2)^{*2}$ and the equation $p^2-dq^2=4a/\alpha$ has a solution (p,q) satisfying (62) (it is "almost primitive"). Each such solution (p,q) defines a unique (if it exists) $\mu \mod 2a^2$ such that $\mu^2\equiv d \mod 4a^2$ and $p\equiv\mu q \mod 2a$. The pair $(d,\,\overline{\mu})$ then gives an element of the set $\mathcal{D}iv(a)$. Thus, to find all possible $\overline{\mu}$ (for the given α, d), it is enough to find all "almost primitive" solutions (p,q) of the equation $p^2-dq^2=4a/\alpha$ and $\nu \mod 2a$ such that $p\equiv\nu q \mod 2a$.

Indeed, let (p,q) be a solution of the equation $p^2-dq^2=4a/\alpha$. For example, assume that it is primitive. Let $\nu\equiv p/q \mod 2a$. Then $\nu^2\equiv d \mod 4a$. We have $\mu=\nu+2ka \mod 2a^2$, $k\in\mathbb{Z}$, and $\mu^2\equiv\nu^2+4k\nu a+4k^2a^2\equiv d \mod 4a^2$. Equivalently,

$$4\nu ka\equiv d-\nu^2 \mod 4a^2.$$

Finally, we get

$$\nu k\equiv\frac{d-\nu^2}{4a} \mod a$$

which determines $\mu \mod 2a^2$ uniquely. If g.c.d$(p,q)=2$, then a is odd. For this case, one should again start with $\nu \mod 2a$ such that $p\equiv\nu q \mod 2a$ and $\nu^2\equiv d \mod 4a$. Then there exists a unique lifting $\mu \mod 2a^2$ such that $\mu\equiv\nu \mod 2a$ and $\mu^2\equiv d \mod 4a^2$.

References

1. Borevich, Z. I. and Shafarevich, I. R., *Number Theory* (3d edition) Nauka, Moscow, 1985, English transl., Academic Press, 1966
2. Cossec, F. R., *Reye Congruences*, Trans. Amer. Math. Soc., **280**, (1983) 737–751.
3. James, D. G., *On Witt's theorem for unimodular quadratic forms*, Pacific J. Math., **26**, (1968), 303–316.
4. Madonna C., and Nikulin, V. V., *On a classical correspondence between K3 surfaces*, Proc. Steklov Math. Inst., **241** (2003) 132-168 (Russian).
5. Mayer, A., *Families of K-3 surfaces*, Nagoya Math. J., **48** (1982), 1–17.
6. Mukai, S., *Symplectic structure of the moduli space of sheaves on an Abelian or K3 surface*, Invent. math., **77**, (1984) 101–116.
7. Mukai, S., *On the muduli space of bundles on K3 surfaces, in Vector bundles on algebraic varieties* (Bombay, 1984), Tata Inst. Fund. Res. Studies in Math. no. 11, Bombay, 1987, 341–413.
8. Mukai, S., *Duality of polarized K3 surfaces*, in *New trends in algebraic geometry. Selected papers presented at the Euro conference*, K. Hulek (ed.), Warwick, UK, July 1996, Cambridge University Press. London Math. Soc. Lect. Notes Ser. 264, Cambridge, 1999, 311–326.
9. Nikulin, V. V., *Finite automorphism groups of Kählerian surfaces of type K3*, Trudy Mosk. Matem. Ob-va, **38**, (1979), 75–137. English transl. in Trans. Moscow Math. Soc., **38**, (1980) 71–135.
10. Nikulin, V. V., *Integral symmetric bilinear forms and some of their geometric applications, Izv. Akad. Nauk SSSR Ser. Mat.*, **43**, (1979) 111–177. English transl. in Math. USSR Izv., **14**, (1980).
11. Nikulin, V. V., *On correspondences between K3 surfaces*, Izv. Akad. Nauk SSSR Ser. Mat., **51**, (1987) 402–411. English transl. in Math. USSR Izv., **30**, (1988).
12. Pjatetski-Shapiro, I. I., and Shafarevich, I. U., *A Torelli theorem for algebraic surfaces of type K3*, Izv. Akad. Nauk SSSR Ser. Mat., **35**, (1971) 530–572. English transl. in Math. USSR Izv., **5**, (1971) 547–588.
13. Saint-Donat, B., *Projective models of K–3 surfaces*, Amer. J. of Mathem., **96**, (1974) 602–639.
14. Shafarevich, I. R. (ed.) *Algebraic surfaces*, Trudy Matem. Inst. Steklov, T. 75, 1965. English transl. in Proc. Stekov Inst. Math., **75** (1965).
15. Shokurov, V. V., *The Noether–Enriques theorem on canonical curves*, Mat. Sbornik, **86**, (1971) 367–408. English transl. in Math. USSR-Sb. **15**, (1971) 361–404.
16. Tyurin, A. N., *The intersection of quadrics*, Uspekhi Mat. Nauk, **30**, (1975) 51–99. English transl. in Russian Math. Surveys, **30** (1975) 51–105.
17. Tyurin, A. N., *Special* 0*-cycles on a polarized K3 surface* Izv. Akad. Nauk SSSR Ser. Mat., **51**, (1987) 131–151. English transl. in Math. USSR Izv., **30**, (1988) 123–143.
18. Tyurin, A. N., *Cycles, curves and vector bundles on algebraic surfaces*, Duke Math. J., **54**, (1987) 1–26.
19. Tyurin, A. N., *Symplectic structures on the varieties of moduli of vector bundles on algebraic surfaces with* $p_g > 0$, Izv. Akad. Nauk SSSR Ser. Mat., **52**, (1988) 139–178. English transl. in Math. USSR Izv., **33**, (1989) 139–177.

Dipartimento di Matematica, Università degli studi di Ferrara, Italia
E-mail address: madonna@mat.uniroma3.it

Dept. of Pure Math., The University of Liverpool, Liverpool L69 3BX, UK

Steklov Mathematical Institute, ul. Gubkina 8, Moscow 117966, GSP-1, Russia
E-mail address: vnikulin@liv.ac.uk slava@nikulin.mian.su

Clay Mathematics Proceedings
Volume 3, 2004

Contractions and monodromy in homological mirror symmetry

Balázs Szendrői

Abstract. This paper discusses the mirror correspondence between contractions and degenerations of Calabi–Yau varieties, originally due to Morrison, in the light of recent developments. In homological mirror symmetry, degenerations lead to symplectomorphisms, whereas contractions give rise to Fourier–Mukai functors. Several explicit examples are treated, many of them conjectural.

Introduction

Kontsevich's Homological Mirror Symmetry conjecture [**14**] connects holomorphic and symplectic geometry in a deep and surprising way. It relates a pair of Calabi–Yau varieties, one with fixed holomorphic (complex) structure, one with fixed symplectic structure, and predicts an equivalence of two very different kinds of categories. On the symplectic side, "the (derived) Fukaya category" has gone through many transfigurations over the years, and there is still no unique definition. On the holomorphic side, the category is well known: it is the derived category of coherent sheaves. The conjectured equivalence between these two categories implies that there should be a correspondence between their symmetries. The obvious symmetries are symplectic, respectively holomorphic automorphisms; however, typically there are many more of the former than the latter. The idea of the Fourier–Mukai functor [**19**] comes to the rescue: the derived category of a Calabi–Yau variety possesses more symmetries than just the obvious ones, just by virtue of the fact that its canonical bundle is globally trivial.

The purpose of this note is to review the dictionary between symplectic isomorphisms and Fourier–Mukai functors in a pedagogical way. Versions of a global correspondence (at least on the cohomological level) are discussed for example in [**3, 10, 30**]; a comprehensive account will presumably be attempted in [**7**]. My aim here is rather more limited, summarizing and slightly extending the range of constructions that are local to singularities of Calabi–Yau varieties. I will discuss surface double points, threefold nodes, isolated threefold singularities and finally curves of singularities on threefolds, proving no new results but posing several open

2000 *Mathematics Subject Classification.* 14J32, 14E15, 18E30.

problems. The presentation is at the level of suggestive analogies, relying on the shape of cohomology actions, relations, toric examples and the like.

The real question this note completely disregards is of course what exactly *is* mirror symmetry for singularities. Recent ideas of Kontsevich, Kapustin–Li [**13**] and Orlov [**20**] on the one hand, related on the physics side to D-branes in Landau–Ginzburg models, and Seidel [**24, 25**] on categories defined by vanishing cycles on the other, will provide the tools to ask and eventually answer this question in much more precise detail than attempted here.

I thank Mark Gross for pointing out the relation between toric Gorenstein singularities, polytopes and degenerations.

1. Homological mirror symmetry in a nutshell

1.1. I shall make no attempt to give a general overview of mirror symmetry. There are several good sources available; most relevant for this note is Kontsevich's 1994 ICM address [**14**] and its 2000 ECM update by Manin [**16**], both with extensive bibliography. As originally formulated in string theory, mirror symmetry relates two Calabi–Yau manifolds (X, ω_X) and (Y, ω_Y), both equipped with complex structures and compatible symplectic (Kähler) forms, and certain structures defined on (families of) them. In fact, at least deep inside the moduli space, mirror symmetry "decouples" the symplectic and the complex structures, and interchanges one with the other. Kontsevich conjectured that eventually all mirror symmetry constructions should be understood as an equivalence of categories depending on symplectic, respectively holomorphic data:

$$m\colon D^b\mathrm{Fuk}(Y,\omega) \longrightarrow D^b(X).$$

Here $D^b\mathrm{Fuk}(Y,\omega)$ is the "derived Fukaya category". It should be constructed purely and functorially in symplectic terms, using Lagrangian submanifolds of (Y,ω) (now simply thought of as a symplectic manifold, with no complex structure), and their Floer homology. The analytic details are highly non-trivial, and in fact to date not completely settled. On the right hand side, $D^b(X)$ is the bounded derived category of coherent sheaves on the smooth Calabi–Yau variety X. This is a perfectly well-defined and relatively well-known (triangulated) category; see [**9**] for a detailed introduction.

Irrespective of the details of its definition, functoriality of the Fukaya category should imply that it carries an action of the group of symplectomorphisms $\mathrm{Symp}(Y,\omega)$. Symplectomorphisms symplectically homotopic to the identity (in the C^∞-topology) are expected to act trivially. The mirror symmetry equivalence m should then give rise to a map

$$\mu\colon \pi_0(\mathrm{Symp}(Y,\omega)) \to \mathrm{AutEq}(D^b(X))/\langle[1]\rangle, \tag{1}$$

which can be studied independently of the intricacies of the Fukaya category. The group of self-equivalences $\mathrm{AutEq}(D^b(X))$ of the derived category (together with its triangulated structure) contains [1], the translation functor, which is expected to correspond to the translation functor in the derived Fukaya category. This corresponds to the difference between "ordinary" and "graded" symplectic geometry, as explained in [**26**, Introduction].

1.2. The previous discussion implies that we need a handy supply of self-equivalences of the derived category. As I already discussed in the introduction, there are two obvious sources: automorphisms of the complex manifold X, as well as line bundles acting by tensor product. However, one soon realizes that this is not enough. For example, a generic quintic has no automorphisms other than the identity, and only a $\mathbb{Z}$ worth of line bundles. Its mirror in turn has many more symplectomorphisms.

A fundamental construction of Mukai [**19**] comes to the rescue. By analogy with the theory of classical correspondences, an object $\mathcal{F} \in D^b(X \times X)$ defines a functor

$$\Phi^{\mathcal{F}}\colon D^b(X) \to D^b(X)$$

by

$$\Phi^{\mathcal{F}}(-) = \mathbf{R}p_{1*}(\mathcal{F} \overset{\mathbf{L}}{\otimes} p_2^*(-)),$$

where $p_i\colon X \times X \to X$ are the two projections, and $\overset{\mathbf{L}}{\otimes}$, $\mathbf{R}p_{1*}$ denote operations in the derived category. It often happens that, under appropriate conditions, $\Phi^{\mathcal{F}}$ is a self-equivalence of the triangulated category $D^b(X)$, and then it is called a *Fourier–Mukai functor* [**19**].

As a special case, this definition includes the example of a self-equivalence defined by an automorphism; here $\mathcal{F}$ is (the structure sheaf of) the graph of the automorphism. The self-equivalence defined by tensoring with a line bundle $\mathcal{L}$ is realized by $\mathcal{F} = \mathcal{O}_\Delta(\mathcal{L})$ where $\Delta \in X \times X$ is the diagonal. However, the main virtue of this framework is that in the Calabi–Yau context, there are many more Fourier–Mukai functors. Several examples will be given below.

1.3. The mirror symmetry map m is expected to be compatible with a cohomology isomorphism

$$\bar{m}\colon H^{\text{middle}}(Y, \mathbb{Q}) \longrightarrow H^{\text{even}}(X, \mathbb{Q})$$

under, on the symplectic side, the map taking a Lagrangian submanifold to its cohomology class, and on the complex side, the map taking a sheaf or complex to its K-theory class and then via the Chern character to cohomology. For technical reasons, one uses a slight modification of the Chern character, the Mukai map $\mathrm{ch}(-)\sqrt{\mathrm{Td_X}}$, to pass from K-theory to cohomology, which makes the map compatible with natural bilinear pairings. This plays no role in this note.

In fact, here I would like to invoke the "symmetry" of mirror symmetry, which says that actually the complex geometry of Y is also expected to be equivalent to the symplectic geometry of (X, ω_X) in the sense of a categorical equivalence. For the examples studied here (K3s, Calabi–Yau threefolds, but also for elliptic curves) this means that one actually expects a full isomorphism

$$\bar{m}\colon H^*(Y, \mathbb{Q}) \longrightarrow H^*(X, \mathbb{Q}) \tag{2}$$

between $\mathbb{Q}$-vector spaces.

A symplectomorphism of (Y, ω) acts on homology by pushforward. In the context of cohomology actions, Y will always be a compact manifold; hence I can use Poincaré duality to think of cohomology as covariant for symplectomorphisms. A Fourier–Mukai functor $\Phi^{\mathcal{F}}\colon D^b(X) \to D^b(X)$ acts on the cohomology of X via the cohomological correspondence induced by the Chern class of the complex $\mathcal{F}$. One expects the correspondence μ of (1) between symplectomorphisms and Fourier–Mukai functors to be compatible with cohomology actions via the isomorphism $\bar{m}$ of (2).

One specific issue in the Calabi–Yau context is that cupping with the holomorphic n-form induces an isomorphism, well-defined up to constant, $H^1(X,T_X) \cong H^1(X,\Omega_X^{n-1})$. Via Hodge theory, the latter is a direct summand of cohomology. The map $\varphi^{\mathcal{F}}$ preserves Hodge structures, so one obtains an action of a Fourier–Mukai functor on $H^1(X,T_X)$. This is the tangent space to deformations of X, and the interpretation of this action is that the Fourier–Mukai self-equivalence only deforms as a self-equivalence along deformation directions fixed by its cohomology action. This issue is spelled out in [**31**, Theorem 2.1].

2. Degenerations and contractions

2.1. The specific issue considered here is the relation of symplectomorphisms and derived equivalences arising from mirror symmetry between degenerations and contractions [**17**]. Fix some projective ambient space $\mathbb{P}$, and consider a family $\mathcal{Y} \subset T \times \mathbb{P}$ with projection $\pi\colon \mathcal{Y} \to T$ over a smooth base T, having smooth total space $\mathcal{Y}$ and smooth Calabi–Yau fibres over an open set $T^0 \subset T$ whose complement is a divisor with normal crossings. Assume also that $\mathbb{P}$ is equipped with a symplectic form Ω, compatible with its complex structure, whose restriction ω_s to every smooth fibre Y_s of π is still non-degenerate. The map π then becomes a Lefschetz fibration [**23**]. In particular, there is a notion of symplectic parallel transport over T^0. Fixing a base point $t \in T^0$ with fibre $(Y,\omega) = (Y_t,\omega_t)$, there is a map

$$\sigma : \pi_1(T^0) \to \pi_0(\mathrm{Symp}(Y,\omega)),$$

the *symplectic monodromy* of the family π.

As always, if the fibre (Y_s,ω_s) is smooth for all $s \in T$, then the symplectic monodromy is trivial. The interesting case is when Y_s is singular for $s \in T \setminus T^0$, in other words when π represents a degeneration of the complex structure on the Calabi–Yau manifold Y_t, and $T \setminus T^0$ lies in the boundary of the space of complex structures. There is a vast literature on degenerations of Calabi–Yau manifolds; I will describe some specific situations below.

As mirror symmetry interchanges complex and Kähler moduli, the construction mirror to degenerations of the complex structure should be a degeneration of the Kähler structure of the mirror X of Y. The way this can be thought of as monodromy in an actual moduli space around some boundary should not concern us here; compare [**18**] and the much more advanced approach [**5**] for details. I will take the old-fashioned view [**17**] which says that in many cases, Kähler degeneration means that a collection of Kähler classes $[\omega_t]$ on the (complex) Calabi–Yau manifold X tends to some limit class $[\omega_0]$ on the boundary of the *Kähler cone* [**33**] of the complex Calabi–Yau manifold X, the cone of Kähler classes on X in the vector space $H^2(X,\mathbb{R})$. Such boundary points, in favorable cases, correspond to *contractions* on the complex manifold X: algebraic morphisms $X \to \bar{X}$ satisfying certain conditions [**33**].

We then get our basic correspondence: the map μ arising from mirror symmetry considerations should in certain situations relate symplectic monodromy around boundary points of the complex moduli space of Y to Fourier–Mukai transforms defined from contractions on the complex Calabi–Yau manifold X. I will further restrict attention to cases where the contraction $X \to \bar{X}$ is *birational*, contracting a locus $E \subset X$ to a locus $C \subset \bar{X}$ and inducing an isomorphism between non-empty open subsets $X \setminus E \cong \bar{X} \setminus C$. As such contractions are more easily classified, at least into broad classes, I will always start from the contraction and discuss the corresponding degeneration. I will attempt to be rigorous in notation: X will be a Calabi–Yau threefold (with fixed complex structure) and $X \to \bar{X}$ a birational contraction. On the mirror side, Y_0 will always be the degenerate complex manifold (possibly local) with a symplectic smoothing (Y, ω) and, if needed, a resolution $\widehat{Y}_0 \to Y_0$.

2.2. Note that the above discussion assumed a number of details about the specific form of the degeneration of the complex structure of Y. Just because the Calabi–Yau variety Y is embedded in some ambient space $\mathbb{P}$ (such as a weighted projective space or a toric variety), it is by no means certain that all deformations of Y can also be embedded in $\mathbb{P}$ or that all symplectic forms on Y arise as restrictions of some form Ω from $\mathbb{P}$. Hence for a general discussion of the relation between diffeomorphisms and Fourier–Mukai functors, one needs to treat both symplectic structure and complex structure in families; compare [**30**]. However, for my present purposes the above considerations suffice.

3. Surface double points

In dimension two, there is only one class of birational contractions on Calabi–Yau varieties: the contraction of a (-2)-curve on a K3 surface, or more generally the contraction of a tree of such curves. I begin by discussing the case of a single curve.

3.1. Let X be a smooth K3 surface containing a rational curve $\mathbb{P}^1 \cong E \subset X$ of square (-2). There is a contraction $X \to \bar{X}$ contracting this curve to a point $P \in \bar{X}$, which is locally analytically isomorphic to the surface node (simple double point)

$$\{x_1^2 + x_2^2 + x_3^2 = 0\} \subset \mathbb{A}^3.$$

The mirror degeneration to this contraction is known to be a degeneration to the same type of singularity: this means a one-dimensional family of K3 surfaces $\widetilde{\pi}\colon \widetilde{\mathcal{Y}} \to \Delta$ with smooth fibres over the punctured disc Δ^* and a singular fibre Y_0 containing a single double point. Locally analytically the family $\widetilde{\mathcal{Y}}$ is isomorphic to the family

$$\{x_1^2 + x_2^2 + x_3^2 = t\} \subset \mathbb{A}^4$$

where t is the coordinate in the base Δ and the symplectic form ω_t on an open set of Y_t is given by the restriction of the standard form on $\mathbb{C}^3$.

Fix $t \neq 0$ as the base point and assume for simplicity that t is positive real. Then $Y = Y_t$ contains a two-sphere $S^2 \simeq S \subset Y_t$ given by

$$S = \{x_1^2 + x_2^2 + x_3^2 = t \mid x_i \in \mathbb{R}\}.$$

In fact this sphere is Lagrangian in (Y, ω), and it represents the *vanishing cycle*: its homology class generates the subspace of homology disappearing under passage from Y to Y_0. The symplectic monodromy of the family $\pi\colon \mathcal{Y} \to \Delta^*$ was constructed explicitly with the help of the vanishing cycle S by Arnold [**2**] as follows.

Consider first a model situation: the cotangent bundle T^*S^2 with its canonical symplectic form η. By means of the standard metric, identify T^*S^2 with the tangent bundle TS^2. The latter has a circle action σ, defined by the normalized geodesic flow, transporting a tangent vector ξ with unit speed along the geodesic emanating from it, irrespective of its length; regard this as a circle action on T^*S^2. Consider also an auxiliary smooth real function ψ such that $\psi(t)+\psi(-t)=2\pi$ for all $t\in\mathbb{R}$, and $\psi(t)=0$ for $t>>0$. Now define $\tau\colon T^*S^2\to T^*S^2$ by

$$\tau(\xi)=\begin{cases}\sigma(e^{i\psi(|\xi|)})(\xi) & \xi\in T^*S^2\setminus S^2\\ A(\xi) & \xi\in S^2\end{cases}$$

where A is the antipodal map on the sphere. It is easy to check that τ is continuous, and acts trivially away from a small neighbourhood of the zero-section. A short argument also shows that it is a symplectomorphism of (T^*S^2,η) which, up to symplectic isotopy, is independent of the choice of ψ. τ is the *model Dehn twist* of T^*S^2 with respect to its zero section.

If now (Y,ω) is a symplectic manifold containing a Lagrangian two-sphere S, then by a theorem of Weinstein a neighbourhood of this two-sphere can be identified with a neighbourhood of the zero-section in (T^*S^2,η). Since the model Dehn twist of T^*S^2 acts trivially outside a neighbourhood of the zero-section which can be made arbitrarily small, there is a symplectomorphism τ_S of Y defined by the model Dehn twist in a suitable neighbourhood of S.

Proposition 3.1. (Arnold) *The symplectic monodromy of the family $\mathcal{Y}\to\Delta$ deforming the surface node is generated by the Dehn twist $[\tau_S]\in\pi_0(\mathrm{Symp}(Y,\omega))$ in the vanishing cycle S. The cohomology action of $[\tau_S]$ is the map $(\tau_S)_*\colon H^*(Y,\mathbb{Q})\to H^*(Y,\mathbb{Q})$ given by*

$$(\tau_S)_*(\alpha)=\alpha+([S]\cdot\alpha)[S]$$

which is a reflection since $[S]^2=-2$.

What is then the mirror of this symplectic monodromy transformation under the map μ of (1)? According to our basic principle, it should be a Fourier–Mukai functor associated to the contraction of E in X.

Proposition 3.2. (after Seidel–Thomas) *Consider the structure sheaf $\mathcal{F}=\mathcal{O}_{X\times_{\bar X}X}$ of the correspondence $X\times_{\bar X}X\subset X\times X$. This sheaf defines a Fourier–Mukai equivalence*

$$\Phi^{\mathcal{F}}\colon D^b(X)\to D^b(X)$$

with cohomology action $\varphi^{\mathcal{F}}\colon H^*(X,\mathbb{Q})\to H^*(X,\mathbb{Q})$ given by

$$\varphi^{\mathcal{F}}(\alpha)=\alpha+([E]\cdot\alpha)[E]$$

which is a reflection since $[E]^2=-2$.

A plausible guess, which under a suitable choice of mirror map can indeed be made precise, is that the mirror $\mu([\tau_S])$ of the symplectic monodromy $[\tau_S]$ is the functor $\Phi^{\mathcal{F}}$. Their cohomology actions are indeed easy to match. I wrote the sheaf $\mathcal{F}$ in the above form to show its explicit dependence on the contraction, but in fact it can be shown that the above definition is equivalent to the original definition of [**26**] as a twist functor. Without going into the details, which can be found in [**26**], I note that this particular transform is the (inverse) twist on X defined by the spherical sheaf $\mathcal{O}_E(-1)$.

3.2. The story for simple nodes can be generalized to other double point singularities. In the local situation, mirror symmetry relates deformations and smoothings of the arbitrary A_n surface singularity

$$\{x_1^2+x_2^2+x_3^{n+1}=0\}\subset\mathbb{A}^3.$$

As shown in [**22**], a smoothing Y of this singularity contains a collection $S_1, \ldots, S_n$ of Lagrangian spheres, with a single transversal intersection point between S_i and S_{i+1} and no other intersections. There are Dehn twists in all these spheres, which satisfy the relations of the braid group on $n+1$ strings [**22**]. On the contraction side, X contains holomorphic spheres (rational curves) $E_1, \ldots, E_n$, and has a corresponding collection of derived self-equivalences which act by the braid group. The actions on cohomology are given on both sides by reflections generating the symmetric (Weyl) group. For details, consult [**26**].

4. Isolated threefold singularities

I now turn to birational contractions of three-dimensional Calabi–Yau varieties. There are now three classes of possibilities, depending on the dimension of the exceptional locus $E \subset X$ and its image. In this section I consider cases where this image is a point.

4.1. The first possibility to consider is when E is one-dimensional, and hence necessarily rational. The simplest case is when $E \cong \mathbb{P}^1$ with normal bundle $N_{E/X} \cong \mathcal{O}_{\mathbb{P}^1}(-1,-1)$, the $(-1,-1)$-curve. The contraction of such a curve leads to a threefold $\bar{X}$ with an ordinary double point singularity

$$\{x_1^2 + x_2^2 + x_3^2 + x_4^2 = 0\} \subset \mathbb{A}^4.$$

This contraction is the first half of the conifold transition in physics, which is considered to be self-mirror [**17**]. Hence its degeneration mirror should be the other half of the conifold transition, the smoothing $\mathcal{Y} \to \Delta^*$ of the node locally given by

$$\{x_1^2 + x_2^2 + x_3^2 + x_4^2 = t\} \subset \mathbb{A}^5.$$

For fixed t positive real, this local geometry contains a Lagrangian three-sphere given by $x_i \in \mathbb{R}$, which is again the vanishing cycle and is Lagrangian. Globally we have a symplectic Calabi–Yau threefold (Y, ω) containing a Lagrangian S^3, and the situation is analogous to the surface case:

Proposition 4.1. (Arnold) *The symplectic monodromy of the smoothing of the threefold node is given by a Dehn twist $[\tau_S]$ in the vanishing cycle $S \simeq S^3$ defined exactly as in the surface case. Its cohomology action is*

$$(\tau_S)_*(\alpha) = \alpha + ([S]\cdot\alpha)[S];$$

as $[S]^2 = 0$, this is a map of infinite order.

One plausible mirror of this symplectic monodromy map is given by

Proposition 4.2. (Seidel–Thomas) *Consider a $(-1,-1)$-curve E in a Calabi–Yau manifold X. Then its structure sheaf $\mathcal{O}_E$ defines a spherical sheaf, and there is a corresponding twist functor*

$$T_{\mathcal{O}_E} \in \mathrm{AutEq}(D^b(X))$$

with cohomology action given by

$$\alpha \mapsto \alpha + (\mathrm{ch}(\mathcal{O}_E)\cdot\alpha)\mathrm{ch}(\mathcal{O}_E).$$

One notable feature of this pair of symmetries is their cohomology action. As the formulae show, $(\tau_S)_*$ acts trivially on even cohomology as $[S]$ is in degree three; mirror to this, $T_{\mathcal{O}_E}$ acts trivially on odd cohomology as $\mathrm{ch}(\mathcal{O}_E)$ lives in even degree. As mentioned in Section 1, this implies that $T_{\mathcal{O}_E}$ acts trivially on the deformation space of X. Thus this Fourier–Mukai functor must exist on all (small) deformations of X; indeed, contractions of $(-1,-1)$-curves are known to have this property [**33**].

I leave the world of the simple node, though of course there would be much more to say regarding the issue of multiple nodes and flops. I leave these to more able hands; [**17, 26**] and [**27**] respectively have all the details.

4.2. The mirror pairs considered so far can be encoded in toric geometry, which suggests generalizations. Namely, consider a convex polytope Π with vertices in a lattice $\mathbb{Z}^{n-1}$. Embed $\mathbb{Z}^{n-1} \subset \mathbb{Z}^n$ as the affine hyperplane with last coordinate 1. Let N be the sublattice of $\mathbb{Z}^n$ spanned by the origin and the vertices of Π, and let τ be the cone over Π. Then the toric variety $\bar{X} = \mathbb{X}_{N,\tau}$ (in the covariant description [**8**]) is an affine n-fold with a canonical Gorenstein singularity at the origin, which in dimension at most three has toric crepant resolution(s) $X \to \bar{X}$.

On the other hand, the polytope Π can be thought of as the Newton polytope of a polynomial f in variables x_i, x_i^{-1} for $i = 1, \ldots, n-1$. Adding two auxiliary variables u, v, one obtains a family of affine n-folds

$$\mathcal{Y} = \{u^2 + v^2 + f(x_i, x_i^{-1}) = 0\} \subset (\mathbb{A} \setminus \{0\})^{n-1} \times \mathbb{A}^2 \times T$$

over the base T defined by the moduli of the polynomial f. The complex structure of this variety will degenerate for particular f, giving rise to a boundary locus $T \setminus T^0$ and symplectic monodromy. Toric geometry suggests that this family is the mirror of the contraction $X \to \bar{X} = \mathbb{X}_{N,\tau}$.

It is a simple matter to see that the surface A_n singularity is the case when $P = [0, n+1]$ is 1-dimensional, whereas the threefold node arises from the two-dimensional polytope $P = [0,1] \times [0,1]$.

4.3. Examples of birational maps contracting a surface $E \subset X$ to a point also arise in this way. For example, consider the polytope in $\mathbb{R}^2$ with vertices at $(-1,-1), (1,0)$ and $(0,1)$. The contraction $X \to \bar{X} = \mathbb{X}_{N,\tau}$ contracts a projective plane $\mathbb{P}^2 \subset X$ to a Gorenstein singularity $P \in \bar{X}$. The mirror of this singularity is the family of threefolds

$$\mathcal{Y} = \{a_0 + a_1 x_1 + a_2 x_2 + a_3 x_1^{-1} x_2^{-1} + u^2 + v^2 = 0\} \subset (\mathbb{A} \setminus \{0\})^2 \times \mathbb{A}^2 \times T$$

over a base $T \subset \mathbb{A}^4$ with coordinates a_i. The symplectic geometry of this family is studied in detail in [**25**, Proposition 3.2].

4.4. The list can be continued with various two-dimensional polytopes. One interesting example is the following: consider a Calabi–Yau threefold $\widehat{Y}_0$ containing a contractible surface $E \subset \widehat{Y}_0$ abstractly isomorphic to $\mathbb{P}^2$ blown up in three points. Then $E^3 = (K_E)^2 = 6$ and after contraction, Y_0 contains a toric Gorenstein singularity $(P \in Y_0)$ corresponding to the polytope Π with vertices at $(-1,-1)$, $(0,-1)$, $(1,0)$, $(1,1)$, $(0,1)$ and $(-1,0)$. The mirror to this singularity is the family

$$\mathcal{X} = \{u^2 + v^2 + f(x_i, x_i^{-1}) = 0\} \subset \mathbb{A}^2 \times (\mathbb{A} \setminus \{0\})^2 \times T$$

defined by polynomials f with Newton polytope Π. It might be interesting to study the symplectic geometry of degenerations in this family in more detail.

On the other hand, the deformation theory of $(P \in Y_0)$ is quite interesting: by [**1**, 2.1 and 8.4], the local first-order deformation space is three-dimensional, but only a one- and a two-dimensional subspace can be realized as actual deformations, the two components corresponding to two essentially different ways in which Π decomposes into the Minkowski sum of two polytopes Π_1 and Π_2. There are therefore two different (not even diffeomorphic) symplectic smoothings $Y_s^{(j)}$ here. On the other side, as Mark Gross points out, for every decomposition $\Pi = \Pi_1 + \Pi_2$ into Minkowski summands, there are degenerate members

$$\bar{X}^{(j)} = \{u^2 + v^2 + f_1(x_i, x_i^{-1}) f_2(x_i, x_i^{-1}) = 0\} \subset \mathbb{A}^2 \times (\mathbb{A} \setminus \{0\})^2,$$

in the family $\mathcal{X}$, where f_i has Newton polytope Π_i. These varieties are singular, with nodes at the points where $u = v = f_1 = f_2 = 0$, and have small resolutions $X^{(j)} \to \bar{X}^{(j)}$ mirroring the two symplectic smoothings $Y_s^{(j)}$.

4.5. Beyond toric cases, one example discussed in the literature [**28**] concerns, on the symplectic side, the degeneration to the simple threefold triple point

$$\{x_1^3 + x_2^3 + x_3^3 + x_4^3 = 0\} \subset \mathbb{A}^4.$$

Denote as usual by (Y, ω) a symplectic smoothing.

Proposition 4.3. (Smith–Thomas) *The symplectic manifold (Y, ω) contains a collection of 16 Lagrangian three-spheres $S_1, \ldots, S_{16}$ meeting in an intricate configuration depicted on* [**28**, Figure 2]. *In particular, there are 16 corresponding Dehn twists*

$$\tau_{S_i} \in \pi_0(\mathrm{Symp}(Y, \omega)).$$

As far as I know, the mirror of this singularity has not been discussed in the literature. It should have 16 Fourier–Mukai transforms corresponding to the Dehn twists of Proposition 4.3.

5. Non-isolated threefold singularities

The final case left out so far is that of a contraction $(E \subset X) \to (C \subset \bar{X})$ with $\dim C = 1$. Assuming that X is smooth, it follows that the curve C is also smooth [**33**].

5.1. Start with the simplest case: a projective Calabi–Yau variety $\bar{X}$, smooth outside of a smooth curve $C \subset \bar{X}$ of genus g, locally analytically isomorphic to

$$\{x_1^2 + x_2^2 + x_3^2 = 0\} \subset \mathbb{A}^4$$

along C (which is locally the line $\mathbb{A}^1_{x_4}$). Blowing up the ideal of C gives a Calabi–Yau resolution $X \to \bar{X}$ containing an exceptional divisor E geometrically ruled over C.

Proposition 5.1. (Horja, Szendrői) *The structure sheaf $\mathcal{F} = \mathcal{O}_{X \times_{\bar{X}} X}$ on the product $X \times X$ gives a Fourier–Mukai self-equivalence*

$$\Phi^{\mathcal{F}} \colon D^b(X) \to D^b(X).$$

The corresponding cohomology action maps *i. a class* $\alpha \in H^2(X,\mathbb{Q})$ *as*

$$\varphi^{\mathcal{F}}(\alpha) = \alpha + ([l]\cdot\alpha)[E] \mod H^4(X,\mathbb{Q})$$

with $[l] \in H^4(X,\mathbb{Q})$ *the class of the ruling of* E *(this is a reflection since* $[l]\cdot[E] = -2$*);*

ii. on third cohomology as an involution with a codimension $2g$ *fixed locus and* (-1)*-eigenspace given by the image of the cylinder homomorphism* $H^1(C,\mathbb{Q}) \to H^3(X,\mathbb{Q})$ *(note that there is no odd cohomology in different degrees; see* [**29**, Proposition 4.6] *for more details).*

The problem is then clear: find the symplectic mirror of this contraction, together with its symplectic monodromy corresponding to $\Phi^{\mathcal{F}}$.

5.2. Before I move on, I discuss a slight generalization, which will be just as easy (or difficult) to study: suppose that $\bar{X}$ contains a curve of A_n-singularities $C \subset \bar{X}$ locally of the form

$$\{x_1^2 + x_2^2 + x_3^{n+1} = 0\} \subset \mathbb{A}^4.$$

Assume also that repeated blowup leads eventually to a smooth Calabi–Yau resolution X containing a collection $E_1, \ldots, E_n$ of smooth ruled surfaces. Then there is a Fourier–Mukai equivalence $\Phi^{\mathcal{F}_i} \in \mathrm{AutEq}(D^b(X))$ for each of the surfaces, which as a matter of fact satisfy the relations of the braid group [**32**] just as in the surface case.

5.3. To approach the problem of finding the mirror, I recall the discussion of the above setup in the toric context from [**15**, Section 3.2]. Assume therefore that $\bar{X}$ is in fact a hypersurface in a toric variety $\mathbb{P}_\Delta$ given by a reflexive polytope Δ spanned by some lattice points in a lattice $M \cong \mathbb{Z}^4$; compare [**4**]. The singularities of $\bar{X}$ are most easily studied in terms of the normal fan Σ consisting of cones spanned by vertices of the dual polytope Δ° in the dual $N = \mathrm{Hom}_{\mathbb{Z}}(M,\mathbb{Z})$. The specific singularity along C then arises from an edge $\tau = \langle \mathbf{v}_0, \mathbf{v}_{n+1}\rangle$ of Δ° containing n interior lattice points $\mathbf{v}_1, \ldots, \mathbf{v}_n$. The genus g can also be read off from the toric data: it is the number of interior lattice points in the dual two-dimensional face $\tau^\circ \subset \Delta \subset M$.

The mirror Y of X is given, by Batyrev's mirror duality [**4**], as a hypersurface in a partial toric resolution $\widehat{\mathbb{P}}_{\Delta^\circ}$ of the toric variety $\mathbb{P}_{\Delta^\circ}$. The parameter space of the hypersurface Y is the vector space

$$H^0\left(\widehat{\mathbb{P}}_{\Delta^\circ}, -K_{\widehat{\mathbb{P}}_{\Delta^\circ}}\right) \cong \bigoplus_{\mathbf{m}\in\Delta^\circ\cap N} \mathbb{C}\,\mathbf{x}^{\mathbf{m}}.$$

Here $\mathbf{x}^{\mathbf{m}}$ represents a monomial $\prod x_i^{m_i}$, where x_i are to be thought of as affine variables in the affine torus $(\mathbb{C}^*)^4 \subset \widehat{\mathbb{P}}_{\Delta^\circ}$, and Y is the closure in $\widehat{\mathbb{P}}_{\Delta^\circ}$ of the subvariety of the torus given by the chosen Laurent polynomial. By [**6**], this vector space can also be viewed dually as the appropriately graded piece of the homogeneous polynomial ring S of the abstract toric variety $\widehat{\mathbb{P}}_{\Delta^\circ}$, with one generating variable y_η for every lattice point $\eta \in N$ on the boundary of the polytope Δ.

Now return to our concrete situation, restricting attention to $g = 1$, the reasons for which will appear presently; let $\eta_1 \in M$ be the unique lattice point in the interior of the face $\tau^\circ \subset \Delta$ with corresponding homogeneous coordinate y_1. The hypersurface Y is given by a homogeneous equation

$$\{y_1 f(y_j) + g(y_j) = 0\} \subset \widehat{\mathbb{P}}_{\Delta^\circ}$$

where I have separated out the monomials not involving the variable y_1. It can however be checked, using the correspondence between Laurent and homogeneous polynomials [**6**], that the only terms not involving the monomial y_1 correspond to Laurent monomials $\mathbf{x}^{\mathbf{m}}$ with $\mathbf{m} \in \tau$, one of the $n+2$ lattice points responsible for the singularity of $\bar{X}$. The linear relations between these lattice points translate to multiplicative relations between the Laurent monomials, which implies that the equation of Y can be written on a suitable affine piece of $\widehat{\mathbb{P}}_{\Delta^\circ}$ as

$$y_1 f + \sum_{j=0}^{n+1} a_j x^j = 0$$

where x is an auxiliary affine variable. Moreover, the contraction of X back to $\bar{X}$ corresponds to the degeneration to the hypersurface with equation

$$y_1 f + (b_1 x + b_0)^{n+1} = 0$$

which (assuming appropriate regularity of f) is a threefold exactly of the studied type: singular along the curve $\{y_1 = f = b_1 x + b_0 = 0\}$, with a curve of A_{n+1} singularities.

This is then the correspondence suggested by toric geometry: the mirror to a contraction of a collection of ruled surfaces to an elliptic curve of A_n singularities should be a degeneration to a single curve of A_n singularities.

5.4. Before moving on to more theory, it might be illustrative to give an example, appearing as [**12**, Example II]. Let

$$\widetilde{\Delta} = \{\mathbf{m} \in M \cong \mathbb{Z}^4 \colon m_i \geq -1, 1 \geq m_1 + 2m_2 + 2m_3 + 2m_4\}$$

with dual polytope $\widetilde{\Delta}^\circ$ spanned by $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3, \mathbf{e}_4$ and $\mathbf{v}_0 = -\mathbf{e}_1 - 2\mathbf{e}_2 - 2\mathbf{e}_3 - 2\mathbf{e}_4$ in $\widetilde{N} = \oplus \mathbb{Z}\mathbf{e}_i$. The toric variety $\mathbb{P}_{\widetilde{\Delta}}$ is simply weighted projective space $\mathbb{P}^4[1^2, 2^3]$, containing the family of octic Calabi–Yau hypersurfaces. The edge $\langle \mathbf{v}_0, \mathbf{e}_1 \rangle$ contains one interior lattice point, giving rise to a curve of A_1 singularities in the general octic. This curve has genus 3.

Now consider a quotient of this family; let

$$N = \widetilde{N} + \left(\frac{1}{4}, \frac{1}{4}, 0, \frac{1}{2}\right)\mathbb{Z} + \left(0, 0, \frac{1}{4}, \frac{3}{4}\right)\mathbb{Z}$$

with $\Delta^\circ = \widetilde{\Delta}^\circ$ now thought of as a polytope in the lattice N. It can be checked that Δ° is still reflexive, and now $\langle \mathbf{v}_0, \mathbf{e}_1 \rangle$ contains seven interior lattice points, indicating singularities of type A_7 along a curve. The dual to (N, Δ°) is the pair (M, Δ), where

$$M = \widetilde{M} + \left(0, \frac{1}{2}, \frac{1}{4}, \frac{1}{4}\right)\mathbb{Z}$$

and $\Delta = \widetilde{\Delta}$ as a polytope with vertices in M. The mirror two-dimensional face to $\langle \mathbf{v}_0, \mathbf{e}_1 \rangle$ looks like the toric diagram for a $\frac{1}{4}(1,1,2)$ singularity, with one interior lattice point. Hence the generic threefold $\bar{X} \subset \mathbb{P}_\Delta$ has a genus-1 curve of A_7 singularities, so falls under the rubric of the above discussion.

5.5. I now consider the symplectic geometry of the mirror (Y, ω) smoothing the hypersurface Y_0. The monodromy of the degeneration found above has in fact been studied already by Seidel [**22**].

Definition/Proposition 5.2. (Seidel) *Let* (M^{2n}, ω) *be a symplectic manifold. A* set of construction data for a generalized Dehn twist on M *consists of a compact symplectic manifold* (M', ω') *of dimension* $2n-2r$*, a fibre bundle* $p\colon S \to M'$ *with fibre* S^r *and structure group* $O(r+1)$*, and an embedding* $i\colon S \hookrightarrow M$ *such that* $i^*(\omega) = p^*(\omega')$*. Given a set of such data, there is a well-defined* generalized Dehn twist

$$[\tau_S] \in \pi_0(\mathrm{Symp}(M, \omega)).$$

Dehn twists as discussed in previous sections correspond to the case $M' = \text{point}$. The general idea should now be clear:

Conjecture 5.3. *Let* $X \to \bar{X}$ *be a contraction of ruled surfaces* $E_1, \ldots, E_n$ *to an elliptic curve* $C \subset \bar{X}$ *of* A_n*-singularities on a Calabi–Yau threefold. Then the symplectic mirror* (Y, ω) *of* X *contains a collection* $p_i\colon S_i \to \Sigma$ *of* S^2*-bundles over a symplectic Riemann surface* (Σ, ω')*, providing a set of data for generalized Dehn twists*

$$[\tau_{S_i}] \in \pi_0(\mathrm{Symp}(Y, \omega)).$$

These symplectomorphisms satisfy the relations of the braid group on $n+1$ *strings in the obvious way. Their cohomology actions are given for* $\alpha \in H^2(Y, \mathbb{Q})$ *by*

$$(\tau_{S_i})_*(\alpha) = \alpha + ([l_i] \cdot \alpha)[S_i]$$

with $[l_i] \in H^4(Y, \mathbb{Q})$ *the class of the fibre of* p_i*;* $(\tau_{S_i})_*$ *is a reflection since* $[l_i] \cdot [S_i] = -2$*. These symplectic automorphisms mirror the action of the Fourier–Mukai transforms of Proposition 5.1.*

Note that the cohomology action is formally identical to that of Proposition 5.1.*i*. However, remembering that we are in the threefold case, the more important point is the relation to the action in Proposition 5.1.*ii*, which brings me back to the $g = 1$ issue. Note that in Proposition 5.1.*i*, the fixed locus is of codimension $2g$. Hence for a simple formula like that in Conjecture 5.3 to hold, with a codimension-one fixed locus, we need to be in the case $g = 1$ (the other half of the fixed locus in the symplectic case is in H^4 by Poincaré duality).

Geometrically, it is abundantly clear where the fibred manifolds S_i come from, if the sketched degeneration picture is correct. The base Σ of the fibration is the singular locus of the degenerate variety Y_0. The S^2 sphere fibres are the vanishing cycles in the local A_n degeneration transverse to that curve.

5.6. In the toric argument, the $g = 1$ assumption manifested itself in the presence of a single coordinate y_1 which can be pulled out of the equation of Y. In the case of higher genus, there are several such coordinates, and the picture suggested by [**15**] is that of a degeneration to a reducible curve of A_n singularities. These curves will however begin to intersect, necessarily producing singularities worse than A_n, and the geometric picture is not so clear any more. The discussion of the cohomology action also suggests a more complicated symplectic automorphism. It might be worthwhile to study this case in more detail.

Incidentally, the $g = 0$ case is also of some interest. In that case the cohomology action of the Fourier–Mukai functors of Proposition 5.1 in odd degree is trivial. Mirror to that, I expect a symplectomorphism induced by a submanifold $S \subset (Y, \omega)$ generically fibred in spheres by $p\colon S \to \Sigma$, so that the spheres collapse over special points $P \in \Sigma$ making their cohomology class trivial.

5.7. Return to the example of 5.4. This is discussed in [**12**] as Example II, where a pair of (special) Lagrangian cycles $S^1 \times S^2 \simeq N_i \subset (Y, \omega)$ is constructed as the fixed locus of a real involution. If Conjecture 5.3 holds, the natural guess is that in fact $N_i = p_4^{-1}(B_i)$ for a pair of Lagrangian circles $S^1 \simeq B_i \subset (\Sigma, \omega')$; remember that the singularity is of type A_7 and the construction using real variables in the complex A_7 equation gives the middle vanishing 2-cycle. The one-dimensional local moduli space of N_i as a special Lagrangian cycle is geometrically realized then as coming simply from moving the circle B_i locally in Σ. Conjecture 5.3 would imply that there is a host of other Lagrangian cycles around, though their realization as special Lagrangians is bound to run into the usual problem of finding sLag representatives of vanishing cycles.

As for the other two examples of [**12**], Example I involves the original octic as X, with a contraction to a $g = 3$ curve. Its mirror contains a complicated (special) Lagrangian cycle N with $b_2(N) = 5$. As I discussed above, I expect this case to be quite complicated. [**12**, Example III] involves on the complex side a contraction to a $g = 0$ curve, and again $S^1 \times S^2$ in the mirror; this would arise naturally from the $g = 0$ speculation at the end of 5.6.

5.8. To conclude, I want to return to one point which was swept under the carpet above. Namely, just because a threefold has a curve of A_n singularities, it does not follow that in its resolution one finds n irreducible surfaces all ruled over the same curve. This is an issue of monodromy (in a different sense now, over the curve C), which is discussed in detail in [**31, 32**]. I only want to point out that [**31**, Example 4.3] constructs an example of a threefold $\bar{X}$ with an elliptic curve of A_3 singularities, where in the resolution there are only two irreducible surfaces $E_1, E_2 \subset X$. There are two corresponding Fourier–Mukai functors $\Phi_1, \Phi_2 \in \mathrm{AutEq}(D^b(X))$, which satisfy the braid relation

$$\Phi_1 \circ \Phi_2 \circ \Phi_1 \circ \Phi_2 = \Phi_2 \circ \Phi_1 \circ \Phi_2 \circ \Phi_1$$

of the C_3 braid group. The analogue of Conjecture 5.3 would suggest that the mirror (Y, ω) of X should contain a pair of S^2-fibred submanifolds together with the necessary symplectic data, giving rise to a pair of Dehn twists $[\tau_i] \in \pi_0(\mathrm{Symp}(Y, \omega))$ which satisfy the same relation

$$\tau_1 \circ \tau_2 \circ \tau_1 \circ \tau_2 = \tau_2 \circ \tau_1 \circ \tau_2 \circ \tau_1$$

up to symplectic isotopy. I leave the problem of filling in details as a final challenge for you, my Dear Reader.

References

1. K. Altmann, *The versal deformation of an isolated toric Gorenstein singularity*, Inv. Math. **128** (1997), 443–479.
2. V. Arnold, *Some remarks on symplectic monodromy of Milnor fibrations*, The Floer memorial volume (ed. Hofer et al), 99–103, Progr. Math., 133, Birkhauser, Basel, 1995.
3. P. Aspinwall and M.R. Douglas, *D-brane stability and monodromy*, JHEP 0205 (2002) 031.
4. V.V. Batyrev, *Dual polyhedra and mirror symmetry for Calabi-Yau hypersurfaces in toric varieties*, J. Alg. Geom. **3** (1994) 493–535.
5. T. Bridgeland, *Stability conditions on triangulated categories*, math.AG/0212237.
6. D.A. Cox, *The homogeneous coordinate ring of a toric variety*, J. Alg. Geom. **4** (1995) 17-50.
7. R. Dijkgraaf et al (eds), *Mirror Symmetry at the Clay School in Cambridge*, in preparation to be in preparation.
8. W. Fulton, *Introduction to toric varieties*, Princeton University Press, Princeton, 1993.

9. S.I. Gelfand and Y.I. Manin, *Methods of homological algebra*, Springer, Berlin, 1996.
10. P. Horja, *Hypergeometric functions and mirror symmetry in toric varieties*, math.AG/9912109.
11. P. Horja, *Derived category automorphisms from mirror symmetry*, math.AG/0103231.
12. S. Kachru, S. Katz, A. Lawrence and J. McGreevy, *Mirror symmetry for open strings*, Phys. Rev. D (3) 62 (2000).
13. A. Kapustin and Y. Li, *D-Branes in Landau-Ginzburg models and algebraic geometry*, hep-th/0210296.
14. M. Kontsevich, *Homological algebra of mirror symmetry*, in: Proceedings of the International Congress of Mathematicians, (Zurich, 1994, ed. S.D. Chatterji), 120–139, Birkhauser, Basel, 1995.
15. S. Katz, D.R. Morrison and R. Plesser, *Enhanced gauge symmetry in type II string theory*, Nucl. Phys. B 477 (1996), 105–140.
16. Y.I. Manin, *Moduli, motives, mirrors*, in: European Congress of Mathematics, Vol. I (Barcelona, 2000), 53–73, Progr. Math., 201, Birkhuser, Basel, 2001.
17. D.R. Morrison, *Through the looking glass*, Mirror Symmetry III (D. H. Phong, L. Vinet, and S.-T. Yau, eds.) 263-277, American Mathematical Society and International Press 1999.
18. D.R. Morrison, *Compactifications of moduli spaces inspired by mirror symmetry*, Journées de Géométrie Algébrique d'Orsay (Orsay, 1992), Astérisque 218 (1993), 243–271.
19. S. Mukai, *Duality between $D(X)$ and $D(\hat{X})$ with its application to Picard sheaves*, Nagoya Math. J. 81 (1981), 153–175.
20. D. Orlov, *Triangulated categories of singularities and D-branes in Landau–Ginzburg models*, math.AG/0302304.
21. M. Reid, *Canonical* 3*-folds*, Journées de Géométrie Algébrique d'Angers (Angers, 1979, ed. A. Beauville), 273–310, Sijthoff and Noordhoff, Alphen aan den Rijn—Germantown, Md., 1980.
22. P. Seidel, *Lagrangian two-spheres can be symplectically knotted*, J. Diff. Geom. **52** (1999), 145–171.
23. P. Seidel, *Symplectic monodromy*, manuscript 1996.
24. P. Seidel, *Vanishing cycles and mutations*, European Congress of Mathematics, Vol. II (Barcelona, 2000), 65–85, Progr. Math., 202, Birkhuser, Basel, 2001.
25. P. Seidel, *More about vanishing cycles and mutation*, Symplectic geometry and mirror symmetry (Seoul, 2000), 429–465, World Sci. Publishing, River Edge, NJ, 2001.
26. P. Seidel and R. Thomas, *Braid group actions on derived categories of coherent sheaves*, Duke Math. J. 108 (2001), 37–108.
27. I. Smith, R. Thomas and S.-T.Yau, *Symplectic conifold transitions*, J. Diff. Geom. **62** (2002) 209-242.
28. I. Smith and R. Thomas, *Symplectic surgeries from singularities*, Turkish J. Math. **27** (2003) 231-250.
29. B. Szendrői, *Calabi–Yau threefolds with a curve of singularities and counterexamples to the Torelli problem*, Int. J. Math. **11** (2000), 449–459.
30. B. Szendrői, *Diffeomorphisms and families of Fourier–Mukai transforms in mirror symmetry*, in: Applications of algebraic geometry to coding theory, physics and computation (Eilat, 2001), 317–337, NATO Sci. Ser. II Math. Phys. Chem. 36, Kluwer, Dordrecht, 2001.
31. B. Szendrői, *Enhanced gauge symmetry and generalized braid group actions*, Comm. Math. Phys. **238** (2003) 35-51.
32. B. Szendrői, *Artin group actions on derived categories of threefolds*, math.AG/0210121, submitted.
33. P.M.H. Wilson, *The Kähler cone on Calabi–Yau threefolds*, Inv. Math. 107 (1992), 561-583. *Erratum*, Inv. Math. 114 (1993), 231–233.

Department of Mathematics, Utrecht University, PO. Box 80010, NL-3508 TA Utrecht, The Netherlands

E-mail address: szendroi@math.uu.nl

Clay Mathematics Proceedings
Volume 3, 2004

Lectures on Supersymmetric Gauge Theory

Nick Dorey

Abstract. These lectures give an informal introduction to recent developments in supersymmetric (SUSY) gauge theories. In particular, we describe how heuristic ideas about the strongly-coupled vacuum of QCD such as electric-magnetic duality and large N string theory have found a precise realization in the context of SUSY gauge theory.

1. Introduction

Since its discovery almost 30 years ago, supersymmetry (SUSY) has become one of the key theoretical ideas in contemporary high-energy physics. Although there is, as yet, no direct experimental evidence for SUSY there are a number of related reasons why it remains close to theorists' hearts. One general motivation is based on the central role that symmetry principles have played so far in the development of the subject. Given this historical perspective, looking for theories with the largest possible symmetry groups seems like a sensible strategy. Under certain fairly general assumptions, supersymmetry is the only possible extension of the known symmetries of particle physics and this fact alone makes it worthy of study. There are also (at least) three more specific motivations for SUSY. Two of these we will only mention in passing as they will not concern us directly in these lectures:

• Supersymmetry offers a potential solution to the hierarchy problem of the Standard Model. In particular, it provides a natural way of keeping scalar particles light by pairing them with fermions whose masses in turn are protected by (approximate) chiral symmetries. For this reason SUSY is an important ingredient in many possible extensions of the Standard Model.

• Supersymmetry plays a key role in string theory. Indeed, spacetime supersymmetry, in the form of the GSO projection, seems to be necessary for removing the tachyon from the spectrum and ensuring a stable ground state for the string.

Either of these subjects would require a lecture course of its own. And so we come to the main subject of these lectures, which is lessons that supersymmetry can teach us about the dynamics of non-Abelian gauge theories in four dimensions. These gauge theories are, of course, the cornerstone of the standard model. However, despite their long familiarity, we still understand very little about them

2000 *Mathematics Subject Classification.* 81T60, 81T13.

outside the regime of perturbation theory. The problem is particularly pressing in QCD-like theories with asymptotic freedom, which flow to strong coupling in the infrared (IR). In particular, the strongly-coupled dynamics in the IR are responsible for the enduring mystery of quark confinement. After many years of effort, we have yet to find analytic methods which can provide quantitative understanding of the confining phase. In the absence of such methods, progress has been made only by resorting to lattice simulations.

Although QCD itself remains as intractable as ever, theorists have often attempted to study the key physical phenomena such as quark confinement and chiral symmetry breaking in the context of simpler models. For example, in the 1970's there were many theoretical studies of field theories in two spacetime dimensions such as the Schwinger model and the Gross-Neveu model, which have some QCD-like features. Recent progress in this direction has come from the realization that supersymmetric non-Abelian gauge theories in four dimensions provide a much more realistic theoretical laboratory. Supersymmetry has the effect of constraining the form of quantum corrections while still allowing pheneomena of physical interest including confinement. As we will review below, interesting qualitative ideas about QCD and confinement such as electric-magnetic duality and the large N string, which remain heuristic in the physical case, can be implemented in a very precise way in the context of SUSY theories. The reader may justifiably ask what supersymmetry has to do with QCD and the strong interactions. At present there is no satisfactory answer to this question. There is instead the hope that some of the lessons we have learnt, particularly about the physics of the large N limit, will be applicable also in the non-supersymmetric case.

In this introdution, we will briefly review two old ideas about confinement in QCD which have recently taken on a much more quantitative form in the supersymmetric context. The first idea is the possibility of a duality between electric and magnetic charges. This idea has its genesis in the obvious symmetry of Maxwell's equations in the vacuum: $\mathbf{E} \to \mathbf{B}$, $\mathbf{B} \to -\mathbf{E}$. In the real world this symmetry is of course broken by the asymmetry between electric charges which are common place and magnetic charges which are yet to be seen in Nature. Despite this, the existence of magnetic charges remains an important theoretical possibility. As we will discuss below, non-Abelian gauge theories often have a Coulomb phase where the unbroken gauge group includes an Abelian factor. This is true, for example, of many candidate GUT theories which embed the $U(1)$ of electromagnetism inside a simple group. These theories often contain classical solitons which carry the corresponding magnetic charge. At weak coupling, these states are very massive (possibly explaining why we have never seen them). On the other hand, the classical mass formula suggests that they may become light at strong coupling, and many authors have speculated on their potential importance for strong coupling phenomena like confinement.

In any case, non-Abelian gauge theories in their Coulomb phase often have both electric and magnetically charged particles in their spectrum and, in some examples, they appear symmetrically in a way which suggests that the trivial electric-magnetic symmetry of Maxwell's equations in vacuum can be extended to include charges. This 'symmetry' exchanges electrically charged elementary particles with magnetically charged solitons. Because of the Dirac quantization condition (see Section 3 below), it also necessarily exchanges weak coupling, where the electric degrees of freedom are light, with strong coupling, where the magnetic degrees of freedom are light. An equivalence between theories with different values of the coupling constant is known as a *duality*. In some special cases, of which the best understood is $\mathcal{N}=4$ supersymmetric Yang-Mills theory, it is now believed that this electric-magnetic duality is an exact property of the theory. If true, this gives us a precise description of the strong coupling regime which was previously out of reach. Of course, to test such a duality one needs an independent calculation of some quantity in the strongly coupled theory.

To understand the possible relevance of electric-magnetic duality to confinement, we must extend our discussion of gauge theory from the Coulomb phase to the Higgs phase. Specifically, we can think about what happens to magnetic monopoles when we spontaeously break the $U(1)$ of electromagnetism. The Higgs mechanism occurs when a massless electrically-charged scalar field acquires a vacuum expectation value. In other words light electrically-charged particles form a condensate in the vacuum. When this occurs, magnetic fields are screened and cannot penetrate the vacuum. This is a precise relativistic analog of the Meissner effect in superconductors, where the Cooper pairs of electrons form an electrically-charged condensate and magnetic fields are excluded from the superconducting sample. In this case also, magnetic flux lines cannot spread out and it is energetically preferred for them to form flux tubes or strings of finite tension. The only configurations of finite energy are those where a monopole and anti-monopole are joined by such a flux tube. Thus magnetic charges are confined by a linear potential in the Higgs phase.

So far we have described the consequences of the Higgs mechanism for our conjectural magnetic charges. Of course we are really interested in the confinement of elementary particles which carry ordinary electric charges or their non-Abelian analog. However, in a theory with electric-magnetic duality, the two things should be related. Duality indicates that magnetic charges can become light at strong coupling. In this regime, it is plausible that they can be encouraged (by the introduction of a suitable potential) to condense in the vacuum. In this dual Higgs phase where the vacuum contains a magnetic condensate, electric charges are confined. Thus the elusive confining phase of gauge theory is mapped onto the Higgs phase, which we can understand easily within conventional perturbation theory.

There are several obvious objections to the idea described above: as stated it really only applies to confinement of Abelian electric charge. One may ask how it is expected to work in the context of QCD where there are no stable monopole solutions and the coupling constant inverted by duality is replaced by the dimensionful parameter Λ_{QCD}. Despite this, there has been considerable progress in realizing these ideas quantitatively in supersymmetric gauge theory. First, there is now substantial evidence that electric-magnetic duality is an exact property of $\mathcal{N}=4$ supersymmetric Yang-Mills theory. The famous Seiberg-Witten solution [**19**] of $\mathcal{N}=2$ supersymmetric Yang-Mills theory shows explicitly that monopoles can become massless at strong coupling in an asymptotically free gauge theory. Further soft breaking to $\mathcal{N}=1$ explicitly exhibits confinement due to monopole condensation. Further, a quantitative duality between the Higgs and confining phases of $\mathcal{N}=1$ theories is now well established. We will review some of these developments in the the following.

The second theoretical idea for the 1970's which has gained renewed importance in the supersymmetric context is that of the large N expansion due to Gerard 't Hooft [**1**]. As mentioned above, because of dimensional transmutation, QCD lacks a dimensionless coupling in which to expand. However, a parameter can be introduced by replacing the three colours of QCD with an arbitrary number N or, in other words, working with gauge group $SU(N)$. Following a strategy used earlier in statistical mechanics, one may then look for simplifications in the $N \to \infty$ limit and attempt to expand around this limit in powers of $1/N$. To illustrate how this idea is implemented in practice, we consider vacuum diagrams in pure $SU(N)$ gauge theory with bare coupling constant g. The correct limit to think about for any non-Abelian gauge theory is one in which we take $N \to \infty$ and $g^2 \to 0$ holding the combination $\lambda = g^2 N$ (known as the 't Hooft coupling) fixed. In this limit a vacuum diagram with E propagators, V vertices and F closed loops scales like N^χ where $\chi = V - E + F$. If we consider the Feynman diagram as a triangulation of the two-dimensional surface on which it can be drawn without internal lines crossing, then $\chi = V - E + F$ is precisely the Euler characteristic of that surface. All compact two-dimensional surfaces are characterized up to homeomorphism by their number of handles, the genus $\mathcal{G}$, which is related to the Euler characteristic by $\chi = 2 - 2\mathcal{G}$. The $1/N$ expansion is therefore an expansion in the number of handles of a two-dimensional surface.

This beautiful result means that the leading behavior at large N is governed by planar diagrams which can be drawn on a sphere without lines crossing. Planar diagrams can have an arbitrary number of loops and calculating their sum is a formidable problem which remains unsolved, except in certain lower-dimensional models. However the complexity of the leading order reflects the fact that large N QCD retains many of the features which are hard to analyse directly in the $N = 3$ case. In particular, large N QCD has asymptotic freedom in the UV and quark confinement in the IR[1]. Despite the complexity there are tantalizing indications that the problem of summing the planar diagrams should, ultimately, be tractable. In particular, Witten has argued that the leading order at large N is essentially semi-classical in the sense that the path integral is dominated by a single field configuration known as the 'master field'. A useful review of this and other aspects of the large N limit is [**2**]. Another indication lies in the topological nature of the $1/N$ expansion described above. In particular, as an expansion over 2D surfaces of different topology, it is similar to string perturbation expansion with string coupling constant proportional to $1/N$. A long-standing conjecture is that large N QCD is dual to a weakly-coupled string theory: the QCD string. If this were true, the large N hadron spectrum could be descibed as the (tree-level) spectrum of a string, confirming other indicators of stringy behaviour in the strong interactions, such as Regge behaviour.

The conjecture described above has been around for over 20 years. The key question, "which string theory describes QCD?" has so far remained unanswered. Obvious problems include the fact that the only weakly-coupled string theories we know about are the critical string theories which live in ten dimensions rather then four. These theories inevitably contain a massless spin two particle, identified with the graviton in the context of fundamental strings. Despite these objections

[1]Or at least, the resons why we believe ordinary QCD is confining, such as the area law for the Wilson loop in lattice QCD, remain valid at large N.

significant progress has been made in the more controlled context of supersymmetric gauge theories. For the first time a precise proposal for the large N string dual to a four-dimensional non-Abelian gauge theory has emerged. As in our discussion of electric-magnetic duality above, the maximally supersymmetric case is understood best. Maldacena [**8**] has conjectured that the large N dual of $\mathcal{N} = 4$ SUSY Yang-Mills theory with gauge group $SU(N)$ is the IIB superstring on the spacetime manifold $AdS_5 \times S^5$. Note that this *is* a theory which lives in ten dimensions and has a massless graviton. As we will describe in the following, this reflects some rather special features of the $\mathcal{N} = 4$ theory. Cases with less supersymmetry have also been analysed successfully, although the QCD string itself remains elusive.

We will see below that, like electric-magnetic duality, Maldacena's conjecture relates the strong coupling behaviour of one theory to the weak coupling behaviour of another. Testing any such conjecture clearly requires us to calculate some quantity in both theories and compare. In particular this requires us to perform an independent calculation on the strongly-coupled side of the correspondence. Fortunately, this is where supersymmetry can help. In supersymmetric theories some quantities are protected against quantum corrections and can be reliably extrapolated to strong coupling. Such quantities can often provide quantitative tests of strong-weak coupling duality. In the next section we will give an overview of this aspect of SUSY.

The remainder of these notes are organized as follows. In Section 2 we give a brief overview of protected quantities in SUSY gauge theory. In Section 3 we consider $\mathcal{N} = 4$ supersymmetric Yang-Mills theory at the classical level. In the following two sections we consider the quantum theory in its Coulomb and conformal phases respectively. This naturally includes a discussion of S-duality and the AdS/CFT correspondence. Section 5 is devoted to a brief sketch of the Seiberg-Witten solution of $\mathcal{N} = 2$ SUSY Yang-Mills theory. Finally, we give our conclusions and a brief bibliography.

2. Supersymmetry and Exact Results

In this section, we will briefly review of some of the special properties of supersymmetric theories which allow exact statements to be made about their dynamics. Some of these features are present even in the simplest possible model: supersymmetric quantum mechanics with a single (complex) supercharge Q. In this case the SUSY algebra takes the simple form

$$\{Q\,, Q^\dagger\} = H \tag{2.1}$$

where H is the Hamiltonian. We also introduce a conserved (modulo-two) fermion number operator $(-1)^F$ which anti-commutes with Q. Bose and Fermi eigenstates have eigenvalues ± 1 respectively. One immediate consequence of the SUSY algebra is that the energy of all states is bounded below by zero: $E \geq 0$. The bound is only saturated for states which are annihilated by Q. It follows that the groundstates of the system have zero energy and are invariant under supersymmetry. In contrast, Q acts non-trivially on states of non-zero energy and it follows immediately that these states necessarily come in degenerate Bose-Fermi pairs.

States in the Hilbert space therefore come in two different 'multiplets' of supersymmetry depending on whether their energies saturate the bound $E \geq 0$. This is a characteristic feature of supersymmetry which we will meet again. Another such feature is the existence of quantities which are protected from quantum corrections. So far we have met the groundstate energy which is strictly zero. A more sophisticated example is the Witten Index, defined by,

$$\mathcal{I} = \mathrm{Tr}\,(-1)^F \, \exp\,(-\beta H)\,. \tag{2.2}$$

Because states of non-zero energy come in Bose-Fermi pairs their contribution to the trace cancels. Thus we have $\mathcal{I} = n_B - n_F$, where n_B and n_F are the numbers of Bose and Fermi groundstates respectively. As we vary the parameters of the model groundstates can only be lifted to non-zero energy in Bose-Fermi pairs. Similarly new groundstates can only appear if a Bose-Fermi pair descends to zero energy. Neither of these processes change the value of $\mathcal{I}$, which therefore remains invariant under all smooth variations of the parameters which preserve SUSY[2].

Now we move on to the more interesting case of supersymmetric quantum field theory. The global supersymmetry algebra in four dimensions has left and right handed Weyl spinor supercharges Q^A_α and $\bar{Q}_{\dot{\alpha}A}$ respectively. Here we are adopting the standard notation of Wess and Bagger [**5**]: α and $\dot{\alpha}$ are left and right-handed Weyl spinor indices and $A = 1, 2, \ldots, \mathcal{N}$ labels distinct families of supercharges. In four-dimensional renormalizable quantum field theory we are only interested in algebras which lead to multiplets of particles with spins ≤ 1: this limits the possiblities to $\mathcal{N} = 1, 2$ and 4. The so-called extended SUSY algebras with $\mathcal{N} = 2$ and 4 have an $SU(\mathcal{N})$ internal symmetry which rotates the supercharges. The left- and right-handed supercharges are in the $\mathcal{N}$ and $\bar{\mathcal{N}}$ of this group, denoted by the raised and lowered index $A = 1, 2, \ldots, \mathcal{N}$. The basic anti-commutator is

$$\{Q^A_\alpha\,, \bar{Q}_{\dot{\beta}B}\} = 2\sigma^\mu_{\alpha\dot{\beta}} P_\mu \delta^A_B. \tag{2.3}$$

For theories with extended SUSY, supercharges of the same chirality also have a non-trivial anti-commutator,

$$\{Q^A_\alpha\,, Q^B_\beta\} = \epsilon_{\alpha\beta}\mathcal{Z}^{AB}. \tag{2.4}$$

Here $\mathcal{Z}^{AB} = -\mathcal{Z}^{BA}$ is a central charge which commutes with all the other elements of the algebra.

We will now briefly list some of the special features of these theories which constrain the form of quantum corrections:

Non-renormalization: As in the example of SUSY quantum mechanics considered above, the SUSY algebra implies that the energy of a supersymmetric vacuum of the theory is exactly equal to zero. Typically in quantum field theory the vacuum energy may be set to zero at the classical level, but it will certainly recieve quantum corrections from loop diagrams. Unbroken supersymmetry implies that these corrections vanish to all orders. This vanishing is due to a cancellation between bosons and fermions, which reflects the additional factor of -1 in the Feynman rules for a fermionic loop.

BPS states: Just as in SUSY quantum mechanics the supersymmetry algebra implies that the energy of all states is bounded below by zero: $E \geq 0$. In the case of extended SUSY, one can actually derive a stronger bound for the masses of particles of the schematic form $M \geq |\mathcal{Z}|$ where $\mathcal{Z}$ is the central charge. This is known as the Bogomol'nyi bound. Another similarity to the quantum mechanical case is the fact

[2]There are several caveats here: the logic described above only applies when the Hamiltonian has a discrete spectrum. In theories with non-compact directions we must also limit ourselves to variations which do not change the asymptotic behaviour of the potential in these directions.

that states which saturate this bound live in smaller representations of the SUSY algebra. Specifically, states which saturate the bound are invariant under exactly half the SUSY generators and typically fill out 'short' or BPS representations of dimension $2^{\mathcal{N}}$. In contrast, generic states transform non-trivially under all the SUSY generators and form 'long' representations of dimension $2^{2\mathcal{N}}$. Considerations like those described above leading to the Witten index in the case of quantum mechanics mean that BPS states cannot be lifted above the Bogomol'nyi bound except in multiples of $2^{\mathcal{N}}$. Typically, the necessary additional BPS states are not present and we conclude that the BPS mass formula $M = |\mathcal{Z}|$ is exact.

Holomorphy: Certain special quantities in SUSY gauge theories are constrained to be holomorphic functions of the fields and parameters. Much recent progress can be traced to the fact that holomorphic functions can be reconstructed exactly from knowledge of their singularities.

Flat directions: In theories with extended SUSY the potential energy typically has flat directions along which scalar fields can acquire expectation values without energetic cost. Fluctuations along these directions are massless scalars. It is important to emphasize that these scalars are not necessarily Goldstone modes for any global symmetry. Their masslessness is instead a consequence of unbroken supersymmetry. This situation is often described by saying that the theory has a *moduli space*, $\mathcal{M}$, of gauge-inequivalent vacua. It is often useful to think of the moduli space as a Riemannian manifold with the scalar fields $\{X_i\}$, with $i = 1, 2, \ldots \dim_{\mathbb{R}}\mathcal{M}$ as real coordinates. The natural metric $\mathcal{G}_{ij}$ on $\mathcal{M}$ is provided by the lowest term in the derivative expansion of the Wilsonian effective action for the massless scalars,

$$\mathcal{L}_{\text{eff}} = \frac{1}{2}\,\mathcal{G}_{ij}\,(X)\;\partial_\mu X^i \partial^\mu X^j \;+\;\ldots. \tag{2.5}$$

Of course we must really consider a supersymmetric version of (2.5) with kinetic terms for the superparteners of the massive fields. It turns out that the existence of such an extension provides stringent constraints on the geometry of $\mathcal{M}$. In particular, for $\mathcal{N} = 1$ SUSY in four dimensions, it is constrained to be a Kähler manifold. In various situations with extended supersymmetry the requirement is for hyper-Kähler or special Kähler geometry.

In the next sections we will review what we know about theories with $\mathcal{N} = 4$ and $\mathcal{N} = 2$ supersymmetry.

3. $\mathcal{N} = 4$ SUSY Yang-Mills: the Classical Theory

$\mathcal{N} = 4$ supersymmetric Yang-Mills theory in four dimensions can be obtained by dimensional reduction of the minimal supersymmetric gauge theory in ten dimensions. In the following we will consider the theory with gauge group $SU(N)$ (or sometimes $U(N)$). The field content is a single massless vector multiplet. In addition to the gauge field A_μ, the multiplet includes four adjoint-valued left-handed Weyl fermions λ^A_α with $A = 1, 2, 3, 4$ which transform in the fundamental representation of the $SU(4)$ R-symmetry mentioned in the introduction. The charge-conjugate degrees of freedom are right-handed Weyl spinors, $\bar{\lambda}_{\dot{\alpha}A}$ in the anti-fundamental of $SU(4)_R$. The multiplet is completed by six adjoint scalar fields ϕ_a, with $a = 1, 2, \ldots, 6$ which form a $\mathbf{6}$ of $SU(4)_R$. As $SU(4)$ is locally the same as $SO(6)$, this can also be thought of as the vector representation of the latter group.

The Lagrangian consists of the Maxwell term for the gauge fields together with minimal coupling for the adjoint scalars and fermions. The Lagrangian also includes a non-trivial potential for the scalar fields,

$$\mathcal{V} = \sum_{a>b} \mathrm{Tr}_N \left([\phi_a,\phi_b]^2 \right) . \tag{3.1}$$

Finally the supersymmetric completion of these terms include Yukawa couplings between the fermions and scalars. All these terms have mass dimension four and the resulting Lagrangian is therefore scale-invariant. In fact the classical Lagrangian is invariant under the group of conformal transformations in four dimensions, $SO(4,2)$. This group includes the Poincaré generators together with dilatations and special conformal transformations. The coefficients of each term in the Lagrangian are uniquely determined by $\mathcal{N}=4$ supersymmetry up to a single overall coefficient, $1/g^2$. Here g is the dimensionless gauge coupling constant. One may also add to the Lagrangian the topological term $\mathrm{Tr}_N(F \wedge F)$ with coefficient the vacuum angle θ. It is customary to combine g^2 and θ in a single complex coupling,

$$\tau = \frac{4\pi i}{g^2} + \frac{\theta}{2\pi}. \tag{3.2}$$

Apart from the choice of gauge group, this is the only free parameter in the theory.

As mentioned above, the theory is invariant under the conformal group and also under $\mathcal{N}=4$ supersymmetry transformations. The closure of these generators results in the largest possible global spacetime symmetry group in four dimensions. This is the $\mathcal{N}=4$ superconformal group, denoted $SU(2,2|4)$. Its bosonic part is $SO(4,2) \times SU(4)_R$. In addition to the sixteen real components of the SUSY generators Q^A_α and $\bar{Q}_{\dot{\alpha}A}$, there are sixteen independent superconformal generators, S^A_α and $\bar{S}_{\dot{\alpha}A}$.

The $\mathcal{N}=4$ theory exhibits a characteristic feature of theories with sixteen supercharges mentioned in the previous section: the potential $\mathcal{V}$ has flat directions. Specifically, if we choose each of the six scalar fields to have vacuum expectation values (VEVs) in the Cartan subalgebra of $U(N)$, then the potential vanishes and we have a supersymmetric vacuum. The corresponding manifold of vacuum states is known as the Coulomb branch. The behaviour of the theory depends crucially on whether these VEVs are non-zero. The two cases correspond to different phases of the system which we describe in turn:

Coulomb Phase: If the scalar VEVs are non-zero then the gauge group $G = U(N)$ is spontaneously broken down to its Cartan subgroup $U(1)^N$. The theory has massless photons corresponding to unbroken gauge generators which give rise to long range Abelian gauge fields (hence the term 'Coulomb'). As the scalar fields have mass dimension, choice of a VEV also spontaneosly breaks the conformal symmetry down to its Poincaré subgroup and also spontaneously breaks part of the $SU(4)_R$ symmetry. Some of the massless scalars which appear as superpartners of the photons can be thought of as Goldstone bosons in this context. In addition to massless neutral fields, the theory also has massive particles which carry electric and magnetic charges with respect to the Cartan subgroup.

Conformal Phase: If the scalar VEVs are zero, then the $U(N)$ gauge symmetry remains unbroken and so does the classical superconformal symmetry. The theory is then in an interacting conformal phase. There is no mass gap and states in the Hilbert space form multiplets of the superconformal group.

Before moving on to describe the quantum theory, we will take a closer look at the spectrum of the theory with gauge group $SU(2)$ on the Coulomb branch. In this theory we can expand the bosonic fields in terms of the three Pauli matrices τ_i, $i = 1, 2, 3$, in the obvious way as $A_\mu = A^i_\mu \tau_i/2$, $\phi^{ia}\tau_i/2$. Up to a gauge rotation, we can choose scalar VEVs of the form $\langle\phi^a\rangle = v^a\tau_3/2$. The Coulomb branch is therefore parameterized by the complex six-component vector v^a with an $SU(4)_R$-invariant length defined by $|v|^2 = \sum_{a=1}^{6} v^a\bar{v}^a$. The Weyl group of $SU(2)$ acts as $v^a \to -v^a$ and, modding this out, the manifold of gauge inequivalent vacua is just $\mathbb{R}^6/Z_2$.

As above, when $|v| > 0$, the gauge group is spontaneously broken down to the Cartan subalgebra, which, in this case, is the $U(1)$ subgroup of $SU(2)$ generated by the third Pauli matrix τ_3. In principle, particle states may carry electric and/or magnetic charges, denoted Q_E and Q_M respectively, with respect to this $U(1)$. On general grounds, the allowed charges are constrained by the Dirac quantization condition $Q_E Q_M \in 2\pi\,\mathbb{Z}$. A semiclassical analysis of the theory reveals the following states:

• A massless $\mathcal{N} = 4$ vector multiplet with $Q_E = Q_M = 0$ containing the photon and its superpartners. These are the elementary quanta which remain massless after the Higgs mechanism. They correspond to fluctuations of the fields living in the unbroken $U(1)$.

• An $\mathcal{N} = 4$ vector multiplet containing massive gauge bosons and their superpartners. Each of these states has $Q_E = \pm g$, $Q_M = 0$ and mass $M = |v|$. The spin one members of the multiplet are fluctuations of the fields $W^\pm_\mu = A^1_\mu \pm iA^2_\mu$ and are referred to as 'W-bosons'.

In addition to these elementary quanta, the classical theory also contains magnetically charged solitons or magnetic monopoles. These are classical solutions of the field equations for the Yang-Mills field minimally coupled to an adjoint scalar. The configurations have finite mass $M = 4\pi|v|/g^2$ and carry magnetic charge $Q_M = 4\pi/g$. Somewhat miraculously, the units of electric and magnetic charge carried by the W-bosons and magnetic monopoles respectively turn out to satisfy the Dirac quantization.[3] At weak coupling the monopole can be quantized in a semiclassical approximation and leads to massive one-particle states in the Hilbert space. These include,

• An $\mathcal{N} = 4$ supermultiplet of states with $Q_E = 0$ and $Q_M = 4\pi/g$.

• An infinite tower of dyonic states which carry integer multiples of the quantum of electric charge in addition to one unit of magnetic charge: $Q_E = ng$, $Q_M = 4\pi/g$ with integer n.

The mass of each of these states is described by the universal mass formula:

$$M = \frac{|v|}{g}\sqrt{Q_E^2 + Q_M^2}. \tag{3.3}$$

[3]This appears miraculous because the Dirac quantization condition is derived from quantum mechanical considerations, while the quantization of magnetic charge observed in this theory happens already at the classical level. A further miracle is the fact that the monopole actually carries twice the minimum allowed quantum of magnetic charge. This means that the theory remains consistent when particles in the fundamental representation of $G = SU(2)$ are introduced which have *half* the electric charge of the W-boson.

4. The Quantum Theory of the Coulomb Phase

In this section and the next we describe some of the remarkable results for the quantum $\mathcal{N}=4$ theory which have emerged over the last few years. As mentioned in the introduction the large space-time symmetry group of the theory places strong restrictions on the form of quantum corrections to the classical results of the last section. Of course an important general issue in any theory is whether the symmetries of the classical theory survive quantization or, in other words, whether the theory has anomalies. For example, the ordinary Maxwell action for any non-Abelian gauge theory is scale invariant at the classical level, manifesting itself in the presence of a dimensionless coupling constant g^2. However, this scale invariance is usually broken at one loop by the running of the coupling with energy scale μ: $g^2 \to g^2(\mu)$. The breaking of scale invariance is measured by the Renormalization Group β-function, $\beta(g) = \mu(\frac{\partial}{\partial\mu})g^2(\mu)$. In asymptotically free gauge theories such as QCD we have $\beta < 0$, which means that the running coupling becomes large in the IR.

A remarkable fact about the $\mathcal{N}=4$ theory is that the β-function is exactly zero. The cancellation occurs order by order in perturbation theory due to a cancellation between bosonic and fermionic loops. This is a striking example of the non-renormalization properties of SUSY gauge theories mentioned in the introduction. It reflects the fact that the $\mathcal{N}=4$ theory is exactly scale-invariant at the quantum level. A related fact is that the theory does not have a chiral anomaly. In theories with massless fermions and a chiral anomaly the vacuum angle can be set to zero by a chiral rotation of the fermions. In contrast, in the $\mathcal{N}=4$ theory, θ remains as a genuine parameter of the theory. More precisely the physics depends periodically on θ via the Yang-Mills instanton factor $q = \exp(2\pi i\tau) = \exp(-8\pi^2/g^2 + i\theta)$ which is invariant under the transformation T: $\theta \to \theta + 2\pi$.

As in the previous section, the physics depends critically on whether the adjoint scalars acquire a vacuum expectation value. As before we have two phases: in the conformal phase where the gauge group is unbroken the theory is invariant under the full $\mathcal{N}=4$ superconformal algebra. In the language of the Renormalization Group, the dimensionless parameter $\tau = 4\pi i/g^2 + \theta/2\pi$ is an exactly marginal coupling which parametrizes a fixed line of interacting superconformal theories. The situation in the Coulomb phase is slightly different. As before, the superconformal symmetry is spontaneously broken by a choice of scalar VEV. As there is no explicit breaking of the symmetry the classical Goldstone bosons remain massless in the full quantum theory. There are remarkable properties of the physics in each of these phases which we will now discuss in turn.

In the classical Coulomb phase, the theory has a spectrum of massive particles carrying magnetic and electric charges as described in the previous section. In a generic QFT, we would be able to say little about the spectrum of the quantum theory apart from computing the radiative corrections to the classical masses in a weak coupling expansion. In the $\mathcal{N}=4$ theory however, we are able to do much better than this, and we can actually make some exact statements about the spectrum. The underlying reason is the presence of a central charge in the $\mathcal{N}=4$ supersymmetry algebra. As above we have the anti-commutator,

$$\{Q^A_\alpha, Q^B_\beta\} = \epsilon_{\alpha\beta}\mathcal{Z}^{AB}, \tag{4.1}$$

where $\mathcal{Z}^{AB}$ is an anti-symmetric 4×4 matrix. By a unitary rotation we can put this matrix in the canonical form,

$$\mathcal{Z} \quad = \quad 2|Z| \begin{pmatrix} 0 & +1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & +1 \\ 0 & 0 & -1 & 0 \end{pmatrix}.$$

By working in the restframe of a massive particle, the algebra can be used to prove a Bogomol'nyi bound for the mass of any state: $M \geq |Z|$. Generic states in the spectrum whose masses lie above the bound live in 'long' multiplets of $\mathcal{N} = 4$ SUSY of dimension $2^{2\mathcal{N}} = 256$. In contrast, states with $M = |Z|$, which saturate the bound, are invariant under half the generators and live in 'short' or BPS multiplets of dimension $2^{\mathcal{N}} = 16$. As mentioned in the introduction, the key fact is that, as we vary the parameters of the theory, BPS-saturated states cannot be lifted above the Bogomol'nyi bound except by combining with fifteen other short multiplets with the same quantum numbers to make a long multiplet. Normally the additional states are simply not available and we can safely conclude that BPS states which are present in one regime of parameters will be present throughout the parameter space. Further, the relation $M = |Z|$ is the exact mass formula for these states.

To proceed further we must identify the central charge. In the $SU(2)$ theory it is given by

$$Z = \frac{v}{g}\left(Q_E + iQ_M\right) \tag{4.2}$$

where Q_E and Q_M are electric and magnetic charges with respect to the unbroken $U(1)$ subgroup of $SU(2)$ on the Coulomb branch. The mass formula for BPS states is then precisely (3.3) above and we can conclude that the entire spectrum of electric and magnetically charged states discussed in the previous section are actually BPS states. The fact that the 't Hooft-Polyakov monopoles in this theory yield BPS states is closely connected with the fact that the corresponding classical soliton solutions obey first order Bogomol'nyi equations of the form $B_i = \pm D_i\phi$. This equation is precisely the condition that the classical field configuration is invariant under half the supersymmetry generators. It is also easy to find hints as to why the W-bosons should be BPS saturated. These states are obtained from the spectrum of massless states via the Higgs mechanism and massless states are automatically BPS saturated.

In the classical analysis of the previous section, we saw that the basic unit of electric charge is that of the W-boson, which is g, while the basic unit of magnetic charge is that carried by the 't Hooft-Polyakov monopole, which is $4\pi v/g$. We can express the charges of an aibitrary state in terms of integer multiples of these unit charges by setting $Q_E = n_E g$ and $Q_M = n_M(4\pi/g)$ where n_E and n_M are integers. In these units the central charge of the $SU(2)$ theory can be written as

$$Z = v\left(n_E + n_M\tau\right). \tag{4.3}$$

Actually this formula is a slight extension of the previous one as we have incorporated the effects of a non-zero θ angle. It reflects a phenomenon first discovered by Witten in [**15**]. In the presence of non-zero θ, monopoles with n_M units of magnetic charge acquire an electric charge $\Delta Q_E = n_M g\theta/2\pi$. The Dirac quantization condition is also modified by the non-zero vacuum angle to accommodate this shifted charge. Thus, in the presence of n_M units of magnetic charge, the shift

$T : \theta \to \theta + 2\pi$ is no longer a trivial invariance of the theory. It shifts the integer-valued electric charge n_E of every state by exactly n_M units. However T will still leave the spectrum of the theory unchanged provided the theory already contains all the possible image states under the repeated application of this transformation (and its inverse).

In fact the BPS spectrum described above has another much more profound invariance. As we have a theory with both electric and magnetic charges, we ask to what extent they appear symmetrically in the theory. One may look for a transformation which interchanges electric and magnetic charges and fields in a way which leaves Maxwell's equations invariant. The action on the charges is $Q_E \to Q_M$ and $Q_M \to -Q_E$. As our units of electric and magnetic charge (for $\theta = 0$) are determined by the Dirac quantization condition to be $Q_E = n_E g$ and $Q_M = n_M 4\pi/g$, this requires the transformations $n_E \to n_M$, $n_M \to -n_E$ together with $g \to 4\pi/g$. Olive and Montonen [**13**] were the first to observe that this is indeed an invariance of the BPS spectrum described above, and they were also the first to conjecture that this could be an exact duality of a non-Abelian gauge theory in its Coulomb phase.

Several comments are in order about the conjectured duality. First of all, as it inverts the coupling constant, it relates the strong coupling behaviour of the theory to its weak coupling behaviour. As discussed in the introduction, this makes the proposal extremely interesting but hard to test without independent knowledge about the strongly-coupled regime. The transformation also interchanges the roles of the electric charges which appear as elementary particles at weak coupling with the magnetic charges which appear as classical solitons. Thus Olive-Montonen duality shares many features with the duality between the sine-Gordon and Thirring models in two dimensions [**4**], where the kink of the former model is identified with the elementary fermion of the latter. The conjecture of Olive and Montonen refers in the first instance to an ordinary bosonic gauge theory without supersymmetry, and in this context there are several obvious objections. First the invariance is only demonstrated at the level of the classical mass formula, which in bosonic gauge theory will certainly be corrected by quantum effects. Quantum effects also lead to the running of the coupling constant and the generation of a dynamical mass scale which is the only true parameter of the theory. Strong and weak coupling both occur in the same theory at different energy scales and it is unclear how to implement the tranformation $g \to 4\pi/g$. Finally the symmetry between the monopole and the electrically charged gauge boson is broken by the fact that the latter states have spin one. In the absence of fermions, the monopole field configuration is essentially spherically symmetric and therefore does not carry angular momentum.

It is a remarkable fact that each of the objections raised above is without force in the context of $\mathcal{N} = 4$ supersymmetric Yang-Mills theory. As mentioned above the BPS mass formula $M = |Z|$ is exact and, together with the stability properties of BPS states described above, it can be used to determine the BPS spectrum at strong coupling. As the β-function of the $\mathcal{N} = 4$ theory vanishes exactly, there are none of the problems associated with the running of the coupling. Perhaps most remarkably, in the $\mathcal{N} = 4$ theory, the W-boson and the BPS monopole lie in supermultiplets with the same spin content [**14**]. Although the classical monopole itself does not carry spin, it can form threshold bound-states with the elementary fermions of the theory which do. The fact that these states fill out a short multiplet is intimately related to the fact mentioned above that the BPS monopole is invariant under half the supersymmetry generators.

It is instructive to consider the extension of electric-magnetic duality occurring when the vacuum angle is non-zero. In this case, duality acts on the complexified coupling $\tau = 4\pi i/g^2 + \theta/2\pi$ as the transformation $S : \tau \to -1/\tau$. Obviously S^2 is the identity and repeated application of this transformation does not generate any thing new. However, we should also remember the transformation T introduced above which shifts the vacuum angle by 2π. This acts on the complexified coupling as $T : \tau \to \tau + 1$. Importantly though, S and T do not commute with each other and, together, they generate an infinite group of transformations known as $SL(2, \mathbb{Z})$. The action of a general element of this group on the coupling is

$$\tau \quad \to \quad \frac{a\tau + b}{c\tau + d} \tag{4.4}$$

where a, b, c and d are integers with $ad - bc = 1$. The corresponding action on the integer-valued electric and magnetic charges is

$$\begin{pmatrix} n_E \\ n_M \end{pmatrix} \quad \to \quad \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} n_E \\ n_M \end{pmatrix}.$$

This group is well known in complex analysis as the modular group of the torus. If complex tori are parameterized by the ratio of their periods $\tau = \omega_2/\omega_1$, then values of τ related by an $SL(2, \mathbb{Z})$ transformation as in (4.4) describe tori which are equivalent as complex manifolds. This geometric realization of the duality group is central to how electric-magnetic duality appears on branes living in toroidal compactifications of M-theory.

The extended electric-magnetic duality described above is known as S-duality. There is now much evidence that S-duality is actually an exact invariance of the $\mathcal{N} = 4$ theory. The simplest evidence in favour of this comes from the spectrum of BPS states. Although the BPS mass formula is certainly invariant under S-duality, invariance of the spectrum clearly requires the existence in the theory of the $SL(2, \mathbb{Z})$ images of all the known BPS states. In particular, it is easy to prove that the $SL(2, \mathbb{Z})$ orbit of the W-boson with electric charge $n_E = 1$ and zero magnetic charge includes states with *all* coprime values of n_E and n_M. For $n_M = 1$, this amounts to the existence of the dyonic states mentioned in the previous section. However, for $n_M > 1$, we also have a prediction for the existence of many new states which can be thought of as bound-states of monopoles and dyons. By the usual reasoning, it suffices to check the existence of these states at weak coupling where we may use a semiclassical quantization of the multi-monopole system. The existence of the required bound-states translates into the existence of certain normalizable harmonic forms on the moduli space of BPS monopoles. This has been checked explicitly by Sen [**16**] in the case of magnetic charge two, where an explicit metric on the monopole moduli space is known. Progress has also been made in demonstrating the existence of the necessary forms in the case of arbitrary magnetic charge [**17**].

5. The Quantum Theory of the Conformal Phase

We now return to the conformally invariant phase where the scalar VEVs are zero and the $SU(N)$ gauge symmetry is unbroken. As mentioned in Section 2, in this phase the theory is characterized by invariance under the supergroup $SU(2,2|4)$, which contains the four-dimensional conformal group $SO(4,2)$ and the R-symmetry group $SU(4)_R$ together with 32 real supercharges. The Hilbert space of the theory is made up of states transforming in irreducible representations of this supergroup. States are labeled by their eigenvalues under the appropriate Casimir operators. These include the left and right moving spins (j_L, j_R), the scaling dimension D, which is the eigenvalue of the dilatation operator, and the Casimirs of $SU(4)_R$. As in any conformal field theory, there is a natural map between local operators and states provided by radial quantization, so we can equivalently think in terms of local, gauge-invariant operators labeled by the same quantum numbers. Roughly speaking, the state corresponding to a given operator is that obtained by acting on a canonically chosen vacuum state with the operator in question.

In the previous sections, we encountered the phenomenon of short or BPS-saturated multiplets of supersymmetry. It turns out that an exactly analogous phenomenon occurs in the representation theory of the superconformal algebra. The easiest case to describe is that of $\mathcal{N}=1$ superconformal symmetry, the conformal extension of the minimal supersymmetry algebra in four dimensions. The $\mathcal{N}=1$ superconformal group contains a $U(1)$ global R-symmetry. States or operators are labeled by a definite scaling dimension D and a definite $U(1)$ R-charge R. There is an absolute lower-bound on these scaling dimensions, $D \geq |R|$. This is analogous to the Bogomol'nyi bound in ordinary supersymmetry and, as in that case, states which saturate the bound are invariant under half of the algebra and consequently live in short representations. The corresponding operators are known as chiral primaries and, by a now familiar argument, they cannot be lifted above the bound except in very special circumstances. In cases where the theory has a weakly-coupled limit, we will typically find a set of these chiral primary operators in the classical theory with classical dimensions which saturate the bound. Their BPS stability property then ensures that the corresponding states remain in the spectrum for all values of the coupling. In the usual language of the Renormalization Group, the exact formula $D = |R|$ ensures that these operators do not acquire anomalous dimensions.

The $\mathcal{N}=4$ superconformal algebra also has chiral primary representations. Roughly speaking, the states are chiral primary with respect to the $\mathcal{N}=4$ superconformal algebra if they are chiral primary with respect to any $\mathcal{N}=1$ subalgebra. The condition on their scaling dimensions reads $D = |R|$, where R is the charge under any $U(1)$ subgroup of $SU(4)_R$. One infinite class of such operators is formed by taking symmetric traceless combinations of the six adjoint scalar fields ϕ_a. For example, the operators

$$O_{a_1 a_2 \ldots a_n} = \mathrm{Tr}_N \left[\phi_{a_1}\phi_{a_2}\ldots\phi_{a_n}\right] \tag{5.1}$$

are chiral primary provided we symmetrize the $SU(4)_R$ indices and subtract off traces.

The quantum behaviour of the $\mathcal{N}=4$ theory in its conformal phase is highly constrained by superconformal symmetry. Apart from the spectrum of states, the natural observables are the correlation functions of local gauge-invariant operators.

These obey superconformal Ward identities which are sufficient to determine all two- and three- point functions of chiral primary operators exactly. However, scaling dimensions and correlation functions of other operators in the theory are much less constrained and, in general, they will receive complicated perturbative and non-perturbative corrections. Despite this, two remarkable properties of the quantum theory have emerged in recent years. The first is the electric-magnetic duality discussed in the previous section. Unlike the Coulomb phase, the $SU(N)$ gauge group is unbroken and we cannot define Abelian electric and magnetic charges. Nevertheless there is still considerable evidence that the conformal theory has an exact $SL(2,\mathbb{Z})$ duality acting on the complexified coupling constant as in (4.4). A precise formulation is that the operators in the theory and their correlation functions transform as modular forms of definite weight[4]. In [**27**], Vafa and Witten evaluated the partition function of a topologically-twisted version of the $\mathcal{N}=4$ theory on various four-manifolds and verified its modular properties explicitly.

The second major development is one we will touch on only briefly. As mentioned in the introduction, there is a long-standing conjecture that the large N limit of non-Abelian gauge theory should be dual to a weakly-coupled string theory. Until recently, the identity of the dual string theory was essentially unknown. In late 1997, Maldacena [**8**] made a remarkable proposal concerning the large N dual of the $\mathcal{N}=4$ theory at its superconformal point. Specifically it concerns the $\mathcal{N}=4$ theory with gauge group $SU(N)$ in the 't Hooft limit $N\to\infty$, $g^2\to 0$ with the 't Hooft coupling $\lambda=g^2N$ held fixed. Maldacena conjectured that the $\mathcal{N}=4$ theory in this limit (and perhaps beyond it) is equivalent to Type IIB superstring theory propagating on the ten-dimensional spacetime, $AdS_5\times S^5$. Here AdS_5 denotes five-dimensional anti-de-Sitter space and S^5 is the five-sphere. This proposal is also known as the AdS/CFT correspondence. The identification of the parameters of the two theories is as follows: the complexified coupling $\tau=4\pi i/g^2+\theta/2\pi$ of the $\mathcal{N}=4$ theory is identified with a corresponding complex parameter $\tau_{IIB}=i/g_{str}+2\pi C^{(0)}$ of the IIB theory. Here g_{str} is the string coupling and $C^{(0)}$ is the VEV of the scalar field arising from the Ramond-Ramond sector of the IIB theory. The radius, L, of both the AdS_5 and S^5 factors of the spacetime, measured in units of the string length $\sqrt{\alpha'}$, scales like a positive power of the 't Hooft coupling: $L^2/\alpha'\sim\sqrt{\lambda}$. Additionally there are N units of R-R fiveform flux through the five-sphere.

Several comments are in order. The first point of agreement between the two sides of the correspondence is the global symmetries appearing on both sides. The $SO(4,2)$ group of conformal transformations on the $\mathcal{N}=4$ side emerges as the isometry group of AdS_5. Similarly, the $SU(4)$ R-symmetry of the $\mathcal{N}=4$ theory corresponds to the isometry group of the five-sphere. The number of supercharges of the IIB theory is thirty-two and all of these remain unbroken in the $AdS_5\times S^5$ background. In fact they enlarge the bosonic symmetry group to form the supergroup $SU(2,2|4)$ which is precisely the $\mathcal{N}=4$ superconformal group in four dimensions. In retrospect, Maldacena's proposal is strongly suggested by the fact that IIB on $AdS_5\times S^5$ is the only known string theory with this symmetry group. More complicated dynamical aspects of the $\mathcal{N}=4$ theory are also reproduced on the IIB side. In particular the IIB theory has an $SL(2,\mathbb{Z})$ duality group acting on τ_{IIB}, including transformations which interchange the NS-NS and R-R sectors of

[4] $f(\tau,\bar{\tau})$ is a modular form of weight $(w,\bar{w})$ if it transforms as $f\to(c\tau+d)^w(c\bar{\tau}+d)^{\bar{w}}f$ under the $SL(2,\mathbb{Z})$ transformation (4.4)

the theory. This corresponds to the S-duality group of the $\mathcal{N} = 4$ theory acting on τ.

The fact that the string coupling is identified with the Yang-Mills coupling constant (squared), which scales like $1/N$, means that the dual string theory is weakly coupled at large N as originally envisaged by 't Hooft. In principle the sum of planar diagrams in the $\mathcal{N} = 4$ theory can be evaluated by solving the tree-level string theory of $AdS_5 \times S^5$. Unfortunately, the latter task has proved to be extremely difficult. The main problem is that the background includes non-zero R-R flux, something which is hard to incorporate in the theory on the string world-sheet. Rather than working with the full string theory on $AdS_5 \times S^5$, one may instead consider only the low-energy approximation to the theory, which amounts to studying IIB supergravity on the same background. This approximation is reliable as long as the stringy excitations are more massive than the Kaluza-Klein modes of the SUGRA fields on the S^5. This is true provided the radius of the sphere is much larger that the string length. Given the identification $L^2/\alpha' = \sqrt{\lambda}$, this corresponds to an interesting strongly-coupled regime of the $\mathcal{N} = 4$ theory. In contrast, Feynman perturbation theory is only valid when the 't Hooft coupling is small. Hence the regimes in which we may calculate reliably on each side of the correspondence are non-overlapping.

In this sense, AdS/CFT is a weak-coupling/strong-coupling duality like the electric-magnetic duality discussed above. In particular, we are confronted with the same problem we encountered in that context: in order to test Maldacena's proposal we need to perform an independent calculation when one of the two theories in question is strongly coupled. As in the case of S-duality, the special properties of supersymmetric gauge theory come to the rescue. In particular, as described above, we know the theory contains a large set of operators whose scaling dimensions are independent of the coupling. One may safely extrapolate the classical spectrum of $\mathcal{N} = 4$ chiral primaries to the regime of large 't Hooft coupling where they can be compared with the supergravity spectrum. In particular, the IIB theory on $AdS_5 \times S^5$ contains an infinite tower of states living in small representations of the $SU(2,2|4)$ superalgebra. These states correspond to the expansion of the massless ten-dimensional fields of IIB supergravity in spherical harmonics on the five-sphere. This comparison was first performed by Witten in [**9**], and yields impressive agreement.

6. Theories with less supersymmetry

So far we have only discussed the $\mathcal{N} = 4$ theory. We have seen that the large degree of supersymmetry gives us enough control over quantum effects to formulate and test highly non-trivial dualities like $SL(2,\mathbb{Z})$ and AdS/CFT. However, the extended supersymmetry also leads to features, like the vanishing β-function, which are rather unrealistic from the point of view of QCD. In fact, powerful exact results also exist for theories with less supersymmetry, most notably the celebrated solution of $\mathcal{N} = 2$ SUSY Yang-Mills due to Seiberg and Witten [**19**]. In the remainder of this article we will provide a brief sketch of the physics of these theories emphasizing the role of electric-magnetic duality.

As in the original Seiberg-Witten paper we will focus on the minimal $\mathcal{N} = 2$ theory with gauge group $SU(2)$. This consists of a single vector multiplet of $\mathcal{N} = 2$ SUSY which contains the $SU(2)$ gauge field, two Weyl fermions and a single complex scalar field Φ. All fields are in the adjoint representation of the gauge group and the theory has a potential term,

$$\mathcal{V} = \mathrm{Tr}\left([\Phi,\Phi^{\dagger}]^2 \right). \tag{6.1}$$

As in the $\mathcal{N}=4$ theory the potential has a flat direction along which the complex scalar acquires a VEV breaking the $SU(2)$ gauge symmetry down to its Cartan subalgebra. By gauge rotation we can choose the generator of the unbroken $U(1)$ to be the third Pauli matrix. The resulting Coulomb branch can be parameterized by a single complex number a, up to Weyl reflection $a \to -a$, as $\langle\Phi\rangle = a\tau_3/2$. Alternatively we may introduce the gauge-invariant modulus $u = \langle \mathrm{Tr}\,\Phi^2\rangle$. Classically we have $u = a^2/2$ although, as u is a composite operator, this relation can receive quantum corrections. The classical theory on the Coulomb branch is very similar to that of the $\mathcal{N}=4$ theory. Apart from the massless photon multiplet, there are also massive $\mathcal{N}=2$ multiplets corresponding to the massive W-bosons, as well as a spectrum of monopoles and dyons. All of these states live in short multiplets of $\mathcal{N}=2$ SUSY. The classical Bogomol'nyi bound is $M \geq |Z|$, with central charge,

$$Z \;=\; a\,(n_E + n_M\tau) \tag{6.2}$$

where n_E and n_M are integer-valued electric and magnetic charges.

The differences between the $\mathcal{N}=4$ and $\mathcal{N}=2$ theories emerge as soon as we compute quantum corrections. In the minimal $\mathcal{N}=2$ theory the β-function is non-zero and negative at one-loop. This means that the $\mathcal{N}=2$ theory is asymptotically free in the UV. The coupling constant runs and is replaced by a dynamical scale Λ. In the absence of scalar VEVs, the theory runs to strong coupling at the scale Λ and, a priori, it is very hard to say what the IR physics is like. However, if the scalar VEV a is non-zero, the gauge group is broken down to $U(1)$ at the scale $|a|$. Below this energy scale the effective theory is Abelian and the coupling does not run. It follows that the theory has two different regimes. When the VEVs are large, $|a| >> \Lambda$, the running coupling is frozen before it has a chance to get large and the theory is weakly coupled at all length scales and semiclassical methods are valid. On the other hand when $|a| \sim \Lambda$ the IR coupling is strong and perturbation theory breaks down. If we think of the Coulomb branch as the complex u-plane, the first regime is the asymptotic region near the point at infinity and the second is the region near the origin.

As in the $\mathcal{N}=4$ theory, to make progress we must concentrate on quantities which are highly constrained by supersymmetry. As in that case these include the spectrum of BPS states. However, although the BPS bound $M = |Z|$ is an exact mass formula in this case also, the $\mathcal{N}=2$ story is complicated by the fact that the central charge can recieve quantum corrections. Indeed we have already commented on the fact that the bare coupling τ appearing in the central charge gets renormalized. The most general possibility for the exact central charge is,

$$Z \;=\; n_E a(u) + n_M a_D(u) \tag{6.3}$$

where a and a_D are *holomorphic* in u and λ. This is an example of the general feature of SUSY gauge theories described earlier, namely that SUSY constrains some observables to be holomorphic functions of the fields and parameters.

The low-energy theory on the Coulomb branch is the $\mathcal{N}=2$ supersymmetric extension of QED. It is characterzed by an effective complexified coupling τ_{eff} which is also constrained to be holomorphic in u and Λ. In the absence of charged matter it has an exact $SL(2,\mathbb{Z})$ duality which acts on the effective coupling τ_{eff} rather than the bare one as in (4.4). This is essentially an extension of the Hodge duality of Maxwell theory to $\mathcal{N}=2$ superspace and it can be implemented by an exact transformation of the path integral. The renormalized BPS mass formula given above is also $SL(2,\mathbb{Z})$ invariant provided that a_D/a transforms in the same way as τ_{eff} under modular transfomations. Although the duality is only a property of the low-energy theory, it plays an important role in the exact solution of the model.

Previously we indicated that a, a_D and τ_{eff} depended holomorphically on u and Λ. However it is not quite precise to say that they are holomorphic functions defined on the u-plane. In particular we should allow for the possiblity that the spectrum changes by an $SL(2,\mathbb{Z})$ transformation as we traverse a closed curve on the u-plane. Thus we should really require that (a,a_D) defines a holomorphic section of a flat $SL(2,\mathbb{Z})$ bundle over the u-plane. In fact, a one-loop perturbative analysis is sufficient to show that this bundle is non-trivial. The classical theory has a $U(1)$ R-symmetry under which the adjoint scalar has charge two. Acting with this generator on a point far from the origin of the Coulomb branch we traverse a large circle on the u-plane. However, at one loop the symmetry suffers from an anomaly. The effect of this is that a $U(1)$ rotation through an angle α is no longer exactly a symmetry of the theory but rather leads to a shift in the effective vacuum angle $\theta_{eff}\to\theta_{eff}+2\alpha$. Via the Witten effect described in the previous section, this shifts the electric charges of $(n_M=1)$ magnetic monopoles in the theory by an amount α/π. Rotation through 2π, accomplished by traversing a large circle in the u-plane, therefore implements the non-trivial $SL(2,\mathbb{Z})$ transformation[5] denoted T^2 in the notation of the previous section. Such a transformation associated with a closed path is known as a *monodromy.*

In summary, the low energy effective action and the BPS spectrum of the theory are determined by the holomorphic quantities $\tau_{eff}(u)$ and $(a(u),a_D(u))$, which together define a holomorphic section of an $SL(2,\mathbb{Z})$ bundle over the u-plane. Like a holomorphic function, this bundle is essentially determined by its behaviour at its singular points in the u-plane. In particular, the allowed singularities are logarithmic cuts which induce $SL(2,\mathbb{Z})$ monodromies associated with closed paths in the u-plane which encircle the singular point. The non-trivial monodromy described in the previous paragraph is associated with a logarithmic branch point at $u=\infty$. The logarithmic divergence as $u\to\infty$ in turn reflects the logarithmic running of the coupling in the UV. In principle the $SL(2,\mathbb{Z})$ bundle can be reconstructed from a knowledge of the monodromies of the BPS spectrum around each singular point. As the branch cut coming from the point at infinity must terminate somewhere there must be at least one singular point in the interior of the u-plane. To make further progress it is necessary to consider the possible physical significance of such points.

The singular points are points at which the couplings in the low-energy effective Lagrangian diverge. This breakdown of the low-energy description is a classic symptom of extra massless degrees of freedom which need to be taken into account. It is reasonable therefore to assume that the singular points in the interior of the u-plane are points at which new massless particles appear. In the classical theory

[5]The real story is slightly more complicated than I have indicated as the transformation also changes the sign of the magnetic charge.

there is only one such point, the point $u = 0$ where all the classical BPS states become massless and the full $SU(2)$ gauge symmetry is restored. One might think that the simplest possibility is that this situation persists in the quantum theory. However, a basic constraint on the solution is that the effective gauge coupling constant $\mathrm{Im}\,\tau_{eff}$ should be positive everywhere to ensure unitarity of the effective action. Using an elementary theorem from complex analysis, Seiberg and Witten were able to rule out scenarios involving only a single singularity at the point $u = 0$.

The next simplest possibility is that of a pair of singularities away from the origin interchanged by the global symmetry $u \to -u$. For various reasons discussed in [**19**], it is hard to associate these points with the restoration of non-Abelian gauge symmetry. However, Seiberg and Witten found an interesting alternative explanation. They considered the possibility of points at which a single BPS state becomes massless. Given the presence of a new light state in the vicinity of the singular point the effective Lagrangian can then be improved by including a charged $\mathcal{N} = 2$ matter multiplet and explicitly evaluating the one-loop running of the effective coupling due to the charged particle. Just as at weak coupling, the perturbative logarithm dictates the $SL(2,\mathbb{Z})$ monodromy associated with contours encircling the singularity. Remarkably, they found a single consistent solution with two singular points at $u = \pm\Lambda^2$. The corresponding massless states at these two singularities are the BPS monopole and the dyon of electric-magnetic charge $(n_E, n_M) = (1, 1)$. The resulting exact formulae for the BPS mass spectrum and effective Lagrangian have been tested precisely in many different ways.

The Seiberg-Witten solution realizes electric-magnetic duality in a quite different way than the $\mathcal{N} = 4$ theory. The magnetic degrees of freedom which are very massive in the semiclassical regime become massless in the vicinity of the singular points. At low energies we have a dual description in terms of an Abelian $\mathcal{N} = 2$ gauge theory with charged matter. In the $\mathcal{N} = 4$ theory, electric-magnetic duality holds at all length scales. The monopoles become massless only in the strong coupling limit and the dual theory in this regime is non-Abelian: it is just the same $\mathcal{N} = 4$ theory again with but with dual coupling $\tau_D = -1/\tau$.

Finally we can return to the issue we hoped to address all along, that of confinement. In particular we can study the effect of soft-breaking of the $\mathcal{N} = 2$ supersymmetry down to $\mathcal{N} = 1$. In terms of $\mathcal{N} = 1$ supersymmetry, the $\mathcal{N} = 2$ vector multiplet splits into an $\mathcal{N} = 1$ vector multiplet containing the gauge field and a single Weyl fermion and a chiral multiplet containing the complex adjoint scalar Φ and its $\mathcal{N} = 1$ superpartner which is another Weyl fermion. Soft breaking of $\mathcal{N} = 2$ supersymmetry down to $\mathcal{N} = 1$ means giving a non-zero mass μ to the adjoint chiral multiplet. At energy scales far below μ, the massive scalar and its superpartner decouple leaving the theory of a single $\mathcal{N} = 1$ vector multiplet. This theory is $\mathcal{N} = 1$ supersymmetric Yang-Mills theory, a theory which is believed to have many features in common with QCD. In particular, it is an asymptotically free theory which runs to strong coupling in the IR. Strong coupling dynamics lead to the dynamical breaking of chiral symmetry by a fermion condensate just as they do in QCD. We can say this with confidence as the so-called gluino condensate can be calculated exactly using the fact that it is a holomorphic function of the parameters [**22**]. As in QCD, we interpret this as the signal of a gauge theory realized in the confining phase. One important difference from QCD is that the non-anomalous chiral symmetry is $\mathbb{Z}_{2N}$ and, as always, spontaneous breaking of a discrete symmetry leads to degenerate vacua. For the pure $\mathcal{N} = 1$ theory with gauge group $SU(N)$, the condensate breaks $\mathbb{Z}_{2N}$ down to $\mathbb{Z}_2$ and we expect to find exactly N degenerate supersymmetric vacuum states.

As soon as we introduce a small non-zero mass μ for the adjoint chiral multiplet, the IR dynamics should be qualitatively similar to the $\mathcal{N} = 1$ super Yang-Mills theory with gauge group $SU(2)$[6]. Hence we expect to find a confining theory with two supersymmetric vacua. However, when μ is small compared to the dynamical scale Λ, we can actually study the dynamics *quantitatively* as a small perturbation of the Seiberg-Witten solution. In the classical theory, the introduction of a mass term for Φ means that this field can no longer acquire a VEV and the only surviving classical vacuum is the one in which the VEV is zero and the non-Abelian symmetry is restored. However, the Seiberg-Witten solution shows that such a vacuum simply does not exist in the full quantum theory. Naively this might suggest that the perturbed theory has no SUSY vacua.

To understand what really happens one must remember to include the effects of the extra light degrees of freedom which appear near the singular points on the u-plane. For example, in the vicinity of the point $u = \Lambda^2$ the low energy theory actually contains three scalar fields. Apart from the modulus field u, we have complex scalar fields M and $\tilde{M}$ whose quanta are the light BPS monopoles. Including these degrees of freedom one finds a supersymmetric vacuum at $u = \Lambda$ with a non-zero value for the VEVs of the monopole fields $\langle M \rangle = \langle \tilde{M} \rangle \sim \sqrt{\mu\Lambda}$. Thus the light monopoles have condensed in the vacuum. This immediately leads to the confinement of electric charges due to the dual Meissner effect! As the dual theory is weakly coupled, one can study the formation of electric flux tubes very explicitly. This is the first explicit demonstration of confinement in a non-Abelian gauge theory in four dimensions.

Another supersymmetric vacuum appears at the other singular point $u = -\Lambda^2$. By similar arguments, one finds that the light dyons condense in the vacuum realizing a phase with oblique confinement. These two SUSY vacua are believed to be continuously connected to the two SUSY vacua of the pure $\mathcal{N} = 1$ theory which emerges as the mass parameter μ is taken to infinity.

7. Conclusion

In these notes we have seen how two old ideas about QCD have been realized very explicitly in the more controlled context of supersymmetric gauge theory. Seiberg-Witten theory conclusively demonstrates that confinement due to the condensation of magnetic monopoles really does occur in four-dimensional field theory. Maldacena's conjecture identifies the large N dual string theory of a four-dimensional non-Abelian gauge theory for the first time. In both cases, we are still a long way from being able to apply these ideas in their original context. However, the author believes that the developments described above represent important steps in the right direction.

The author would like to thank the organisers of the TMR Winter school in Utrecht and the Clay Institute/INI School on "Complex Geometry and String Theory" where these lectures were given.

[6]Quantitative agreement with the $\mathcal{N} = 1$ theory should only be obtained when we fully decouple the extra matter and this requires taking the limit $\mu \to \infty$ holding the dynamical scale on the $\mathcal{N} = 1$ theory fixed.

8. Further Reading

A (very) partial list of references for my lectures is as follows. The classic textbook on $\mathcal{N}=1$ SUSY is the book of Wess and Bagger [**5**]. The basic facts about the representations of extended SUSY, including reduction from higher dimensions are reviewed in Lykken's 1996 TASI lectures [**6**].

Different aspects of $\mathcal{N}=4$ SUSY Yang-Mills are reviewed by different authors. The superconformal phase and the AdS/CFT correspondence are extensively discussed in the review of Aharony et al [**7**]. Another very useful reference for this subject is [**10**]. The relation between theories with 16 supercharges in different dimensions is covered in [**11**].

The original conjecture of electric-magnetic duality in non-Abelian gauge theory was made by Montonen and Olive in [**13**]. Its particular relevance to the $\mathcal{N}=4$ theory was suggested by Osborn [**14**]. Other important ingredients for the S-duality conjecture were supplied in [**12**] (BPS states) and [**15**] (the "Witten effect"). An impressive semiclassical test of S-duality is described by Sen in [**16**]. These developments are reviewed in Harvey's lecture notes [**18**].

The original paper by Seiberg and Witten on $\mathcal{N}=2$ SUSY Yang-Mills is [**19**]. The generalization to include matter is described in their second paper [**20**]. An important earlier work is by Seiberg [**21**], where the general form of the prepotential is derived. Direct semiclassical tests of the Seiberg-Witten solution are given in [**23**] and [**24**]. There are many reviews of Seiberg-Witten theory, but the one which is most relevant to these lectures is Peskin's 1996 TASI lectures [**22**]. This also includes an excellent review of $\mathcal{N}=1$ theories. A more stringy perspective on SW theory is given by Lerche [**25**].

References

1. G. 't Hooft, *A planar diagram theory for the strong interactions*, Nucl. Phys. **B72** (1974) 461.
2. S. R. Coleman, *1/N*, in *C79-07-31.7* SLAC-PUB-2484 Presented at 1979 Int. School of Subnuclear Physics, Pointlike Structures Inside and Outside Hadrons, Erice, Italy, Jul 31-Aug 10, 1979.
3. S. Coleman and J. Mandula, *All Possible Symmetries of the S-matrix*, Phys. Rev. **159** (1967) 1251.
4. S. R. Coleman, *Quantum Sine-Gordon Equation As The Massive Thirring Model*, Phys. Rev. D **11** (1975) 2088.
5. J. Wess and J. Bagger, *Supersymmetry and Supergravity*, Princeton University Press (1992).
6. J. D. Lykken, *Introduction to supersymmetry*, Fields, Strings and duality (1997) 85-153 World Sci. [arXiv:hep-th/9612114].
7. O. Aharony, S. S. Gubser, J. Maldacena, H. Ooguri and Y. Oz, Phys. Rept. **323** (2000) 183 [arXiv:hep-th/9905111].
8. J. M. Maldacena, *The large N limit of superconformal field theories and supergravity*, Adv. Theor. Math. Phys. **2** (1998) 231 [Int. J. Theor. Phys. **38** (1999) 1113] [arXiv:hep-th/9711200].
9. E. Witten, *Anti-de Sitter space and holography*, Adv. Theor. Math. Phys. **2** (1998) 253 [arXiv:hep-th/9802150].
10. K. A. Intriligator, *Bonus symmetries of $N=4$ super-Yang-Mills correlation functions via AdS duality*, Nucl. Phys. B **551** (1999) 575 [arXiv:hep-th/9811047].
11. N. Seiberg, *Notes on theories with 16 supercharges*, Nucl. Phys. Proc. Suppl. **67** (1998) 158 [arXiv:hep-th/9705117].
12. E. Witten and D. I. Olive, *Supersymmetry Algebras That Include Topological Charges*, Phys. Lett. B **78** (1978) 97.

13. C. Montonen and D. I. Olive, *Magnetic Monopoles As Gauge Particles?*, Phys. Lett. B **72** (1977) 117.
14. H. Osborn, *Topological Charges For N=4 Supersymmetric Gauge Theories And Monopoles Of Spin 1*, Phys. Lett. B **83** (1979) 321.
15. E. Witten, *Dyons Of Charge* $e\theta/2\pi$, Phys. Lett. B **86** (1979) 283.
16. A. Sen, *Dyon - monopole bound states, selfdual harmonic forms on the multi-monopole moduli space, and* $SL(2,\mathbb{Z})$ *invariance in string theory*, Phys. Lett. B **329** (1994) 217 [arXiv:hep-th/9402032].
17. G. Segal and A. Selby, *The Cohomology Of The Space Of Magnetic Monopoles*, Commun. Math. Phys. **177** (1996) 775.
18. J. A. Harvey, *Magnetic monopoles, duality, and supersymmetry*, arXiv:hep-th/9603086. Fields, Strings and duality. World Sci. (1997) 157-216.
19. N. Seiberg and E. Witten, *Electric - magnetic duality, monopole condensation, and confinement in N=2 supersymmetric Yang-Mills theory*, Nucl. Phys. B **426** (1994) 19 [Erratum-ibid. B **430** (1994) 485] [arXiv:hep-th/9407087].
20. N. Seiberg and E. Witten, *Monopoles, duality and chiral symmetry breaking in N=2 supersymmetric QCD*, Nucl. Phys. B **431** (1994) 484 [arXiv:hep-th/9408099].
21. N. Seiberg, *Supersymmetry And Nonperturbative Beta Functions*, Phys. Lett. B **206** (1988) 75.
22. M. E. Peskin, *Duality in supersymmetric Yang-Mills theory*, Fields, Strings and duality. World Sci. (1997) 729-809.
23. D. Finnell and P. Pouliot, *Instanton calculations versus exact results in four-dimensional SUSY gauge theories*, Nucl. Phys. B **453** (1995) 225 [arXiv:hep-th/9503115].
24. N. Dorey, V. V. Khoze and M. P. Mattis, *Multi-Instanton Calculus in N=2 Supersymmetric Gauge Theory*, Phys. Rev. D **54** (1996) 2921 [arXiv:hep-th/9603136].
25. W. Lerche, *Introduction to Seiberg-Witten theory and its stringy origin*, Nucl. Phys. Proc. Suppl. **55B** (1997) 83 [Fortsch. Phys. **45** (1997) 293] [arXiv:hep-th/9611190].
26. R. Donagi and E. Witten, *Supersymmetric Yang-Mills Theory And Integrable Systems*, Nucl. Phys. B **460** (1996) 299 [arXiv:hep-th/9510101].
27. C. Vafa and E. Witten, *A Strong coupling test of S duality*, Nucl. Phys. B **431** (1994) 3 [arXiv:hep-th/9408074].

Department of Physics, University of Wales, Swansea, Singleton Park, Swansea, SA2 8PP, UK

E-mail address: n.dorey@swansea.ac.uk

Clay Mathematics Proceedings
Volume 3, 2004

The Geometry of A-branes

Anton Kapustin

Abstract. I discuss the geometry of A-branes in the context of the Homological Mirror Symmetry conjecture. I argue that there exist A-branes which are coisotropic, rather than Lagrangian, submanifolds of a Calabi-Yau. I propose a formal analogue of Floer homology for coisotropic A-branes.

Mirror Symmetry

The goal of my talk is to describe the current state of knowledge about topological D-branes of type A, also known as A-branes. The most important motivation for studying topological D-branes is the Homological Mirror Symmetry conjecture put forward by Maxim Kontsevich. Therefore I will begin by reviewing this conjecture.

A weak Calabi-Yau manifold X is a compact complex manifold of Kähler type whose canonical line bundle is trivial. In physics, weak Calabi-Yau manifolds are usually equipped with additional structure: a Kähler form ω and a B-field $B \in H^2(X,\mathbb{R})/H^2(X,\mathbb{Z})$. The triviality of the canonical line bundle is equivalent to the requirement that the holonomy group of the Kähler metric belong to $SU(n)$, where $n = \dim_{\mathbb{C}} X$. If the holonomy group is *exactly* $SU(n)$, then according to standard terminology X is called a Calabi-Yau manifold. But in this talk, for reasons of brevity, I will use the term "Calabi-Yau" to mean "weak Calabi-Yau."

It is believed that to any weak Calabi-Yau equipped with a Kähler form and a B-field one can naturally associate an $N = 2$ superconformal field theory. An $N = 2$ superconformal field theory is a special kind of quantum field theory in two dimensions which describes propagation of superstrings on X. Its mathematical definition is complicated (see e.g. [**23, 8**]). The most important feature of an $N = 2$ SCFT is that its symmetry algebra contains the direct sum of two copies of the $N = 2$ super-Virasoro algebra with equal central charges. These two copies are referred to as left-moving and right-moving super-Virasoro algebras. In the case of an $N = 2$ SCFT associated to a Calabi-Yau manifold the central charge is equal to $3n$. The $N = 2$ super-Virasoro algebra has a canonical $N = 1$ super-Virasoro subalgebra, and a certain involution, called the mirror automorphism, which acts as the identity on the $N = 1$ subalgebra. If two Calabi-Yaus are complex conjugate,

2000 *Mathematics Subject Classification.* Primary 14J32, 81T45; Secondary 53D40.
Key words and phrases. Mirror symmetry, topological field theory, D-branes.
The author was supported in part by the DOE grant DE-FG03-92-ER40701.

then the corresponding $N = 2$ SCFTs are "almost isomorphic": they are isomorphic as $N = 1$ SCFTs, but the isomorphism induces the mirror automorphism on both copies of the $N = 2$ super-Virasoro algebra. The idea of mirror symmetry is "to take a square root" of the operation of complex conjugation. That is, one says that two Calabi-Yau manifolds are mirror if there exists an isomorphism of the corresponding $N = 1$ SCFTs which acts by the mirror automorphism on one copy of the $N = 2$ super-Virasoro algebra, and by the identity on the other one.

An explicit construction of the $N = 2$ SCFT corresponding to X is currently possible only when X is a complex torus with a flat metric. The reason for this is that $N = 2$ SCFTs are very complicated objects depending on the Ricci-flat metric on X, whose explicit form is unknown. To facilitate the discussion of the mirror relation, E. Witten introduced two "truncated" versions of $N = 2$ SCFT called the A-model and the B-model [**24, 25**]. The truncation procedure used by Witten is known as topological twisting. The main idea of topological twisting is to regard a certain element in the super-Virasoro algebra (supercharge) as a BRST operator, and focus on its cohomology. There are two essentially different choices of the BRST operator, which are exchanged by the mirror automorphism, so one gets two topologically twisted versions of the $N = 2$ SCFT. As indicated by the name, topologically twisted theories are independent of the metric on the 2d world-sheet, i.e. they are topological field theories (TFT). 2d TFTs are much simpler than 2d SCFTs; for a definition and discussion, see e.g. [**2, 6**]. From the mathematical viewpoint, the notion of a 2d TFT is equivalent to the notion of a super-commutative Frobenius algebra, i.e. a super-commutative algebra with an invariant inner product. For example, the algebra corresponding to the A-model of a Calabi-Yau X is the quantum cohomology ring of X, while the algebra corresponding to the B-model is

$$\oplus_{p,q=0}^{n} H^p(\Lambda^q T^{1,0} X).$$

The advantage of the "topologically twisted" viewpoint is that it allows one to treat complex and symplectic aspects of a Calabi-Yau separately. That is, the A-model is independent of the complex structure of X and the $(0, 2)$ and $(2, 0)$ parts of the B-field, while the B-model is independent of the Kähler form ω and the $(1, 1)$ part of the B-field. Since the mirror automorphism exchanges the A and B-twists, mirror symmetry exchanges the A and B-models. This observation is the starting point for relating the counting of holomorphic curves in X with periods of the mirror X'.

In view of the above, mirror symmetry should be regarded as some kind of duality between complex and symplectic manifolds, which implies the equivalence of the B and A-models of the respective manifolds. The origin of this duality was greatly clarified by the Homological Mirror Symmetry Conjecture of M. Kontsevich [**16**]. Kontsevich proposed that the equivalence of B and A-models reflects an equivalence between certain triangulated categories attached to complex and symplectic manifolds. On the complex side, the relevant category is the bounded derived category of coherent sheaves on X, which we will denote $D^b(X)$.[1] On the symplectic side,

[1]This statement is strictly true only if the $(0, 2)$ part of the B-field is zero. Otherwise, one has to consider the bounded derived category of "twisted" sheaves on X [**12, 3**]. It is denoted $D^b(X, B)$.

the relevant category is the so-called derived Fukaya category $DF(X)$. Its definition is due to K. Fukaya and M. Kontsevich and will be sketched below. One can regard these categories as enriched versions of the B and A-models, respectively. It is believed that the Homological Mirror Symmetry Conjecture completely captures the mathematical meaning of the mirror relation. That is, Calabi-Yau manifolds X and X' are mirror if and only if $D^b(X)$ (more precisely, $D^b(X, B)$, see the preceding footnote) is equivalent to the derived Fukaya category $DF(X')$, and vice versa.

Topological D-branes

Now let us discuss the physical interpretation of the Homological Mirror Symmetry Conjecture. From the physical viewpoint, a natural way to enrich the A and B-models is to consider 2d TFTs with boundaries. Recall [**2, 6**] that a 2d TFT assigns to any closed connected one-dimensional manifold (i.e. to a circle) a vector space V, and to any oriented 2d bordism from a collection of k circles to a collection of m circles a linear map from $V^{\otimes k}$ to $V^{\otimes m}$. These maps must satisfy certain compatibility conditions [**2, 6**]. To define a 2d TFT with boundaries, we replace closed connected one-dimensional manifolds with compact connected one-dimensional manifolds with boundaries. There are two such manifolds: a circle and an interval. Thus we have two vector spaces, V and W. Instead of bordisms between collections of circles, we now have bordisms between collections of circles and intervals. Whereas for usual 2d TFTs bordisms are compact oriented 2d manifolds with boundaries, for 2d TFTs with boundaries bordisms are compact oriented 2d manifolds with boundaries and "corners." To any such bordism a 2d TFT assigns a linear map from a tensor product of several copies of V and W to another such tensor product. Whereas ordinary 2d TFT describes propagation of closed strings (circles), a 2d TFT with boundaries describes propagation of open and closed strings (intervals and circles). For a detailed discussion of 2d TFTs with boundaries, see [**21, 17**].

In physics, 2d TFTs are defined by means of a path integral over spaces of fields defined on the 2d world-sheet. In the case of a 2d TFT with boundaries, one must specify boundary conditions on the fields. In most cases there are many different choices of boundary conditions, and this suggests that the two endpoints of the open string should not be treated symmetrically: the boundary conditions may be different on the two endpoints. If we denote the set of boundary conditions for a 2d TFT by I, then we will have a vector space analogous to W for any ordered pair (i, j) of elements of I. This strongly suggests that a 2d TFT with boundaries produces a category, instead of an algebra. Objects of this category are boundary conditions, while morphisms are vector spaces W_{ij}. Using the axioms of the 2d TFT with boundaries, one can show that this is indeed the case [**17**]. One usually refers to boundary conditions for a 2d TFT as topological D-branes. There are also non-topological D-branes, which are boundary conditions for a 2d SCFT which are compatible with the super-Virasoro algebra.

Applying these observations to the A and B-models, we get two categories of topological D-branes, which are called, naturally enough, the categories of A-branes and B-branes. Alternatively, one can view A-branes and B-branes as non-topological D-branes for the underlying $N = 1$ SCFT which are compatible with the A and B-twist, respectively. Topological D-branes were introduced by E. Witten [**26**]. He showed that the category of A-branes does not depend on the complex structure of X, while the category of B-branes is independent of the symplectic structure. One can think of these categories as the enriched versions of the A and B-models: while A and B-models describe closed strings, A and B-models with boundaries describe closed and open strings with all possible boundary conditions. Since the mirror automorphism exchanges the A and B-twists, it is clear that the category of A-branes for X is equivalent to the category of B-branes for its mirror X', and vice versa.

Now we are ready to explain Kontsevich's conjecture in physical terms. According to Witten [**26**], holomorphic vector bundles are examples of B-branes, and spaces of morphisms between vector bundles E and F are global Ext groups $Ext^k(E,F) = H^k(F \otimes E^*)$ [**26**]. Further, Witten showed that examples of A-branes are provided by Lagrangian submanifolds equipped with vector bundles with flat connections, and spaces of morphisms between them are Floer homology groups (the definition of the Floer homology groups is sketched below). Now, Ext groups are spaces of morphisms in the derived category of coherent sheaves, therefore if is natural to conjecture that arbitrary complexes of coherent sheaves are also examples of B-branes, and morphisms between them are morphisms in the derived category. Similarly, it is reasonable to assume that arbitrary "complexes" of Lagrangian submanifolds with flat vector bundles (in the sense explained in [**16**]) are examples of A-branes. The Homological Mirror Symmetry Conjecture is basically the statement that *all* topological B-branes and A-branes arise in this way (as complexes of coherent sheaves or complexes of Lagrangian submanifolds with flat vector bundles). In other words, the Homological Mirror Symmetry Conjecture would follow if we could prove that the category of A-branes (resp. B-branes) is equivalent to $DF(X)$ (resp. $D^b(X)$).

Category of A-branes versus the Fukaya category

What is the current status of the Homological Mirror Symmetry Conjecture? On the mathematical side, a certain modification of the conjecture has been proved in the case when X and X' are elliptic curves [**22**]. Unfortunately the methods of [**22**] do not extend to higher-dimensional Calabi-Yaus. On the physical side, there is now a good understanding of why complexes of coherent sheaves can be regarded as B-branes and why morphisms between B-branes coincide with morphisms in the derived category [**7, 18, 1, 5, 14**]. Similar arguments can be made on the symplectic side to show that objects of the derived Fukaya category are valid A-branes [**19, 20**]. There is also some evidence that the category of B-branes is equivalent to the derived category of (twisted) coherent sheaves [**12**]. On the other hand, it has been shown recently that the Fukaya category is only a full subcategory of the category of A-branes [**13**], i.e. there are A-branes which are not related to Lagrangian submanifolds. This implies that the symplectic side of Kontsevich's conjecture needs substantial modification. In this section we outline the argument of [**13**] showing that in the case of flat tori there are not enough Lagrangian submanifolds for the Homological Mirror Symmetry Conjecture to be true, and that there are non-Lagrangian A-branes.

The idea is to work on the level of K-theory, and to show that for certain mirror pairs X and X' the K-theory of $D^b(X)$ is strictly bigger than the K-theory of the Fukaya category $DF(X')$. In fact, to simplify life, we will tensor K-theory with $\mathbb{Q}$ and use the Chern character to map the rational K-theory to an appropriate cohomology group. In the case of $D^b(X)$, the Chern character takes values in the intersection of $H^*(X,\mathbb{Q})$ and $\oplus_{p=0}^n H^{p,p}(X)$, which are both subgroups of $H^*(X,\mathbb{C})$. (The Hodge conjecture says that the image of the Chern character map coincides with this intersection.) In the case of the Fukaya category, the situation is somewhat less clear. Mirror symmetry maps $H^*(X,\mathbb{C})$ into $H^*(X',\mathbb{C})$, therefore the Chern character for the Fukaya category should take values in some subgroup of $H^*(X,\mathbb{C})$. The only obvious candidate for the Chern character of an object of the Fukaya category is the Poincaré dual of the corresponding Lagrangian submanifold (taken over $\mathbb{Q}$). This guess can be physically motivated, and we will assume it in what follows.

Now we will exhibit an example showing that the Fukaya category is not big enough for the Homological Mirror Symmetry Conjecture to hold for all known mirror pairs. Let E be an elliptic curve, e be an arbitrary point of E, and $\mathrm{End}_e(E)$ be the ring of endomorphisms of E which preserve e. For a generic E we have $\mathrm{End}_e(E) = \mathbb{Z}$, but for certain special E $\mathrm{End}_e(E)$ is strictly larger than $\mathbb{Z}$. Such special E's are called elliptic curves with complex multiplication. One can show that E has complex multiplication if and only if its Teichmüller parameter τ is a root of a quadratic polynomial with integral coefficients. Let E be an elliptic curve with complex multiplication. Consider the Abelian variety $X = E^n$, $n \geq 2$. One can show that for such a variety the dimension of the image of the map

$$ch : K(D^b(X)) \otimes \mathbb{Q} \longrightarrow H^*(X,\mathbb{Q}) \bigcap \oplus_{p=0}^n H^{p,p}(X)$$

is

$$\dim_{\mathbb{Q}} \mathbf{Im}(ch) = \binom{2n}{n}.$$

On the other hand, X is related by mirror symmetry to a symplectic torus X' of real dimension $2n$. Cohomology classes Poincaré-dual to Lagrangian submanifolds in X' lie in the kernel of the map

$$H^n(X',\mathbb{R}) \xrightarrow{\wedge\omega} H^{n+2}(X',\mathbb{R}). \tag{1}$$

This map is an epimorphism, and therefore the dimension of the kernel is equal to $\binom{2n}{n} - \binom{2n}{n+2}$. Thus the image of the Chern character map for the Fukaya category of X' has dimension less than or equal than $\binom{2n}{n} - \binom{2n}{n+2}$. Therefore $K(DF(X'))$ cannot be isomorphic to $K(D^b(X))$.

This leaves us with a question: if not all A-branes are Lagrangian submanifolds, what *are* they? On the level of cohomology, if the Chern character of A-branes does not take values in the kernel of the map Eq. (1), where *does* it take values? In the case of flat tori, we can answer the second question. In this case we know that a mirror torus is obtained by dualizing a Lagrangian sub-torus, and can infer how the cohomology classes transform under this operation. The answer is the following [**10**]. Suppose the original torus is of the form $X = A \times B$, where A and B are Lagrangian sub-tori, and the mirror torus is $X' = \hat{A} \times B$, where $\hat{A}$ is the dual of A. Consider a torus $Z = A \times \hat{A} \times B$. It has two obvious projections π and π' to X and X'. On $A \times \hat{A}$ we also have the Poincaré line bundle whose Chern character will be denoted P. Given a cohomology class $\alpha \in H^*(X,\mathbb{Q})$, we pull it back to Z using π, multiply by P, and then push forward to X' using π'. This gives a cohomology class $\alpha' \in H^*(X', \mathbb{Q})$ which is mirror to α. The requirement that α be in the intersection of $H^*(X, \mathbb{Q})$ and $\oplus_p H^{p,p}(X)$ implies that α' is orthogonal to the class $\exp(i\omega)$ [**10**].[2] Cohomology classes dual to Lagrangian submanifolds satisfy this condition, but one can also give examples of coherent sheaves on X such that α' is not even a middle-dimensional cohomology class. For example, in [**13**] we have constructed a holomorphic line bundle on X such that α' has the form $\alpha' = e^a$, where $a \in H^2(X', \mathbb{Z})$. This suggests that in this case the mirror of the holomorphic line bundle on X is a line bundle on a symplectic torus X' whose Chern character e^a is orthogonal to $e^{i\omega}$. We will see in the next section that this guess is correct.

Coisotropic A-branes

To make further progress in understanding A-branes, we need to rely on physical arguments. As explained above, an A-brane is a boundary condition for an $N = 2$ SCFT which is compatible with the A-twist. In [**13**] we analyzed this condition assuming that an A-brane is a submanifold Y of a Kähler manifold X, and carries a Hermitian line bundle E equipped with a unitary connection d_E. (Our analysis was on the classical level; some comments on possible quantum effects are made below.) We showed that in order for a triple (Y, E, d_E) to be an A-brane, the following three conditions are necessary and sufficient.

(i) Y must be a coisotropic submanifold of X. This means that the restriction of the symplectic form ω to Y must have constant rank, and its kernel is an integrable distribution $\mathcal{L}Y \subset TY$. We will denote by $\mathcal{N}Y$ the quotient bundle $TY/\mathcal{L}Y$. Note that $\omega|_Y$ is only a pre-symplectic form on the vector bundle TY (i.e. it has non-trivial kernel), but it descends to a symplectic form σ on the vector bundle $\mathcal{N}Y$.

(ii) The curvature 2-form $F = (2\pi i)^{-1} d_E^2$, regarded as a bundle map from TY to TY^*, annihilates $\mathcal{L}Y$. (The factor $(2\pi i)^{-1}$ is included to make F a real 2-form with integral periods). This implies that F descends to a section f of $\Lambda^2(\mathcal{N}Y^*)$.

(iii) The symplectic form σ and the skew-symmetric bilinear form f, regarded as bundle maps from $\mathcal{N}Y$ to $\mathcal{N}Y^*$, satisfy $(\sigma^{-1} f)^2 = \mathbf{id}_{\mathcal{N}Y}$. This means that $J = \sigma^{-1} f$ is a complex structure on the bundle $\mathcal{N}Y$.

Let us make some comments on these three conditions. The condition (i) implies the existence of a foliation of Y whose dimension is equal to the codimension of Y in X. It is known as the characteristic foliation of Y. The form σ is a basic 2-form with respect to it. If the characteristic foliation happens to be a fiber bundle with a smooth base Z, then $\mathcal{N}Y$ is simply the pull-back of TZ to Y. In general, $\mathcal{N}Y$ is a foliated vector bundle over the foliated manifold Y, and it makes sense to talk about local sections of $\mathcal{N}Y$ locally constant along the leaves of the foliation. It is useful to think of such sections as vector fields on the generally non-existent quotient manifold Z. In the same spirit, the 2-form σ should be interpreted as a symplectic form on Z. One can summarize the situation by saying that Y is a foliated manifold with a transverse symplectic structure σ. $\mathcal{N}Y$ will be called the transverse tangent bundle of Y.

The condition (ii) says that for any section v of $\mathcal{L}Y$ we have $i_v F = 0$. Since $dF = 0$, this implies that the Lie derivative of F along such v vanishes. In other words, F is also a basic 2-form. In the case when the characteristic foliation is a fibration, this is equivalent to saying that f is a pull-back of a closed 2-form on the base Z. In general, it is useful to think of f as a pull-back from a non-existent Z.

[2]We assume that the B-field vanishes, for simplicity.

The condition (iii) implies, first of all, that f is non-degenerate. Thus f is another transverse symplectic structure on Y. Second, the condition (iii) says that the ratio of the two transverse symplectic structures f and σ is a complex structure on the transverse tangent bundle $\mathcal{N}Y$. In other words, the ratio is a transverse almost complex structure. If the characteristic foliation is a fibration with base Z, then $J = \sigma^{-1} f$ is simply an almost complex structure on Z.

An easy consequence of these conditions is that the dimension of Y must be $n + 2k$, where $n = \frac{1}{2} \dim_{\mathbb{R}} X$, and k is a non-negative integer. k has the meaning of "transverse complex dimension." If $k = 0$, then Y is a Lagrangian submanifold, and the second condition forces F to vanish. (The third condition is vacuous in this case). Another extreme case occurs when $Y = X$ (this is possible only if n is even). In this case the leaves of the characteristic foliation are simply points, the second condition is vacuous, and the third condition says that $\omega^{-1}F$ is an almost complex structure on X.

A less obvious property is that the transverse almost complex structure J is integrable [**13**]. This follows easily from the well-known Gelfand-Dorfman theorem [**9**] which plays an important role in the theory of integrable systems. Thus Y is a transverse complex manifold, i.e. it has complex structure in the directions transverse to the leaves of the characteristic foliation. It is also easy to see that both f and σ have type $(0,2)+(2,0)$ with respect to J. In fact, $f+i\sigma$ is a transverse holomorphic symplectic form on the transverse complex manifold Y.

The somewhat mysterious condition (iii) can be rewritten in several equivalent forms. For example, an equivalent set of conditions is

$$\wedge^r (f + i\sigma) \neq 0, \quad r < k, \quad \wedge^k (f + i\sigma) = 0.$$

Here k is related to the dimension of Y as above. This form is convenient for comparison with the conditions on the Chern character of A-branes on tori explained above. For example, let us set $n = 2$ and let $Y = X$. In this case the above conditions are equivalent to

$$F \wedge \omega = 0, \quad F \wedge F = \omega \wedge \omega.$$

On the level of cohomology, this is equivalent to the condition that the Chern character e^F be orthogonal to $e^{i\omega}$.

It is not yet known how to generalize these considerations to the case when A-branes carry a vector bundle of rank higher than one (except in the case when Y is a Lagrangian submanifold, and E is flat). The difficult part is to find a meaningful generalization of (iii). For general E the curvature 2-form is a form with values in $\mathrm{End}(E)$, and an (almost) complex structure on $\mathcal{N}Y$ does not seem to arise. For example, it is clear that if both (Y, E, d_E) and $(Y, E', d_{E'})$ are rank-one A-branes satisfying the conditions above, then their sum is also a valid A-brane. However, instead of one transverse almost complex structure on Y, we now get two! If a rank-two bundle E on Y does not decompose into a direct sum of two line bundles, it is not clear what conditions on the curvature of d_E ensure that it is an A-brane. In principle, the physical definition of an A-brane as a boundary condition for a sigma-model should enable one to find these conditions, but there are some technical difficulties when the gauge group is non-Abelian.

So far our discussion of A-branes was classical. The main source of worry is the possibility that $N = 2$ super-Virasoro used to define the A and B-twist is broken by quantum anomalies. One can argue that any possible anomaly must arise from non-perturbative effects on the world-sheet, i.e. from the contributions to the path integral from Riemann surfaces in X whose boundaries lie in Y and which cannot be continuously deformed to a point. In the case when Y is Lagrangian, the conditions for the absence of anomalies have been analyzed by K. Hori [**11**]. The result is that there are no anomalies if and only if the so-called Maslov class of Y vanishes. Let us recall how the Maslov class is defined. Let us choose a holomorphic section Ω of the canonical line bundle (which is trivial for Calabi-Yau manifolds). Restricting it to Y, we obtain a nowhere vanishing n-form. On the other hand, we also have a volume form vol on Y, which comes from the Kähler metric on X. This is also a nowhere vanishing n-form on Y, and therefore $\Omega|_Y = h \cdot vol$, where h is a nowhere vanishing complex function on Y. h can be thought of as an element of $H^0(\mathbb{C}^*_Y)$, where $\mathbb{C}^*_Y$ is the sheaf of $\mathbb{C}^*$-valued functions on Y. The standard exponential exact sequence gives a homomorphism from $H^0(\mathbb{C}^*_Y)$ to $H^1(Y, \mathbb{Z})$, and the Maslov class of Y is defined as the image of h under this homomorphism. (Explicitly, the Čech cocycle representing the Maslov class is constructed as follows: choose a good cover of Y, take the logarithm of h on each set of the cover, divide by $2\pi i$, and compare the results on double overlaps). Although the definition of the Maslov class seems to depend both on the complex and symplectic structures on X, in fact it is independent of the choice of complex structure. Note also that if the Maslov class vanishes, the logarithm of h exists as a function, and is unique up to addition of $2\pi i m, m \in \mathbb{Z}$. A Lagrangian submanifold Y together with a choice of the branch of $\log h$ is called a graded Lagrangian submanifold [**16**]. In the Fukaya category, all Lagrangian submanifolds are graded.

For coisotropic Y the condition for anomaly cancellation is not yet known, but there is an obvious guess. Let F be the curvature 2-form of the line bundle on Y, and let the dimension of Y be $n+2k$, as before. It is easy to see that the $(n+2k)$-form $\Omega|_Y \wedge F^k$ is nowhere vanishing, and therefore we have $\Omega|_Y \wedge F^k = h \cdot vol$. where vol is the volume form, and h is an element of $H^0(\mathbb{C}^*_Y)$. We propose that the vanishing of the image of h in $H^1(Y, \mathbb{Z})$ is the condition of anomaly cancellation. It would be interesting to prove or disprove this guess.

In the case when X is a torus with a constant symplectic form, Y is an affine sub-torus, and the curvature 2-form F is constant, one can quantize the sigma-model and verify directly that the $N = 2$ super-Virasoro algebra is preserved on the quantum level. This shows that non-Lagrangian A-branes exist on the quantum level.

Now let us suppose that X is a non-toroidal Calabi-Yau manifold. The analysis of [**13**] applies to such cases just as well, but it is harder to find explicit examples of non-Lagrangian A-branes. If X has even complex dimension, we expect that there exist non-Lagrangian A-branes with $Y = X$, i.e. vector bundles on X. For example, let X be a hyperkähler manifold with three symplectic forms $\omega_1, \omega_2, \omega_3$. If we set $\omega = \omega_1$, we can solve the condition (iii) by setting $F = \omega_1 \cos\theta + \omega_2 \sin\theta$ with $\theta \in \mathbb{R}$. If one can find θ such that F has integral periods, then a line bundle with curvature F is a non-Lagrangian A-brane on X. If X has odd complex dimension, all non-Lagrangian A-branes must have non-zero codimension. As a particularly simple example, let us take X to be a product of a K3 surface and an elliptic curve. If K3 admits a non-Lagrangian A-brane as above (a line bundle), we may consider a product of this A-brane on K3 with a 1-cycle on the elliptic curve. This brane has real codimension 1, and one can easily check that all our conditions are satisfied.

In the most-often discussed case when X is a simply-connected Calabi-Yau 3-fold, there seem to be no non-Lagrangian A-branes of rank one. Indeed, our conditions imply that such an A-brane would have real codimension one, and therefore would be homologically trivial. We suspect that the situation remains the same for A-branes of rank higher than one.

Morphisms between coisotropic A-branes

We hope that we have given convincing arguments that non-Lagrangian A-branes exist, and must be included in order for the Homological Mirror Symmetry Conjecture to be true. Unfortunately, we do not have a proposal for what should replace the Fukaya category. In the remainder of this lecture we will describe ideas which could help to solve this problem..

As we saw above, we know two kinds of A-branes. First, we have objects of the Fukaya category, i.e. triples (Y, E, d_E) where Y is a graded Lagrangian submanifold, E is a trivial vector bundle on Y with Hermitian metric, and d_E is a flat unitary connection of E.[3] Second, we have triples (Y, E, d_E), where (E, d_E) is a Hermitian *line* bundle on Y with a unitary connection, and the conditions (i)-(iii) are satisfied. For objects of the Fukaya category we also know how to compute morphisms and their compositions. Let us try to guess what the recipe should be for objects of the second kind.

To begin with, let us recall how morphisms in the Fukaya category are defined. Suppose we are given two objects (Y_1, E_1, d_1) and (Y_2, E_2, d_2). We will assume that Y_1 and Y_2 intersect transversally at a finite number of points; if this is not the case, we should deform one of the objects by flowing along a Hamiltonian vector field, until the transversality condition is satisfied. Let I be the set of intersection points of Y_1 and Y_2. Now we consider the Floer complex. As a vector space, it is a direct sum of vector spaces

$$V_i = \mathrm{Hom}(E_1(e_i), E_2(e_i)), \quad i \in I.$$

The grading is defined as follows. At any point $p \in Y$ the space T_pY defines a point q in the Grassmannian of Lagrangian planes in T_pX. Let us denote by $\widetilde{Lag}_p$ the universal cover of the Lagrangian Grassmannian of T_pX. On a Calabi-Yau X, these spaces fit into a fiber bundle over X denoted by $\widetilde{Lag}$ [**16**]. Grading of Y provides a canonical lift of q to $\widetilde{Lag}_p$ for all p; these lifts assemble into a section of the restriction of $\widetilde{Lag}$ to Y [**16**]. Thus for each intersection point e_i we have a pair of points $q_1, q_2 \in \widetilde{Lag}_{e_i}$. The grade of the component of the Floer complex corresponding to e_i is the Maslov index of q_1, q_2 (see [**4**] for a definition of the Maslov index.) Finally, we need to define the differential. Let e_i and e_j be a pair of points whose grades differ by one. The component of the Floer differential which maps V_i to V_j is defined by counting holomorphic disks in X with two marked points, so that the two marked points are e_i and e_j (the Maslov index of e_j is the Maslov index of e_i plus one), and the two intervals which make up the boundary of the disks are mapped to Y_1 and Y_2. Note that in order to compute the differential one has to choose an (almost) complex structure J on X such that the form $\omega(\cdot, J\cdot)$ is a Hermitian form on the tangent bundle of X. For a precise definition of the

[3] We assume that the B-field is trivial. The modification needed when $B \neq 0$ is explained in [**12**].

Floer differential, see [**16**]. The space of morphisms in the Fukaya category is defined to be the Floer complex. The composition of morphisms can be defined using holomorphic disks with three marked points and boundaries lying on three Lagrangian submanifolds. It is associative only up to homotopy, given by a triple product of morphisms. Actually, there is an infinite sequence of higher products in the Fukaya category, which are believed to satisfy the identities of an A_∞ category (see [**15**] for a review of A_∞ categories). It is also believed that changing the almost complex structure J gives an equivalent A_∞ category, so that the equivalence class of the Fukaya category is a symplectic invariant.

What we need is a generalization of the Floer complex to non-Lagrangian A-branes. To guess the right construction, the following heuristic viewpoint on the Floer complex (due to A. Floer himself) is very useful. Consider the space of smooth paths in X, which we will denote PX. This space is infinite-dimensional, but let us treat it as if it were a finite-dimensional manifold. Since X is symplectic, we have a natural 1-form α obtained by integrating ω along the path. More precisely, if $\gamma : I \to X$ is a path, and β is a tangent vector to PX at the point γ (i.e. a vector field along $\gamma(t)$), then the value of α on β is defined to be

$$\int_I \omega(\dot{\gamma}(t), \beta(t))dt.$$

Note that the space PX has two natural projections to X, which we denote π_1 and π_2. It is easy to see that $d\alpha = \pi_1^*\omega - \pi_2^*\omega$. Thus α is closed if we restrict it to the subspace of PX consisting of paths beginning and ending on isotropic submanifolds of X.

In particular, let us consider the subspace $PX(Y_1, Y_2)$ consisting of paths which begin at Y_1 and end at Y_2. Since Y_1 and Y_2 are Lagrangian, the restriction of α to $PX(Y_1, Y_2)$ is closed. Thus the operator $d + 2\pi\alpha$ on the space of differential forms on $PX(Y_1, Y_2)$ squares to zero, and we may try to compute the corresponding cohomology groups (in the finite-dimensional case this complex is called the twisted de Rham complex.) Since $PX(Y_1, Y_2)$ is infinite-dimensional, it is not easy to make sense of this. Floer solved this problem by a formal application of Morse theory to this complex. Namely, the Morse-Smale-Witten-Novikov complex in this case is precisely the Floer complex for the pair Y_1, Y_2.

This construction ignores the bundles E_1 and E_2, but it is easy to take them into account. We can pull back the bundles E_1^* and E_2, together with their connections, to $PX(Y_1, Y_2)$ using π_1 and π_2, respectively. This gives us a pair of Hermitian vector bundles on $PX(Y_1, Y_2)$ with unitary flat connections. We tensor them, and then add the 1-form $2\pi\alpha$ to the connection on the tensor product. The resulting connection is still flat, but no longer Hermitian. Finally, we formally apply Morse theory to compute the cohomology of the resulting twisted de Rham complex on $PX(Y_1, Y_2)$. This gives the Floer complex for a pair of objects of the Fukaya category.

Now consider a pair of coisotropic A-branes, instead of a pair of Lagrangian A-branes. We assume that the bundles E_1 and E_2 are line bundles, for reasons discussed above. By $PX(Y_1, Y_2)$ we still denote the space of smooth curves in X beginning at Y_1 and ending on Y_2. The first difficulty, as compared to the Lagrangian case, is that the restriction of α to $PX(Y_1, Y_2)$ is not closed, so we cannot use it to define a complex. The second difficulty is that connections on E_1 and E_2 are not flat, and neither are their pull-backs to $PX(Y_1, Y_2)$. However, these two difficulties cancel each other, as we will see in a moment. So let us proceed as in the Lagrangian case: pull back (E_2, d_2) by π_2, pull back the dual of (E_1, d_1) by π_1, tensor them, and add $2\pi\alpha$ to the connection. The resulting connection on the bundle on $PX(Y_1, Y_2)$ is not flat, but it has the following interesting property. Note that since Y_1 and Y_2 are foliated manifolds with transverse complex structures, so is $PX(Y_1, Y_2)$. A leaf of this foliation consists of all smooth paths in X which begin on a fixed leaf in Y_1 and end on a fixed leaf of Y_2. The codimension of the foliation is finite and equal to the sum of the codimensions of the characteristic foliations of Y_1 and Y_2. The connection on the bundle on $PX(Y_1, Y_2)$ has the following properties:

(a) it is flat along the leaves of the foliation;

(b) its curvature $\pi_2^*(F_2 + i\omega) - \pi_1^*(F_1 + i\omega)$ has type $(2, 0)$ in the transverse directions.

According to property (b), if we consider sections of our bundle which are covariantly constant along the leaves of the foliation, tensored with basic differential forms, then the anti-holomorphic part of the transverse covariant derivative squares to zero. Thus we have a natural substitute for the cohomology of the twisted de Rham complex: the cohomology of the sheaf of sections which are covariantly constant along the leaves and holomorphic in the transverse directions. This is our formal proposal for the space of morphisms between a pair coisotropic A-branes.

This formal proposal must be properly interpreted, before one gets a concrete recipe for computing spaces of morphisms. In the case when Y_1 and Y_2 are Lagrangian submanifolds, the above sheaf becomes the sheaf of covariantly constant sections of a flat line bundle on $PX(Y_1, Y_2)$, and one can interpret its cohomology using Morse-Smale-Witten-Novikov theory. We do not know how to make sense of our formal proposal in general.

The difficulty of defining A-branes and morphisms between them in geometric terms suggests that perhaps the approach based on Floer homology and its generalizations is not the right way to proceed. An analogy which comes to mind is the notion of a holomorphic line bundle, as compared to the general notion of a holomorphic vector bundle. One can study line bundles in terms of their divisors, but this approach does not extend easily to higher rank bundles. Perhaps objects of the Fukaya category, as well as coisotropic A-branes of rank one, are symplectic analogues of divisors, and in order to make progress one has to find a symplectic analogue of the notion of a holomorphic vector bundle (or a coherent sheaf). This analogy is strengthened by the fact that both divisors and geometric representatives of A-branes provide a highly redundant description of objects in the respective categories: line bundles correspond to divisors modulo linear equivalence, while objects of the Fukaya category are unchanged by flows along Hamiltonian vector fields. We believe that a proper understanding of the mirror phenomenon will require a profound change in our viewpoint on the Fukaya category and its generalizations, and that this new viewpoint may be important for symplectic geometry in general.

References

1. P. S. Aspinwall and A. E. Lawrence, *Derived categories and zero-brane stability,* JHEP **0108** (2001), 004 [arXiv:hep-th/0104147].
2. M. F. Atiyah, *Topological quantum field theories,* Publ. Math. IHES **68** (1988).
3. A. Caldararu, S. Katz, E. Sharpe, *D-branes, B-fields, and Ext groups*, arXiv:hep-th/0302099.
4. S. Cappell, R. Lee, and E. Y. Miller, *On the Maslov index*, Comm. Pure and Appl. Math. **47** (1994), 121.

5. D. Diaconescu, *Enhanced D-brane categories from string field theory,* JHEP **0106** (2001), 016 [arXiv:hep-th/0104200].
6. R. Dijkgraaf, *Intersection theory, integrable hierarchies and topological field theory,* arXiv:hep-th/9201003.
7. M. R. Douglas, *D-branes, categories and $N = 1$ supersymmetry,* J. Math. Phys. **42** (2001), 2818 [arXiv:hep-th/0011017].
8. K. Gawędzki, *Lectures On Conformal Field Theory,* preprint IHES-P-97-2.
9. I. M. Gelfand and I. Ya. Dorfman, *The Schouten bracket and Hamiltonian operators,* Funkts. Anal. Prilozh. **14** (1980), 71.
10. V. Golyshev, V. Lunts, and D. Orlov, *Mirror symmetry for Abelian varieties,* J. Alg. Geom. **10** (2001), 433 [arXiv:math.ag/9812003].
11. K. Hori et al., *Mirror Symmetry*, Lectures at the Clay Mathematical School on Mirror Symmetry, Harvard, Spring 2000, Clay Mathematics Monographs, **1** AMS (2003).
12. A. Kapustin and D. Orlov, *Vertex algebras, mirror symmetry, and D-branes: The case of complex tori,* Commun. Math. Phys. **233** (2003), 79 [arXiv:hep-th/0010293].
13. A. Kapustin and D. Orlov, *Remarks on A-branes, mirror symmetry, and the Fukaya category,* to appear in J. Geom. Phys., arXiv:hep-th/0109098.
14. S. Katz and E. Sharpe, *D-branes, open string vertex operators, and Ext groups,* arXiv:hep-th/0208104.
15. B. Keller, *Introduction to A_∞ algebras and modules,* arXiv:math.ra/9910179.
16. M.Kontsevich, *Homological Algebra of Mirror Symmetry*, talk at the International Congress of Mathematicians, Zürich 1994, arXiv:alg-geom/9411018.
17. C. I. Lazaroiu, *On the structure of open-closed topological field theory in two dimensions*, Nucl. Phys. B **603** (2001), 497 [arXiv:hep-th/0010269].
18. C. I. Lazaroiu, *Unitarity, D-brane dynamics and D-brane categories,* JHEP **0112** (2001), 031 [arXiv:hep-th/0102183].
19. C. I. Lazaroiu, *Generalized complexes and string field theory,* JHEP **0106** (2001), 052 [arXiv:hep-th/0102122].
20. C. I. Lazaroiu, *Graded Lagrangians, exotic topological D-branes, and enhanced triangulated categories,* JHEP **0106** (2001), 064 [arXiv:hep-th/0105063].
21. G. Moore and G. B. Segal, unpublished; see also lectures by G. Moore at http://online.itp.ucsb.edu/online/mp01.
22. A. Polishchuk and E. Zaslow, *Categorical mirror symmetry: The Elliptic curve,* Adv. Theor. Math. Phys. **2** (1998), 443 [arXiv:math.ag/9801119].
23. G. B. Segal, *The Definition Of Conformal Field Theory,* in *Differential Geometrical Methods in Theoretical Physics,* Como, 1987, pp. 165-171.
24. E. Witten, *Topological Sigma Models,* Commun. Math. Phys. **118** (1988), 411.
25. E. Witten, *Mirror manifolds and topological field theory,* in *Essays on Mirror Manifolds,* ed. S. T. Yau, International Press, Hong Kong, 1991, pp. 120-158. [arXiv:hep-th/9112056]
26. E. Witten, *Chern-Simons gauge theory as a string theory,* Prog. Math. **133** (1995), 637 [arXiv:hep-th/9207094].

California Institute of Technology, Pasadena, CA 91125, USA
E-mail address: `kapustin@theory.caltech.edu`

Clay Mathematics Proceedings
Volume 3, 2004

Low Energy D-brane Actions

Robert C. Myers

Abstract. The low energy action describing the dynamics of D-branes consists of two parts: the Born-Infeld action and the Wess-Zumino action. This nonlinear action reliably captures the physics of D-branes with great accuracy. One remarkable feature is the appearance of a non-Abelian gauge theory in the description of several (nearly) coincident branes. This non-Abelian structure plays an important role in realizing various geometric effects with D-branes. In particular, the branes' transverse displacements are described by matrix-valued scalar fields and so noncommutative geometry naturally appears in this framework.

1. Introduction

Dirichlet branes have played a central role in all of the major advances in string theory in the past seven years [**78**]. A primary reason for this is that there are many diverse perspectives from which these objects can be studied. Initially D-branes were discovered within perturbative string theory as surfaces which supported open string excitations [**33, 69, 70, 56, 107**]. It was later found that the same systems had a description in terms of gravitational solutions within low energy supergravity [**71**] — see also, *e.g.,* [**105, 92**] for a more extensive review. More recently, however, D-branes have been described in terms of boundary conformal field theory, *e.g.,* [**50, 109, 111**], K theory, *e.g.,* [**132, 96, 133**], derived categories, *e.g.,* [**86, 42, 117**], tachyon condensation, *e.g.,* [**115, 97, 62**] and various approaches from noncommutative geometry, *e.g.,* [**11, 27, 114**]. This paper will discuss the worldvolume action describing the low energy dynamics of D-branes. This approach emerges naturally from the perturbative description of D-branes, but it provides an exceptionally reliable framework to study D-branes and allows one to develop useful physical intuition for these remarkable systems.

One of the most interesting aspects of the physics of D-branes is the appearance of a non-Abelian gauge symmetry when several D-branes are brought together. Of course, we understand that this symmetry emerges through the appearance of new massless states corresponding to open strings stretching between the D-branes [**131**]. Thus, while the number of light degrees of freedom is proportional

2000 *Mathematics Subject Classification.* 81T30, 81T75.

The author thanks David Winters for a thorough proofreading of this manuscript. This research was supported in part by NSERC of Canada and Fonds FCAR du Québec.

to N for N widely separated D-branes, this number grows like N^2 for N coincident D-branes. The non-Abelian symmetry and the rapid growth in massless states are crucial elements in the statistical mechanical entropy counting for black holes [**105, 34, 35, 92**], in Maldacena's conjectured duality between type IIB superstrings in $AdS_5 \times S^5$ and four-dimensional $\mathcal{N}$=4 super-Yang-Mills theory [**2**], and in the development of M(atrix)-theory as a nonperturbative description of M-theory [**120**].

Referring to the worldvolume theory for N (nearly) coincident D-branes as a non-Abelian U(N) gauge theory emphasizes the massless vector states, which one might regard as internal excitations of the branes. Much of the present discussion will focus, however, on the scalar fields describing the transverse displacements of the branes. For coincident branes, these coordinate fields become matrix-valued, appearing in the adjoint representation of the U(N) gauge group. These matrix-valued coordinates then provide a natural framework in which to consider noncommutative geometry. What has become evident, after a detailed study of the worldvolume action governing the dynamics of the non-Abelian U(N) theory [**98, 123**], is that noncommutative geometries appear dynamically in many physical situations in string theory. One finds that D-branes of one dimension metamorphose into branes of a higher dimension through noncommutative configurations. These configurations provide hints of dualities relating gauge theories in different dimensions, and point to a symbiotic relationship between D-branes of all dimensions. The emerging picture is reminiscent of the 'brane democracy' speculations in very early investigations of D-branes [**124, 79**].

One important example of this brane transmogrification is the 'dielectric effect' in which a set of D-branes are polarized into a higher dimensional noncommutative geometry by nontrivial background fields [**98**]. String theory seems to employ this brane expansion mechanism in a variety of circumstances to regulate spacetime singularities [**106, 99, 10, 87**]. So not only do string theory and D-branes provide a natural physical framework for noncommutative geometry, but it also seems that they may provide a surprising realization of the old speculation that noncommutative geometry should play a role in resolving the chaotic short-distance structure of spacetime in quantum gravity.

An outline of this paper is as follows: First, section 2 presents a preliminary discussion of the worldvolume actions governing the low energy physics of a single D-brane (the Abelian case). Then section 3 discusses the extension to the non-Abelian action, relevant for a system of several (nearly) coincident D-branes. Section 4 provides a brief outline of the dielectric effect for D-branes. In appendix A, we give a short discussion of various aspects of noncommutative geometry and fuzzy spheres, which in particular are relevant for the physical system described in the last section.

2. Worldvolume D-brane actions

Within the framework of perturbative string theory, a Dp-brane is a $(p+1)$-dimensional extended surface in spacetime which supports the endpoints of open strings [**78**]. The massless modes of this open string theory form a supersymmetric U(1) gauge theory with a vector A_a, $9-p$ real scalars Φ^i and their superpartner fermions — for the most part, the latter are ignored throughout the following discussion. At leading order, the low-energy action corresponds to the dimensional reduction of that for ten-dimensional U(1) super-Yang-Mills theory. However, as is usual in string theory, there are higher order $\alpha' = \ell_s^2$ corrections — ℓ_s is the string length scale. For constant field strengths, these stringy corrections can be resummed to all orders, and the resulting action takes the Born-Infeld form [**88**]

$$S_{BI} = -T_p \int d^{p+1}\sigma \ \left(e^{-\phi}\sqrt{-\det(P[G+B]_{ab} + \lambda\, F_{ab})}\right) \ , \tag{2.1}$$

where T_p is the Dp-brane tension and λ denotes the inverse of the (fundamental) string tension, *i.e.,* $\lambda = 2\pi\ell_s^2$. This Born-Infeld action describes the couplings of the Dp-brane to the massless Neveu-Schwarz fields of the bulk closed string theory, *i.e.,* the (string-frame) metric $G_{\mu\nu}$, dilaton ϕ and Kalb-Ramond two-form $B_{\mu\nu}$. The symbol $P[\ldots]$ denotes the pull-back of the bulk spacetime tensors to the D-brane worldvolume.

The interactions of the Dp-brane with the massless Ramond-Ramond (RR) fields are incorporated in a second part of the action, the Wess-Zumino term [**45, 89, 57**]

$$S_{WZ} = \mu_p \int P\left[\sum C^{(n)}\, e^{B}\right] e^{\lambda\, F} \ . \tag{2.2}$$

Here $C^{(n)}$ denote the n-form RR potentials. Eq. (2.2) shows that a Dp-brane is naturally charged under the (p+1)-form RR potential, with charge μ_p, and supersymmetry dictates that $\mu_p = \pm T_p$. If we consider the special case of the D0-brane (a point particle), the Born-Infeld action reduces to the familiar worldline action of a point particle, where the action is proportional to the proper length of the particle trajectory. Actually, the string theoretic D0-brane action is not quite this simple geometric action, rather it is slightly embellished with the additional coupling to the dilaton which appears as a prefactor to the standard Lagrangian density. (Note, however, that the tensors B and F drop out of the action since the determinant is implicitly over a one-dimensional matrix.) Turning to the Wess-Zumino action, we see that a D0-brane couples to $C^{(1)}$ (a vector). Then eq. (2.2) reduces to the familiar coupling of a Maxwell field to the worldline of a point particle, *i.e.,*

$$\mu_0 \int P\left[C^{(1)}\right] \simeq q \int A_\mu \frac{dx^\mu}{d\tau} d\tau \ . \tag{2.3}$$

Higher dimensional Dp-branes can also support a flux of $B+F$, which complicates the worldvolume actions above. From eq. (2.2), we see that such a flux allows a Dp-brane to act as a charge source for RR potentials with a lower form degree than p+1 [**45**]. Such configurations represent bound states of D-branes of different dimensions [**131**]. To illustrate this point, let us assume that B vanishes and then we may expand the Wess-Zumino action (2.2) for a D4-brane as

$$\mu_4 \int \left(C^{(5)} + \lambda\, C^{(3)} \wedge F + \frac{\lambda^2}{2} C^{(1)} \wedge F \wedge F\right) \ . \tag{2.4}$$

Hence, the D4-brane is naturally a source for the five-form potential $C^{(5)}$. However, by introducing a worldvolume gauge field with a nontrivial first Chern class, *i.e.,* exciting a nontrivial magnetic flux on the worldvolume, the D4-brane also sources $C^{(3)}$, which is the potential associated with D2-branes. Hence a D4-brane with magnetic flux is naturally interpreted as a D4-D2 bound state. Similarly a D4-brane that supports a gauge field with a nontrivial second Chern class will source the vector potential $C^{(1)}$ and is interpreted as a D4-D0 bound state.

Hence, as already alluded to above, the Born-Infeld action (2.1) has a geometric interpretation, *i.e.,* it is essentially the proper volume swept out by the Dp-brane, which is indicative of the fact that D-branes are actually dynamical objects. This dynamics becomes more evident with an explanation of the static gauge choice implicit in constructing the above action. To begin, we employ spacetime diffeomorphisms to position the worldvolume on a fiducial surface defined as $x^i = 0$ with $i = p+1, \ldots, 9$. With worldvolume diffeomorphisms, we then match the worldvolume coordinates with the remaining spacetime coordinates on this surface, $\sigma^a = x^a$ with $a = 0, 1, \ldots, p$. Now the worldvolume scalars Φ^i play the role of describing the transverse displacements of the D-brane, through the identification

$$x^i(\sigma) = 2\pi\ell_s^2 \Phi^i(\sigma) \qquad \text{with } i = p+1, \ldots, 9. \tag{2.5}$$

With this identification the general formula for the pull-back reduces to

$$\begin{aligned} P[E]_{ab} &= E_{\mu\nu} \frac{\partial x^\mu}{\partial \sigma^a} \frac{\partial x^\nu}{\partial \sigma^b} \\ &= E_{ab} + \lambda\, E_{ai}\, \partial_b \Phi^i + \lambda\, E_{ib}\, \partial_a \Phi^i + \lambda^2\, E_{ij} \partial_a \Phi^i \partial_b \Phi^j \ . \end{aligned} \tag{2.6}$$

In this way, the expected kinetic terms for the scalars emerge to leading order in an expansion of the Born-Infeld action (2.1). Note that our conventions are such that both the gauge fields and worldvolume scalars have the dimensions of $length^{-1}$ — hence the appearance of the string scale in eq. (2.5).

Although it was mentioned above, we want to stress that these worldvolume actions are low energy effective actions for the massless states of the open and closed strings, which incorporate interactions from all disk amplitudes (all orders of tree level for the open strings). The Born-Infeld action was originally derived [**88**] using standard beta function techniques applied to worldsheets with a boundary [**1, 21, 22**]. In principle, they could also be derived from a study of open and closed string scattering amplitudes and it has been verified that this approach yields the same interactions to leading order [**52, 51, 64**]. As a low energy effective action then, eqs. (2.1) and (2.2) include an infinite number of stringy corrections, which essentially arise through integrating out the massive modes of the string — see the discussion in [**59**]. For example, consider the Born-Infeld action evaluated for a flat Dp-brane in empty Minkowski space

$$\begin{aligned} S_{BI} &\simeq -T_p \int d^{p+1}x\ \sqrt{-\det(\eta_{ab} + 2\pi\ell_s^2\, F_{ab})} \\ &\simeq -T_p \int d^{p+1}x\ \left(1 + \frac{(2\pi\ell_s^2)^2}{4} F^2 + (2\pi\ell_s^2)^4 F^4 + (2\pi\ell_s^2)^6 F^6 + \cdots\right) \end{aligned} \tag{2.7}$$

where the precise structure of the F^4 and F^6 terms may be found in [**126, 60**] and [**85**], respectively. Hence, as well as the standard kinetic term for the worldvolume gauge field, eq. (2.7) includes an infinite series of higher dimension interactions, which are suppressed as long as the typical components $\ell_s^2 F_{ab}$ (referred to an orthonormal frame) are small. The square-root expression in the first line resums this infinite series and so one may consider arbitrary values of $\ell_s^2 F_{ab}$ in working with this action. However, the full effective action would also include stringy correction terms involving derivatives of the field strength, *e.g.,* $\partial_a F_{bc}$ or $\partial_a \partial_b F_{cd}$ — see, for example, [**85**] or [**134**]. None of these have been incorporated in the Born-Infeld action and so one must still demand that the variations in the field strength are relatively small, *e.g.,* components of $\ell_s \partial_a F_{bc}$ are much smaller than those of F_{ab}.

Of course, this discussion extends in the obvious way to derivatives of the scalar fields Φ^i.

At this point, we should also note that the bulk supergravity fields appearing in eqs. (2.1) and (2.2) are in general functions of all of the spacetime coordinates, and so they are implicitly functionals of the worldvolume scalars. In static gauge, the bulk fields are evaluated in terms of a Taylor series expansion around the fiducial surface $x^i = 0$. For example, the metric functional appearing in the D-brane action would be given by

$$
\begin{aligned}
G_{\mu\nu} &= \exp\left[\lambda\Phi^i\,\partial_{x^i}\right] G^0_{\mu\nu}(\sigma^a, x^i)|_{x^i=0} \\
&= \sum_{n=0}^{\infty} \frac{\lambda^n}{n!}\,\Phi^{i_1}\cdots\Phi^{i_n}\,(\partial_{x^{i_1}}\cdots\partial_{x^{i_n}})G^0_{\mu\nu}(\sigma^a, x^i)|_{x^i=0}\ .
\end{aligned}
\tag{2.8}
$$

Hence the worldvolume action implicitly incorporates an infinite class of higher dimension interactions involving derivatives of the bulk fields as well. However, beyond this class of interactions incorporated in eqs. (2.1) and (2.2), once again the full effective action includes other higher derivative bulk field corrections [**57, 24, 8, 31, 32, 113**]. It is probably fair to say that the precise domain of validity of the D-brane action from the point of view of the bulk fields is poorly understood.

3. Non-Abelian D-brane action

As N parallel D-branes approach each other, the ground state modes of strings stretching between the different D-branes become massless. These extra massless states carry the appropriate charges to fill out representations under a U(N) symmetry. Hence the $\mathrm{U(1)^N}$ of the individual D-branes is enhanced to the non-Abelian group U(N) for the coincident D-branes [**131**]. The vector A_a becomes a non-Abelian gauge field

$$
A_a = A_a^{(n)} T_n\ , \qquad F_{ab} = \partial_a A_b - \partial_b A_a + i[A_a, A_b]\ , \tag{3.1}
$$

where T_n are N^2 hermitian generators with $\mathrm{Tr}(T_n\,T_m) = \mathrm{N}\,\delta_{nm}$. Central to the following is that the scalars Φ^i are also matrix-valued, transforming in the adjoint of U(N). The covariant derivative of the scalar fields is given by

$$
D_a\Phi^i = \partial_a\Phi^i + i[A_a, \Phi^i]\ . \tag{3.2}
$$

Understanding how to accommodate this U(N) gauge symmetry in the worldvolume action is an interesting puzzle. For example, the geometric meaning (or even the validity) of eq. (2.5) or (2.8) seems uncertain when the scalars on the right hand side are matrix-valued. In fact, the identification of the scalars as transverse displacements of the branes does remain roughly correct. Some intuition comes from the case where the scalars are commuting matrices and the gauge symmetry can be used to simultaneously diagonalize all of them. In this case, one interprets the N eigenvalues of the diagonal Φ^i as representing the displacements of the N constituent D-branes — see, *e.g.,* [**120**]. Further the gauge symmetry may be used to simultaneously interchange any pair of eigenvalues in each of the scalars and so to ensure that the branes are indistinguishable. Of course, to describe noncommutative geometries, we will be more interested in the case where the scalars do not commute and so cannot be simultaneously diagonlized.

In refs. [**98**] and [**123**], progress was made in constructing the worldvolume action describing the dynamics of non-Abelian D-branes. The essential strategy in both of these papers was to construct an action which was consistent with the familiar string theory symmetry of T-duality [**55**]. Acting on D-branes, T-duality acts to change the dimension of the worldvolume [**78**]. The two possibilities are: *(i)* if a coordinate transverse to the Dp-brane, *e.g.,* $y = x^{p+1}$, is T-dualized, it becomes a D$(p+1)$-brane where y is now the extra worldvolume direction; and *(ii)* if a worldvolume coordinate on the Dp-brane, *e.g.,* $y = x^p$, is T-dualized, it becomes a D$(p-1)$-brane where y is now an extra transverse direction. Under these transformations, the roles of the corresponding worldvolume fields change according to

$$(i)\ \Phi^{p+1} \rightarrow A_{p+1}\,, \qquad (ii)\ A_p \rightarrow \Phi^p\,, \tag{3.3}$$

while the other scalars, Φ^i, and the remaining components of A are left unchanged. Hence, in constructing the non-Abelian action, one can begin with the D9-brane theory, which contains no scalars since the worldvolume fills the entire spacetime. In this case, the non-Abelian extension of eqs. (2.1) and (2.2) is given by simply introducing an overall trace over gauge indices of the non-Abelian field strengths appearing in the action [**41, 40**]. Then applying T-duality transformations on $9-p$ directions yields the non-Abelian action for a Dp-brane. Of course, in this construction, one also T-dualizes the background supergravity fields according to the known transformation rules [**55, 18, 19, 20, 16, 94**]. As in the Abelian theory, the result for the non-Abelian action has two distinct pieces [**98, 123**]: the Born-Infeld term

$$\begin{aligned} S_{BI} &= -T_p \int d^{p+1}\sigma\, \mathrm{STr}\left(e^{-\phi}\sqrt{\det(Q^i{}_j)}\right. \\ &\qquad \left.\times\sqrt{-\det\left(P\left[E_{ab}+E_{ai}(Q^{-1}-\delta)^{ij}E_{jb}\right]+\lambda\,F_{ab}\right)}\right), \end{aligned} \tag{3.4}$$

with $E_{\mu\nu} = G_{\mu\nu} + B_{\mu\nu}$ and $Q^i{}_j \equiv \delta^i{}_j + i\lambda\,[\Phi^i,\Phi^k]\,E_{kj}$; and the Wess-Zumino term

$$S_{WZ} = \mu_p \int \mathrm{STr}\left(P\left[e^{i\lambda\,\mathrm{i}_\Phi \mathrm{i}_\Phi}(\sum C^{(n)}\,e^B)\right]e^{\lambda\,F}\right)\,. \tag{3.5}$$

Let us enumerate the non-Abelian features of this action:

1. *Non-Abelian field strength*: The F_{ab} appearing explicitly in both terms is now non-Abelian, of course.

2. *Non-Abelian Taylor expansion*: The bulk supergravity fields are again functions of all of the spacetime coordinates, and so they are implicitly functionals of the non-Abelian scalars. In the action given by eqs. (3.4) and (3.5), these bulk fields are again interpreted in terms of a Taylor expansion as in eq. (2.8); however the transverse displacements are now matrix-valued.

3. *Non-Abelian Pullback*: As was noted in refs. [**72, 39**], the pullbacks of various spacetime tensors to the worldvolume must now involve covariant derivatives of the non-Abelian scalars in order to be consistent with the U(N) gauge symmetry. Hence eq. (2.6) is replaced by

$$P[E]_{ab} = E_{ab} + \lambda\,E_{ai}\,D_b\Phi^i + \lambda\,E_{ib}\,D_a\Phi^i + \lambda^2\,E_{ij}D_a\Phi^iD_b\Phi^j\,. \tag{3.6}$$

4. *Non-Abelian Interior Product*: In the Wess-Zumino term (3.5), i_Φ denotes the interior product with Φ^i regarded as a vector in the transverse space, *e.g.*, acting

on an n-form $C^{(n)} = \frac{1}{n!} C^{(n)}_{\mu_1 \cdots \mu_n} dx^{\mu_1} \cdots dx^{\mu_n}$, we have

$$\mathrm{i}_\Phi \mathrm{i}_\Phi C^{(n)} = \frac{1}{2(n-2)!} \, [\Phi^i, \Phi^j] \, C^{(n)}_{ji\mu_3 \cdots \mu_n} dx^{\mu_3} \cdots dx^{\mu_n} \; . \tag{3.7}$$

Note that when acting on forms, the interior product is an anticommuting operator and hence for an ordinary vector (*i.e.*, a vector v^i with values in $\mathbb{R}^{9-p}$): $\mathrm{i}_v \mathrm{i}_v C^{(n)} = 0$. It is only because the scalars Φ are matrix-valued that eq. (3.7) yields a nontrivial result.

5. *Non-Abelian Gauge Trace*: As is evident above, both parts of the action are highly nonlinear functionals of the non-Abelian fields, and so eqs. (3.4) and (3.5) would be incomplete without a precise definition for the ordering of these fields under the gauge trace. Above, STr denotes the maximally symmetric trace [**127**]. To be precise, the trace includes a symmetric average over all orderings of F_{ab}, $D_a\Phi^i$, $[\Phi^i, \Phi^j]$ and the individual Φ^k appearing in the non-Abelian Taylor expansions of the background fields. This choice matches that inferred from Matrix theory [**121**], and a similar symmetrization arises in the leading order analysis of the boundary beta functions [**39**]. However, we should note that with this definition an expansion of the Born-Infeld term (2.1) does agree with the string theory to fourth order in F [**127, 128**], but it does not seem to capture the full physics of the non-Abelian fields in the infrared limit at higher orders [**65**] — we expand on this point below.

Some general comments on the non-Abelian action are as follows: In the Born-Infeld term (3.4), there are now two determinant factors as compared to one in the Abelian action (2.1). The second determinant in eq. (3.4) is a slightly modified version of that in eq. (2.1). One might think of this as the kinetic factor, since to leading order in the low energy expansion it yields the familiar kinetic terms for the gauge field and scalars. In the same way, one can think of the new first factor as the potential factor, since to leading order in the low energy expansion it reproduces the non-Abelian scalar potential expected for the super-Yang-Mills theory — see eq. (4.1) below. Further, note that the first factor reduces to simply one when the scalar fields are commuting, even for general background fields.

The form of the action and in particular the functional dependence of the bulk fields on the adjoint scalars can be verified in a number of independent ways. Douglas [**41, 40**] observed on general grounds that, whatever their form, the non-Abelian worldvolume actions should contain a single gauge trace, as do both eqs. (3.4) and (3.5). This observation stems from the fact that the action should encode only the low energy interactions derivable from disk amplitudes in superstring theory. Since the disk has a single boundary, the single gauge trace arises from the standard open string prescription of tracing over Chan-Paton factors on each worldsheet boundary. Further then, one may note [**41, 40**] that the only difference in the superstring amplitudes between the U(1) and the U(N) theories is that the amplitudes in the latter case are multiplied by an additional trace of Chan-Paton factors. Hence, up to commutator 'corrections', the low energy interactions should be the same in both cases. Hence, since the background fields are functionals of the neutral U(1) scalars in the Abelian theory, they must be precisely the same functionals of the adjoint scalars in the non-Abelian theory, up to commutator corrections. The interactions involving the non-Abelian inner product in the Wess-Zumino action (3.5) provide one class of commutator corrections. The functional dependence on the adjoint scalars also agrees with the linearized couplings for the bulk fields derived from Matrix theory [**121**]. As an aside, we would add that the technology developed in Matrix theory remains useful in gaining intuition and manipulating these non-Abelian functionals [**80, 122, 100, 101**]. Finally we add that by the direct examination of string scattering amplitudes using the methods of refs. [**52**] and [**64**], one can verify at low orders the form of the non-Abelian interactions in eqs. (3.4) and (3.5), including the appearance of the new commutator interactions in the non-Abelian Wess-Zumino action [**53, 54**].

As noted above, the symmetric trace perscription is known *not* to agree with the full effective string action [**65**]. Rather, at sixth order and higher in the worldvolume field strength, additional terms involving commutators of field strengths must be added to the action [**9**]. The source of this shortcoming is clear. Recall from the discussion following eq. (2.7) that in the Abelian action we have disgarded all interactions involving derivatives of the field strength. However, this prescription is ambiguous in the non-Abelian theory since

$$[D_a, D_b]F_{cd} = i[F_{ab}, F_{cd}] \ . \tag{3.8}$$

It is clear that with the symmetric trace we have eliminated all derivative terms, including those antisymmetric combinations that might contribute commutators of field strengths. One might choose to improve the action by reinstating these commutators. This problem has been extensively studied and the commutator corrections at order F^6 are known [**9**]. Considering the supersymmetric extension of the non-Abelian action [**38, 116, 83, 15, 85**] has lead to a powerful iterative technique relying on stable holomorphic bundles [**84, 36, 25**] which seems to provide a constructive approach to determine the entire effective open string action, including all higher derivative terms and fermion contributions as well. A similar iterative procedure seems to emerge from studying the Seiberg-Witten map [**114**] in the context of a noncommutative worldvolume theory [**30**].

As described below eq. (2.2), an individual Dp-brane couples not only to the RR potential with form degree $n = p + 1$, but also to the RR potentials with $n = p - 1, p - 3, \ldots$ through the exponentials of B and F appearing in the Wess-Zumino action (2.2). Above in eq. (3.5), $\mathrm{i}_\Phi\mathrm{i}_\Phi$ is an operator of form degree –2, and so worldvolume interactions appear in the non-Abelian action (3.5) involving the higher RR forms. Hence in the non-Abelian theory, a Dp-brane can also couple to the RR potentials with $n = p + 3, p + 5, \ldots$ through the additional commutator interactions. To make these couplings more explicit, consider the D0-brane action (for which F vanishes):

$$\begin{aligned} S_{CS} \;=\; \mu_0 \int \mathrm{STr}\Bigg(P\Bigg[& C^{(1)} + i\lambda\, \mathrm{i}_\Phi\mathrm{i}_\Phi \left(C^{(3)} + C^{(1)}B \right) \\ & -\frac{\lambda^2}{2}(\mathrm{i}_\Phi\mathrm{i}_\Phi)^2 \left(C^{(5)} + C^{(3)}B + \frac{1}{2}C^{(1)}B^2 \right) \\ & -i\frac{\lambda^3}{6}(\mathrm{i}_\Phi\mathrm{i}_\Phi)^3 \left(C^{(7)} + C^{(5)}B + \frac{1}{2}C^{(3)}B^2 + \frac{1}{6}C^{(1)}B^3 \right) \\ & +\frac{\lambda^4}{24}(\mathrm{i}_\Phi\mathrm{i}_\Phi)^4 \left(C^{(9)} + C^{(7)}B + \frac{1}{2}C^{(5)}B^2 + \frac{1}{6}C^{(3)}B^3 + \frac{1}{24}C^{(1)}B^4 \right) \Bigg] \Bigg) . \end{aligned} \tag{3.9}$$

Of course, these interactions are reminiscent of those appearing in Matrix theory [**11, 12**]. For example, eq. (3.9) includes a linear coupling to $C^{(3)}$, which is the potential corresponding to D2-brane charge,

$$i\lambda\,\mu_0 \int \mathrm{Tr}\, P\left[\mathrm{i}_\Phi \mathrm{i}_\Phi C^{(3)}\right]$$

$$= i\frac{\lambda}{2}\mu_0 \int dt\ \mathrm{Tr}\left(C^{(3)}_{tjk}(\Phi,t)\,[\Phi^k,\Phi^j] + \lambda C^{(3)}_{ijk}(\Phi,t)\,D_t\Phi^i\,[\Phi^k,\Phi^j]\right)\ , \tag{3.10}$$

where we assume that $\sigma^0 = t$ in static gauge. Note that the first term on the right hand side has the form of a source for D2-brane charge. This is essentially the interaction central to the construction of D2-branes in Matrix theory with the large N limit [**11, 12**]. Here, however, with finite N, this term would vanish upon taking the trace if $C^{(3)}_{tjk}$ was simply a function of the worldvolume coordinate t (since $[\Phi^k,\Phi^j] \in \mathrm{SU(N)}$). However, in general these three-form components are functionals of Φ^i. Hence, while there would be no 'monopole' coupling to D2-brane charge, nontrivial expectation values of the scalars can give rise to couplings to an infinite series of higher 'multipole' moments [**80, 122**].

4. Dielectric Branes

In this section, we wish to consider certain physical effects arising from the new non-Abelian interactions in the worldvolume action, given by eqs. (3.4) and (3.5). To begin, consider the scalar potential for Dp-branes in flat space, *i.e.,* $G_{\mu\nu} = \eta_{\mu\nu}$ with all other fields vanishing. In this case, the entire scalar potential originates in the Born-Infeld term (3.4) as

$$V = T_p\,\mathrm{Tr}\sqrt{\det(Q^i{}_j)} = \mathrm{N}T_p - \frac{T_p\lambda^2}{4}\mathrm{Tr}([\Phi^i,\Phi^j]\,[\Phi^i,\Phi^j]) + \dots \tag{4.1}$$

The commutator-squared term corresponds to the potential for ten-dimensional U(N) super-Yang-Mills theory reduced to $p+1$ dimensions. A nontrivial set of extrema of this potential is given by taking the $9-p$ scalars as constant commuting matrices, *i.e.,*

$$[\Phi^i,\Phi^j] = 0 \tag{4.2}$$

for all i and j. Since they are commuting, the Φ^i may be simultaneously diagonalized and, as discussed above, the eigenvalues are interpreted as the separated positions in the transverse space of N fundamental Dp-branes. This solution reflects the fact that a system of N parallel Dp-branes is supersymmetric, and so they can sit in static equilibrium with arbitrary separations in the transverse space [**78**].

From the results described in the previous section, it is clear that in going from flat space to general background fields, the scalar potential is modified by new interactions and so one should reconsider the analysis of the extrema. It turns out that this yields an interesting physical effect that is a precise analog for D-branes of the dielectric effect in ordinary electromagnetism. That is, when Dp-branes are placed in a nontrivial background field for which the Dp-branes would normally be regarded as neutral, *e.g.,* nontrivial $F^{(n)}$ with $n > p+2$, new terms will be induced in the scalar potential, and generically one should expect that there will be new extrema beyond those found in flat space, *i.e.,* eq. (4.2). In particular, there can be nontrivial extrema with noncommuting expectation values of the Φ^i, *e.g.,* with $\mathrm{Tr}\,\Phi^i = 0$ but $\mathrm{Tr}(\Phi^i)^2 \neq 0$. This would correspond to the external fields 'polarizing' the Dp-branes to expand into a (higher dimensional) noncommutative worldvolume geometry. This is the analog of the familiar electromagnetic process where an external field may induce a separation of charges in neutral materials. In this case, the polarized material will then carry an electric dipole (and possibly higher multipoles). The latter is also seen in the D-brane analog. When the worldvolume theory is at a noncommutative extremum, the gauge traces of products of scalars will be nonvanishing in various interactions involving the supergravity fields. Hence, at such an extremum, the Dp-branes act as sources for the latter bulk fields.

To make these ideas explicit, we will now illustrate the process with a simple example. We consider N D0-branes in a constant background RR field $F^{(4)}$, *i.e.,* the field strength associated with D2-brane charge. We find that the D0-branes expand into a noncommutative two-sphere which represents a spherical bound state of a D2-brane and N D0-branes.

Consider a background where only the RR four-form field strength is nonvanishing, with

$$F^{(4)}_{tijk} = -2f\varepsilon_{ijk} \qquad \text{for } i,j,k \in \{1,2,3\}\ , \tag{4.3}$$

where f is a constant (of dimensions $length^{-1}$). Since $F^{(4)} = dC^{(3)}$, we must consider the coupling of the D0-branes to the RR three-form potential, which is given above in eq. (3.10). If one explicitly introduces the non-Abelian Taylor expansion (2.8), one finds that the leading order interaction may be written as

$$\frac{i}{3}\lambda^2\mu_0 \int dt\, \mathrm{Tr}\left(\Phi^i\Phi^j\Phi^k\right) F^{(4)}_{tijk}(t)\ . \tag{4.4}$$

This final form might have been anticipated since one should expect that the worldvolume potential can only depend on gauge invariant expressions of the background field. Given that we are considering a constant background $F^{(4)}$, the higher order terms implicit in eq. (3.10) will vanish, as they can only involve spacetime derivatives of the four-form field strength. Combining eq. (4.4) with the leading order Born-Infeld potential (4.1) yields the scalar potential of interest for the present problem

$$V(\Phi) = \mathrm{N}T_0 - \frac{\lambda^2 T_0}{4}\mathrm{Tr}([\Phi^i,\Phi^j]^2) - \frac{i}{3}\lambda^2\mu_0 \mathrm{Tr}\left(\Phi^i\Phi^j\Phi^k\right) F^{(4)}_{tijk}(t)\ . \tag{4.5}$$

Substituting in the (static) background field (4.3) and $\mu_0 = T_0$, the extremisation condition $\delta V(\Phi)/\delta\Phi^i = 0$ yields

$$0 = [[\Phi^i,\Phi^j],\Phi^j] + i\, f\varepsilon_{ijk}[\Phi^j,\Phi^k]\ . \tag{4.6}$$

Note that commuting matrices (4.2) describing separated D0-branes still solve this equation. The value of the potential for these solutions is simply $V_0 = \mathrm{N}T_0$, the mass of N D0-branes. Another interesting solution of eq. (4.6) is

$$\Phi^i = \frac{f}{2}\,\alpha^i \tag{4.7}$$

where the α^i are any N×N matrix representation of the SU(2) algebra

$$[\alpha^i,\alpha^j] = 2i\,\varepsilon_{ijk}\,\alpha^k\ . \tag{4.8}$$

For the moment, let us focus on the irreducible representation for which one finds

$$\mathrm{Tr}[(\alpha^i)^2] = \frac{\mathrm{N}}{3}(\mathrm{N}^2-1) \quad \text{for } i=1,2,3. \tag{4.9}$$

Now evaluating the value of the potential (4.5) for this new solution yields

$$V_{\rm N} = {\rm N}T_0 - \frac{T_0\lambda^2 f^2}{6}\sum_{i=1}^{3}{\rm Tr}[(\Phi^i)^2] = {\rm N}T_0 - \frac{\pi^2\ell_s^3 f^4}{6g}{\rm N}^3\left(1-\frac{1}{{\rm N}^2}\right) \tag{4.10}$$

using $T_0 = 1/(g\ell_s)$. Hence the noncommutative solution (4.7) has lower energy than a solution of commuting matrices, and so the latter configuration of separated D0-branes is unstable towards condensing out into this noncommutative solution. One can also consider reducible representations of the SU(2) algebra (4.8); however, one finds that the corresponding energy is always larger than that in eq. (4.10). Hence it seems that the irreducible representation describes the ground state of the system.

Geometrically, one can recognize the SU(2) algebra as that corresponding to the noncommutative or fuzzy two-sphere [**68, 37, 90**]. The physical size of the fuzzy two-sphere is given by

$$R = \lambda\left(\sum_{i=1}^{3}{\rm Tr}[(\Phi^i)^2]/{\rm N}\right)^{1/2} = \pi\ell_s^2 f{\rm N}\left(1-\frac{1}{{\rm N}^2}\right)^{1/2} \tag{4.11}$$

in the ground state solution. From the Matrix theory construction of Kabat and Taylor [**23**], one can infer that this ground state is not simply a spherical arrangement of D0-branes, rather the noncommutative solution actually represents a spherical D2-brane with N D0-branes bound to it. In the present context, the latter can be verified by seeing that this configuration has a 'dipole' coupling to the RR four-form field strength. The precise form of this coupling is calculated by substituting the noncommutative scalar solution (4.7) into the worldvolume interaction (4.4), which yields

$$-\frac{R^3}{3\pi g\ell_s^3}\left(1-\frac{1}{{\rm N}^2}\right)^{-1/2}\int dt\, F^{(4)}_{t123} \tag{4.12}$$

for the ground state solution. Physically this $F^{(4)}$-dipole moment arises because antipodal surface elements on the sphere have the opposite orientation and so form small pairs of separated membranes and anti-membranes. Of course, the spherical configuration carries no net D2-brane charge.

Given that the noncommutative ground state solution corresponds to a bound state of a spherical D2-brane and N D0-branes, one might attempt to match the above results using the dual formulation. That is, this system can be analyzed from the point of view of the (Abelian) worldvolume theory of a D2-brane. In this case, one would consider a spherical D2-brane carrying a flux of the U(1) gauge field strength representing the N bound D0-branes and, at the same time, sitting in the background of the constant RR four-form field strength (4.3). In fact, one does find stable static solutions, but what is more surprising is how well the results match those calculated in the framework of the D0-branes. The results for the energy, radius and dipole coupling are the same as in eqs. (4.10), (4.11) and (4.12), respectively, except that the factors of $(1-1/{\rm N}^2)$ are absent [**98**]. Hence, for large N, the two calculations agree up to $1/{\rm N}^2$ corrections.

One expects that the D2-brane calculations would be valid when $R \gg \ell_s$ while naively the D0-brane calculations would be valid when $R \ll \ell_s$. Hence it appears there is no common domain where the two pictures can both produce reliable results. However, a more careful consideration of range of validity of the D0-brane calculations only requires that $R \ll \sqrt{\rm N}\ell_s$. This estimate is found by requiring that the scalar field commutators appearing in the full non-Abelian potential (4.1) are small so that the Taylor expansion of the square root converges rapidly. Hence, for large N, there is a large domain of overlap where both of the dual pictures are reliable. Note that the density of D0-branes on the two-sphere is $\mathrm{N}/(4\pi R^2)$. However, even if R is macroscopic it is still bounded by $R \ll \sqrt{\mathrm{N}}\ell_s$ and so this density must be large compared to the string scale, *i.e.,* the density is much larger than $1/\ell_s^2$. With such large densities, one can imagine the discreteness of the fuzzy sphere is essentially lost and so there is good agreement with the continuum sphere of the D2-brane picture. More discussion on the noncommutative geometry appears in Appendix A.

Finally note that the Born-Infeld action contains couplings to the Neveu-Schwarz two-form which are similar to that in eq. (4.4). From the expansion of $\sqrt{\det(Q)}$, one finds a cubic interaction

$$\frac{i}{3}\lambda^2 T_0 \int dt\,\mathrm{Tr}\left(\Phi^i\Phi^j\Phi^k\right) H_{ijk}(t)\ . \tag{4.13}$$

Hence the noncommutative ground state, which has $\mathrm{Tr}\left(\Phi^i\Phi^j\Phi^k\right) \neq 0$, also acts as a source of the B field through the worldvolume coupling

$$-\frac{R_0^3}{3\pi g\ell_s^3}\left(1-\frac{1}{\mathrm{N}^2}\right)\int dt\, H_{123}\ . \tag{4.14}$$

This coupling is perhaps not so surprising given that the noncommutative ground state represents the bound state of a spherical D2-brane and N D0-branes. Explicit supergravity solutions describing D2-D0 bound states with a planar geometry have been found [**17, 110**], and are known to carry a long-range H field with the same profile as the RR field strength $F^{(4)}$. One can also derive this coupling from the dual D2-brane formulation. Furthermore, we observe that the presence of this coupling (4.13) means that we would find an analogous dielectric effect if the N D0-branes were placed in a constant background H field. This mechanism plays a role in describing D-branes in the spacetime background corresponding to a WZW model [**7, 102, 91, 5, 3, 4, 47**]. It seems that quantum group symmetries may be useful in understanding these noncommutative configurations [**103, 104**].

The example considered above must be considered simply a toy calculation demonstrating the essential features of the dielectric effect for D-branes. A more complete calculation would require analyzing the D0-branes in a consistent supergravity background. For example, the present case could be extended to consider the asymptotic supergravity fields of a D2-brane, where the RR four-form would be slowly varying but the metric and dilaton fields would also be nontrivial. Alternatively, one can find solutions with a constant background $F^{(4)}$ in M-theory, namely the $\mathrm{AdS}_4{\times}\mathrm{S}^7$ and $\mathrm{AdS}_7{\times}\mathrm{S}^4$ backgrounds — see, *e.g.,* [**46**]. In lifting the D0-branes to M-theory, they become gravitons carrying momentum in the internal space. Hence the expanded D2-D0 system considered here corresponds to the 'giant gravitons' of ref. [**93**]. The analog of the D2-D0 bound state in a constant background $F^{(4)}$ corresponds to M2-branes with internal momentum expanding into AdS_4 [**58, 63**], while that in a constant H field corresponds to the M2-branes expanding on S^4 [**93**]. The original analysis of these M-theory configurations was made in terms of the Abelian world-volume theory of the M2-brane [**93**] — a Matrix theory description of such states in terms of noncommutative geometry was only developed recently [**77, 6**]. Further, one finds that M5-branes will expand in a similar way for these backgrounds, and that expanded D3-branes arise in the type IIB supergravity background $AdS_5 \times S^5$. Remarkably, a detailed analysis [**58, 93, 63**] shows that these expanded branes are BPS states with the quantum numbers of a graviton. Ref. [**95**] extends this discussion to more general expanded configurations.

Alternatively, the dielectric effect has been found to play a role in other string theory contexts, for example, in the resolution of certain singularities in the AdS/-CFT correspondence [**106**]. Further, one can consider more sophisticated background field configurations, which, through the dielectric effect, generate more complicated noncommutative geometries [**125, 48**]. There is also an interesting generalization to open dielectric branes, in which the extended brane emerging from the dielectric effect ends on another D-brane [**129**]. Other interesting applications of the dielectric effect for D-branes can be found in refs. [**73, 74, 75**].

Appendix A. Noncommutative geometry

The idea that noncommutative geometry should play a role in physical theories is an old one [**119, 118, 26**]. Suggestions have been made that such noncommutative structure may resolve the ultraviolet divergences of quantum field theories, or appear in the description of spacetime geometry at the Planck scale. In the past few years, it has also become a topic of increasing interest to string theorists. From one point of view, the essential step in realizing a noncommutative geometry is replacing the spacetime coordinates by noncommuting operators: $x^\mu \to \hat{x}^\mu$. In this replacement, however, there remains a great deal of freedom in defining the nontrivial commutation relations which the operators $\hat{x}^\mu$ must satisfy. Some explicit choices that have appeared in physical problems are as follows:

(i) Canonical commutation relations:

$$[\hat{x}^\mu, \hat{x}^\nu] = i\theta^{\mu\nu} , \qquad \theta^{\mu\nu} \in \mathbb{C} .$$

Such algebras have appeared in the Matrix theory description of planar D-branes [**11**] — for a review, see [**120**]. This work also stimulated an ongoing investigation by string theorists of noncommutative field theories, which arise in the low energy limit of a planar D-brane with a constant B-field flux — see, *e.g.,* [**114, 27, 43, 44**].

(ii) Quantum space relations:

$$\hat{x}^\mu\, \hat{x}^\nu = q^{-1}\, R^{\mu\nu}{}_{\rho\tau}\, \hat{x}^\rho\, \hat{x}^\tau , \qquad R^{\mu\nu}{}_{\rho\tau} \in \mathbb{C} .$$

These algebras received some attention from physicists in the early 1990's — see, *e.g.,* [**130, 112**] — and have appeared more recently in the geometry of the moduli space of N=4 super-Yang-Mills theory [**13, 14**].

(iii) Lie algebra relations:

$$[\hat{x}^\mu, \hat{x}^\nu] = i f^{\mu\nu}{}_\rho\, \hat{x}^\rho , \qquad f^{\mu\nu}{}_\rho \in \mathbb{C} .$$

Such algebras naturally arise in the description of fuzzy spheres, as was discovered in early attempts to quantize the supermembrane [**68, 37**]. These noncommutative geometries have also been applied in Matrix theory to describe spherical D-branes [**23, 81**]. As discussed in the main text, noncommutative geometries with a Lie-algebra structure and these noncommutative spheres, in particular, arise very naturally in various D-brane systems since the transverse scalars are matrix-valued in the adjoint representation of U(N).

Beginning with a prescribed set of commutation relations for the coordinates on a given manifold, the bulk of the problem in noncommutative geometry is to understand the algebra of functions in this framework. Of course, mathematicians are typically careful in defining the specific class of functions with which they wish to work; however, these details are usually glossed over in physical models. That is, as physicists, we are usually confident that the physics will guide the choice of functions. For a fuzzy sphere, one finds that not only is the product structure modified but that the space of functions is naturally truncated to be finite dimensional. The remainder of the discussion in the appendix will focus on these noncommutative spheres [**68, 37, 90**], in part because they have been found to play an interesting physical role in string theory — see, *e.g.,* [**98, 28, 29, 81, 23**]. We also elaborate on these examples because, in contrast to the above discussion, the natural presentation given below de-emphasizes the role of the commutation relations. In particular, the fuzzy four-sphere will provide an intriguing example below.

To begin the construction of a fuzzy sphere, we begin with the standard definition of a k-sphere using the embedding in $(k+1)$-dimensional Cartesian space

$$\sum_{i=1}^{k+1} \left(x^i\right)^2 = R^2 \ , \qquad\qquad x^i \in \mathbb{R}^{k+1} \ . \tag{A.1}$$

Now, functions on S^k can be expanded in terms of spherical harmonics as

$$f(x^i) = \sum_{\ell=0}^{\infty} f_{i_1\cdots i_\ell} x^{i_1} \cdots x^{i_\ell} \ , \tag{A.2}$$

where $f_{i_1\cdots i_\ell}$ are completely symmetric and traceless tensors. We could be more precise in defining a basis of these tensors, but here we will be satisfied with noting that each term in the sum is a linear combination of spherical harmonics with principal quantum number ℓ. Showing this is straightforward: Denoting the individual terms in the sum as f_ℓ and setting aside the constraint (A.1), it is clear that the Laplacian on the Cartesian space annihilates each of these terms, *i.e.,* $\nabla^2 f_\ell = 0$. Now, in spherical polar coordinates on $\mathbb{R}^{k+1}$, the Laplacian may be written as

$$\nabla^2 = R^{-k}\partial_R \left(R^k \partial_R\right) + R^{-2}\nabla^2_\Omega \ , \tag{A.3}$$

where ∇^2_Ω is the angular Laplacian on the unit k-sphere. Hence it follows that $\nabla^2_\Omega f_\ell = \ell(\ell+k-1) f_\ell$.

In general, to produce a fuzzy sphere, one might proceed by replacing the $k+1$ continuum coordinates above by finite dimensional matrices, $x^i \to \hat{x}^i$, whose commutation relations we leave aside for the moment. The matrices are chosen to satisfy a constraint analogous to eq. (A.1):

$$\sum_{i=1}^{k+1} \left(\hat{x}^i\right)^2 = R^2\, \mathbf{1}_\mathrm{N} \ . \tag{A.4}$$

Similarly the continuum functions are replaced by

$$\hat{f}(\hat{x}^i) = \sum_{\ell=0}^{\ell_{max}} f_{i_1\cdots i_\ell} \hat{x}^{i_1} \cdots \hat{x}^{i_\ell} \ , \tag{A.5}$$

where $f_{i_1\cdots i_\ell}$ are the same symmetric and traceless tensors considered in (A.2). Notice that the 'noncommutative' sum is truncated at some ℓ_{max} because for finite dimensional matrices, such products will only yield a finite number of linearly independent matrices. Thus this matrix construction truncates the full algebra of functions on the sphere to those with $\ell \leq \ell_{max}$, and the star product on the fuzzy sphere differs from that obtained by the deformation quantization of the Poisson structure on the embedding space, *i.e.,* the latter acts on the space of all square integrable functions on the sphere [**49, 66**].

The simplest example of this construction is the fuzzy two-sphere, which was already encountered in section 4. In this case, one chooses $\hat{x}^i = \lambda\, \alpha^i$ with $i = 1, 2, 3$, where the α^i are the generators of the *irreducible* N×N representation of SU(2) satisfying the commutation relations given in eq. (4.8). These generators satisfy the Casimir relation

$$\sum \left(\alpha^i\right)^2 = (\mathrm{N}^2 - 1)\, \mathbf{1}_\mathrm{N} \tag{A.6}$$

and so, in order to satisfy the constraint (A.4), the normalization constant should be chosen as

$$\lambda = \frac{R}{\sqrt{\mathrm{N}^2 - 1}}\ . \tag{A.7}$$

With these N×N matrices, one finds the cutoff in eq. (A.5) is $\ell_{max} = \mathrm{N} - 1$. So heuristically, we might say that with this construction we can only resolve distances on the noncommutative sphere for $\Delta d \gtrsim R/N$. Hence, in the limit $N \to \infty$, one expects to get agreement with the continuum theory, as was illustrated with the physical models in the main text.

We should mention that the entire space of functions (A.5) plays a role in the stringy constructions. To illustrate this point, we consider the example of the fuzzy two-sphere appearing in the example of the dielectric effect in section 4. In the static ground state configuration, three of the transverse scalars have a noncommutative expectation value: $\Phi^i = (f/2)\, \alpha^i$ for $i = 1, 2, 3$. Now one might consider excitations of this system. In particular, it is natural to expand fluctuations of the scalars in terms of the noncommutative spherical harmonics

$$\delta\Phi^m(t) = \psi^m_{i_1\ldots i_\ell}(t)\, \alpha^{i_1} \cdots \alpha^{i_\ell} \tag{A.8}$$

where as above the coefficients $\psi^m_{i_1\cdots i_\ell}$ are completely symmetric and traceless. In the case of overall transverse scalars, *i.e.,* $m \neq 0, 1, 2, 3$, the linearized equations of motion reduce to

$$\begin{aligned} \partial_t^2 \delta\Phi^m(t) &= -[\Phi^i, [\Phi^i, \delta\Phi^m(t)]] \\ &= -\ell(\ell+1) f^2\ \delta\Phi^m(t)\ . \end{aligned} \tag{A.9}$$

Hence inserting the ansatz $\delta\Phi^m(t) \propto e^{-i\omega t}$, we find excitations with frequencies $\omega^2 = \ell(\ell+1)f^2$. Of course, these fluctuations inherit the cutoff $\ell \leq \ell_{max}$ from the noncommutative framework. The analysis of the fluctuations in the $i = 1, 2, 3$ directions is more involved but a nice description is given in [**76**]. In the interesting regime where the dielectric calculations are expected to be valid, *i.e.,* large N and $R \ll \sqrt{\mathrm{N}}\ell_s$, we have $f \ll 1/\sqrt{\mathrm{N}}\ell_s$. Hence the low energy excitations are well below the string scale, giving further corroboration that the low energy effective action provides an adequate description of the physics [**98**]. We might add that these frequencies match the results found from calculations in the dual continuum framework of the expanded D2-brane, again up to $1/\mathrm{N}^2$ corrections. Of course, the latter description gives no upper cutoff on the angular momentum. A similar discussion [**28**] applies for the excitations of the D3⊥D1-system.

We now turn to a discussion of the fuzzy four-sphere, which is relevant for the construction of the D-string description of the D5⊥D1 system [**29**]. This construction also played a role in ref. [**23**] for the Matrix theory description of spherical D4-branes (longitudinal M5-branes). At first sight, the construction of the fuzzy four-sphere appears very similar to the fuzzy two-sphere above, but in fact it yields a very different object. One begins by choosing $\hat{x}^i = \lambda\, G^i$ where the G^i are an appropriate set of N×N matrices. These matrices were first constructed in ref. [**61**] — see also ref. [**23**]. The G^i are given by the totally symmetric n-fold tensor product of 4×4 gamma matrices:

$$G^i = \big(\Gamma^i\otimes\mathbf{1}\otimes\cdots\otimes\mathbf{1}+\mathbf{1}\otimes\Gamma^i\otimes\mathbf{1}\otimes\cdots\otimes\mathbf{1} + \cdots + \mathbf{1}\otimes\cdots\otimes\mathbf{1}\otimes\Gamma^i\big)_{\mathrm{Sym}}\ , \tag{A.10}$$

where Γ_i, $i=1,\ldots,5$, are 4×4 Euclidean gamma matrices, and $\mathbf{1}$ is the 4×4 identity matrix. The subscript 'Sym' means the matrices are restricted to the completely symmetric tensor product space. With the latter restriction, the dimension of the matrices becomes

$$N = \frac{(n+1)(n+2)(n+3)}{6}\ , \tag{A.11}$$

where n is the integer denoting the size of the tensor product in eq. (A.10). The 'Casimir' associated with the G^i matrices, *i.e.,* $G^iG^i = c\,\mathbf{1}_{\mathrm{N}}$, is given by

$$c = n(n+4)\ . \tag{A.12}$$

Hence, to satisfy eq. (A.4), we choose the normalization constant in $\hat{x}^i = \lambda G^i$ as $\lambda = R/\sqrt{n(n+4)}$.

These matrices were presented as a representation of the fuzzy four-sphere on the basis of a discussion of representations of SO(5) in ref. [**61**]. Working within Matrix theory, ref. [**23**] provided a series of physical arguments towards the same end. That is, the G^i produce a spherical locus, are rotationally invariant under the action of SO(5) and give an appropriate spectrum of eigenvalues.

With these G^i, one can construct matrix harmonics (A.5) with $\ell \le \ell_{max} = n$, as before [**61**]. However, a key difference between the fuzzy two-sphere and the fuzzy four-sphere is that the G^i do *not* form a Lie algebra (in contrast to the α^i used to construct the fuzzy two-sphere). As a result the algebra of these matrix harmonics does not close! [**23, 61**] In particular, one finds that the commutators $G^{ij}\equiv[G^i,G^j]/2$ define linearly independent matrices. The commutators of the G^i and G^{jk} are easily obtained:[1]

$$\begin{aligned}
[G^{ij},G^k] &= 2(\delta^{jk}G^i-\delta^{ik}G^j),\\
[G^{ij},G^{kl}] &= 2(\delta^{jk}G^{il}+\delta^{il}G^{jk}-\delta^{ik}G^{jl}-\delta^{jl}G^{ik}).
\end{aligned}$$

Note then that the G^{ik} are the generators of $SO(5)$ rotations. Hence, combined, the G^i and G^{ij} give a representation of the algebra $SO(1,5)$, as can be seen from the definition of the G^{ij} and the commutators in eq. (A.13). Hence a closed algebra of matrix functions would given by

$$\tilde{a}_{a_1a_2\cdots a_\ell}\,\widetilde{G}^{a_1}\widetilde{G}^{a_2}\cdots\widetilde{G}^{a_\ell}\ , \tag{A.13}$$

[1]We refer the interested reader to refs. [**29, 23**] for more details.

where the $\widetilde{G}^a$ are generators of $SO(1,5)$ with $a = 1, \ldots, 15$, and the $\tilde{a}$ are naturally symmetric in the $SO(1,5)$ indices. Identifying $\widetilde{G}^a = G^a$ for $a = 1 \ldots 5$, the desired matrix harmonics would correspond to the subset of $\tilde{a}$ with nonvanishing entries only for indices $a_i \leq 5$. Thus, while the fuzzy four-sphere construction introduces an algebra that contains a truncated set of the spherical harmonics on S^4, the algebra also contains a large number of elements transforming under other representations of the $SO(5)$ symmetry group that acts on the four-sphere. The reader may find a precise description of the complete algebra in ref. [**61**], in terms of representations of $SO(5)$ (or rather $Spin(5) = Sp(4)$).

Given this extended algebra, or alternatively the appearance of 'spurious' modes, one might question whether the above construction provides a suitable noncommutative description of the four-sphere. Nonetheless, in the context of non-Abelian D-branes (or Matrix theory), the G^i certainly form a basis for the description of physically interesting systems such as a spherical D4-D0 bound state [**23**] or a D5⊥D1 system [**29**]. In the latter case, N D-strings open up into a collection of n perpendicular D5-branes and the fuzzy four-sphere forms the cross-section of the funnel describing this geometry. From the point of view of the D5-brane theory, the four-sphere is endowed with a nontrivial SU(n) bundle. In this context, one can show that the relation between N and n in eq. (A.11) essentially arises from demanding that the gauge field configuration is homogeneous on the four-sphere [**29**]. Given that there are extra instantonic or gauge field degrees of freedom in the full physical system, it is natural that the 'spurious' modes above should be related to non-Abelian excitations of the D5-brane theory. Evidence for this identification can be found by studying the linearized excitations of the fuzzy funnel describing the D5⊥D1 system and their couplings to the bulk RR fields [**29**]. A thorough analysis of the noncommutative geometry [**108, 67, 82**] also indicates that this identification is correct. Hence this fuzzy four-sphere construction has a natural physical interpretation within string theory.

References

1. A. Abouelsaood, C. G. Callan, C. R. Nappi, and S. A. Yost, *Open strings in background gauge fields*, Nuclear Phys. B **280** (1987), no. 4, 599–624.
2. Ofer Aharony, Steven S. Gubser, Juan Maldacena, Hirosi Ooguri, and Yaron Oz, *Large N field theories, string theory and gravity*, Phys. Rep. **323** (2000), no. 3-4, 183–386.
3. Anton Yu. Alekseev, Andreas Recknagel, and Volker Schomerus, *Non-commutative world-volume geometries: branes on* SU(2) *and fuzzy spheres*, J. High Energy Phys. (1999), no. 9, Paper 23, 20 pp. (electronic).
4. ——, *Brane dynamics in background fluxes and non-commutative geometry*, J. High Energy Phys. (2000), no. 5, Paper 10, 25.
5. Anton Yu. Alekseev and Volker Schomerus, *D-branes in the WZW model*, Phys. Rev. D (3) **60** (1999), no. 6, 061901, 2.
6. B. Janssen, Y. Lozano, and D. Rodriguez-Gomez, *A microscopical description of giant gravitons. II: The $AdS_5 \times S^5$ background*, Nucl. Phys. B **669** (2003) 363-378.
7. Constantin Bachas, Michael R. Douglas, and Christoph Schweigert, *Flux stabilization of D-branes*, J. High Energy Phys. (2000), no. 5, Paper 48, 18.
8. Constantin P. Bachas, Pascal Bain, and Michael B. Green, *Curvature terms in D-brane actions and their M-theory origin*, J. High Energy Phys. (1999), no. 5, Paper 11, 30 pp. (electronic).
9. Pascal Bain, *On the non-Abelian Born-Infeld action*, Progress in string theory and M-theory (Dordrecht, The Netherlands) (Marco Picco, Laurent Baulieu, Michael Green and Paul Windey, eds.), Kluwer, 2001.

10. Vijay Balasubramanian and Asad Naqvi, *Giant gravitons and a correspondence principle*, Phys. Lett. B **528** (2002), no. 1-2, 111–120.
11. T. Banks, W. Fischler, S. H. Shenker, and L. Susskind, *M theory as a matrix model: a conjecture*, Phys. Rev. D (3) **55** (1997), no. 8, 5112–5128.
12. Tom Banks and Nathan Seiberg, *Strings from matrices*, Nuclear Phys. B **497** (1997), no. 1-2, 41–55.
13. David Berenstein, Vishnu Jejjala, and Robert G. Leigh, *Marginal and relevant deformations of $N = 4$ field theories and non-commutative moduli spaces of vacua*, Nuclear Phys. B **589** (2000), no. 1-2, 196–248.
14. ———, *Non-commutative moduli spaces, dielectric tori and T-duality*, Phys. Lett. B **493** (2000), no. 1-2, 162–168.
15. E. A. Bergshoeff, M. de Roo, and A. Sevrin, *Non-Abelian Born- Infeld and kappa-symmetry*, J. Math. Phys. **42** (2001), no. 7, 2872–2888.
16. Eric Bergshoeff, Chris Hull, and Tomás Ortín, *Duality in the type-II superstring effective action*, Nuclear Phys. B **451** (1995), no. 3, 547–575.
17. J. C. Breckenridge, G. Michaud, and R. C. Myers, *More D-brane bound states*, Phys. Rev. D (3) **55** (1997), no. 10, 6438–6446.
18. T. H. Buscher, *Quantum Corrections and Extended Supersymmetry in New Sigma Models*, Phys. Lett. B **159** (1985), 127.
19. ———, *A symmetry of the string background field equations*, Phys. Lett. B **194** (1987), no. 1, 59–62.
20. ———, *Path-integral derivation of quantum duality in nonlinear sigma-models*, Phys. Lett. B **201** (1988), no. 4, 466–472.
21. C. G. Callan, C. Lovelace, C. R. Nappi, and S. A. Yost, *Adding holes and crosscaps to the superstring*, Nuclear Phys. B **293** (1987), no. 1, 83–113.
22. ———, *Loop corrections to superstring equations of motion*, Nuclear Phys. B **308** (1988), no. 2-3, 221–284.
23. Judith Castelino, Sangmin Lee, and Washington Taylor, *Longitudinal* 5*-branes as* 4*-spheres in Matrix theory*, Nuclear Phys. B **526** (1998), no. 1-3, 334–350.
24. Yeuk-Kwan E. Cheung and Zheng Yin, *Anomalies, branes, and currents*, Nuclear Phys. B **517** (1998), no. 1-3, 69–91.
25. Andres Collinucci, Mees de Roo, and Martijn G. C. Eenink, *Supersymmetric Yang-Mills theory at order α'^3*, J. High Energy Phys. (2002), no. 6, no. 24, 16.
26. Alain Connes, *Noncommutative geometry*, Academic Press Inc., San Diego, CA, 1994.
27. Alain Connes, Michael R. Douglas, and Albert Schwarz, *Noncommutative geometry and matrix theory: compactification on tori*, J. High Energy Phys. (1998), no. 2, Paper 3, 35 pp. (electronic).
28. Neil R. Constable, Robert C. Myers, and Øyvind Tafjord, *Noncommutative bion core*, Phys. Rev. D (3) **61** (2000), no. 10, 106009, 14.
29. ———, *Non-Abelian brane intersections*, J. High Energy Phys. (2001), no. 6, Paper 23, 37.
30. L. Cornalba, *On the general structure of the non-Abelian Born- Infeld action*, Adv. Theor. Math. Phys. **4** (2000), no. 6, 1259–1318.
31. Ben Craps and Frederik Roose, *Anomalous D-brane and orientifold couplings from the boundary state*, Phys. Lett. B **445** (1998), no. 1-2, 150–159.
32. ———, *(Non-)anomalous D-brane and O-plane couplings: the normal bundle*, Phys. Lett. B **450** (1999), no. 4, 358–362.
33. Jin Dai, R. G. Leigh, and Joseph Polchinski, *New connections between string theories*, Modern Phys. Lett. A **4** (1989), no. 21, 2073–2083.
34. S.R. Das and S.D. Mathur, *The quantum physics of black holes: Results from string theory*, Ann. Rev. Nucl. Part. Sci. **50** (2000), no. 6, 153–206.
35. Justin R. David, Gautam Mandal, and Spenta R. Wadia, *Microscopic formulation of black holes in string theory*, Phys. Rep. **369** (2002), no. 6, 549–686.
36. Mees de Roo, Martijn G. C. Eenink, Paul Koerber, and Alexander Sevrin, *Testing the fermionic terms in the non-Abelian D-brane effective action through order α'^3*, J. High Energy Phys. (2002), no. 8, no. 11, 9.
37. B. de Wit, J. Hoppe, and H. Nicolai, *On the quantum mechanics of supermembranes*, Nuclear Phys. B **305** (1988), no. 4, FS23, 545–581.

38. Frederik Denef, Alexander Sevrin, and Jan Troost, *Non-Abelian Born-Infeld versus string theory*, Nuclear Phys. B **581** (2000), no. 1-2, 135–155.
39. H. Dorn, *Non-Abelian gauge field dynamics on matrix D-branes*, Nuclear Phys. B **494** (1997), no. 1-2, 105–118.
40. Michael R. Douglas, *D-branes in curved space*, Adv. Theor. Math. Phys. **1** (1997), no. 2, 198–209.
41. ______, *D-branes and Matrix theory in curved space*, Nuclear Phys. B Proc. Suppl. **68** (1998), 381–393.
42. ______, *D-branes, categories and $N = 1$ supersymmetry*, J. Math. Phys. **42** (2001), no. 7, 2818–2843.
43. Michael R. Douglas and Chris Hull, *D-branes and the noncommutative torus*, J. High Energy Phys. (1998), no. 2, Paper 8, 5 pp. (electronic).
44. Michael R. Douglas and Nikita A. Nekrasov, *Noncommutative field theory*, Rev. Modern Phys. **73** (2001), no. 4, 977–1029.
45. M.R. Douglas, *Branes within Branes*, hep-th/9512077, in Strings, Branes and Dualities (Cargèse 1997) 267-275.
46. M. J. Duff, *TASI lectures on branes, black holes and anti-de Sitter space*, Strings, branes and gravity, TASI 99 (Boulder, CO), World Sci. Publishing, River Edge, NJ, 2001, pp. 3–125.
47. Stefan Fredenhagen and Volker Schomerus, *Branes on group manifolds, gluon condensates, and twisted K-theory*, J. High Energy Phys. (2001), no. 4, Paper 7, 30.
48. ______, *D-branes in coset models*, J. High Energy Phys. (2002), no. 2, No. 5, 27.
49. Laurent Freidel and Kirill Krasnov, *The fuzzy sphere $\star$-product and spin networks*, J. Math. Phys. **43** (2002), no. 4, 1737–1754.
50. Jürgen Fuchs and Christoph Schweigert, *Branes: from free fields to general backgrounds*, Nuclear Phys. B **530** (1998), no. 1-2, 99–136.
51. Mohammad R. Garousi, *Superstring scattering from D-branes bound states*, J. High Energy Phys. (1998), no. 12, Paper 8, 28 pp. (electronic).
52. Mohammad R. Garousi and Robert C. Myers, *Superstring scattering from* D*-branes*, Nuclear Phys. B **475** (1996), no. 1-2, 193–224.
53. ______, *World-volume interactions on D-branes*, Nuclear Phys. B **542** (1999), no. 1-2, 73–88.
54. ______, *World-volume potentials on D-branes*, J. High Energy Phys. (2000), no. 11, Paper 32, 28.
55. A. Giveon, M. Porrati, and E. Rabinovici, *Target space duality in string theory*, Phys. Rep. **244** (1994), no. 2-3, 77–202.
56. Michael B. Green, *Space-time duality and Dirichlet string theory*, Phys. Lett. B **266** (1991), no. 3-4, 325–336.
57. Michael B. Green, Jeffrey A. Harvey, and Gregory Moore, *I-brane inflow and anomalous couplings on D-branes*, Classical Quantum Gravity **14** (1997), no. 1, 47–52.
58. Marcus T. Grisaru, Robert C. Myers, and Øyvind Tafjord, *SUSY and Goliath*, J. High Energy Phys. (2000), no. 8, Paper 40, 28.
59. David J. Gross and John H. Sloan, *The quartic effective action for the heterotic string*, Nuclear Phys. B **291** (1987), no. 1, 41–89.
60. David J. Gross and Edward Witten, *Superstring modifications of Einstein's equations*, Nuclear Phys. B **277** (1986), no. 1, 1–10.
61. H. Grosse, C. Klimčík, and P. Prešnajder, *On finite 4D quantum field theory in non-commutative geometry*, Comm. Math. Phys. **180** (1996), no. 2, 429–438.
62. Jeffrey A. Harvey and Per Kraus, *D-branes as unstable lumps in bosonic open string field theory*, J. High Energy Phys. (2000), no. 4, Paper 12, 14.
63. Akikazu Hashimoto, Shinji Hirano, and N. Itzhaki, *Large branes in AdS and their field theory dual*, J. High Energy Phys. (2000), no. 8, Paper 51, 20.
64. Akikazu Hashimoto and Igor R. Klebanov, *Scattering of strings from D-branes*, Nuclear Phys. B Proc. Suppl. **55B** (1997), 118–133.
65. Akikazu Hashimoto and Washington Taylor, *Fluctuation spectra of tilted and intersecting D-branes from the Born-Infeld action*, Nuclear Phys. B **503** (1997), no. 1-2, 193–219.
66. E. Hawkins, *Noncommutative regularization for the practical man*, Tech. Report hep-th/9908052, August 1999.
67. Pei-Ming Ho and Sanjaye Ramgoolam, *Higher-dimensional geometries from matrix brane constructions*, Nuclear Phys. B **627** (2002), no. 1-2, 266–288.

68. Jens Hoppe, *Quantum theory of a massless relativistic surface and a two-dimensional bound state problem*, Ph.D. thesis, MIT, July 1982 (available at http://www.aei-potsdam.mpg.de/ hoppe).
69. Petr Hořava, *Background duality of open-string models*, Phys. Lett. B **231** (1989), no. 3, 251–257.
70. ———, *Strings on world-sheet orbifolds*, Nuclear Phys. B **327** (1989), no. 2, 461–484.
71. Gary T. Horowitz and Andrew Strominger, *Black strings and p-branes*, Nuclear Phys. B **360** (1991), no. 1, 197–209.
72. Chris M. Hull, *Matrix theory, U-duality and toroidal compactifications of M-theory*, J. High Energy Phys. (1998), no. 10, Paper 11, 22 pp. (electronic).
73. Yoshifumi Hyakutake, *Torus-like dielectric D2-brane*, J. High Energy Phys. (2001), no. 5, Paper 13, 22.
74. ———, *Expanded strings in the background of NS5-branes via a M2-brane, a D2-brane and D0-branes*, J. High Energy Phys. (2002), no. 1, No. 21, 31.
75. ———, *Notes on the construction of the D2-brane from multiple D0-branes*, Nucl. Phys. B **675** (2003) 241-269.
76. Satoshi Iso, Yusuke Kimura, Kanji Tanaka, and Kazunori Wakatsuki, *Noncommutative gauge theory on fuzzy sphere from matrix model*, Nuclear Phys. B **604** (2001), no. 1-2, 121–147.
77. B. Janssen and Y. Lozano, *A microscopical description of giant gravitons*, Nuclear Phys. B **658** (2003), 281.
78. Clifford Johnson, *D-branes*, Cambridge University Press, Cambridge, 2002.
79. Clifford V. Johnson, Nemanja Kaloper, Ramzi R. Khuri, and Robert C. Myers, *Is string theory a theory of strings?*, Phys. Lett. B **368** (1996), no. 1-2, 71–77.
80. Daniel Kabat and Washington Taylor, *Linearized supergravity from Matrix theory*, Phys. Lett. B **426** (1998), no. 3-4, 297–305.
81. ———, *Spherical membranes in M(atrix) theory*, Adv. Theor. Math. Phys. **2** (1998), no. 1, 181–206.
82. Yusuke Kimura, *Noncommutative gauge theory on fuzzy four-sphere and matrix model*, Nuclear Phys. B **637** (2002), no. 1-3, 177–198.
83. Paul Koerber and Alexander Sevrin, *The non-Abelian open superstring effective action through order* α'^3, J. High Energy Phys. (2001), no. 10, Paper 3, 30.
84. ———, *Testing the* α'^3 *term in the non-Abelian open superstring effective action*, J. High Energy Phys. (2001), no. 9, Paper 9, 13.
85. ———, *The non-Abelian D-brane effective action through order* α'^4, J. High Energy Phys. (2002), no. 10, 46.
86. Maxim Kontsevich, *Homological algebra of mirror symmetry*, Proceedings of the International Congress of Mathematicians, Vol. 1, 2 (Zürich, 1994) (Basel), Birkhäuser, 1995, pp. 120–139.
87. Frédéric Leblond, Robert C. Myers, and David C. Page, *Superstars and giant gravitons in M-theory*, J. High Energy Phys. (2002), no. 1, No. 26, 16.
88. R. G. Leigh, *Dirac-Born-Infeld Action from Dirichlet Sigma Model*, Modern Phys. Lett. A **4** (1989), no. 21, 2767.
89. Miao Li, *Boundary states of D-branes and dy-strings*, Nuclear Phys. B **460** (1996), no. 2, 351–361.
90. J. Madore, *An introduction to noncommutative differential geometry and its physical applications*, London Mathematical Society Lecture Note Series, vol. 206, Cambridge University Press, Cambridge, 1995.
91. Juan Maldacena, Nathan Seiberg, and Gregory Moore, *Geometrical interpretation of D-branes in gauged WZW models*, J. High Energy Phys. (2001), no. 7, Paper 46, 63.
92. Juan M. Maldacena, *Black holes in string theory*, Ph.D. thesis, Princeton University, July 1996.
93. John McGreevy, Leonard Susskind, and Nicolaos Toumbas, *Invasion of the giant gravitons from anti-de Sitter space*, J. High Energy Phys. (2000), no. 6, Paper 8, 14.
94. Patrick Meessen and Tomás Ortín, *An* $\mathrm{Sl}(2,\mathbf{Z})$ *multiplet of nine-dimensional type II supergravity theories*, Nuclear Phys. B **541** (1999), no. 1-2, 195–245.
95. Andrei Mikhailov, *Giant gravitons from holomorphic surfaces*, J. High Energy Phys. (2000), no. 11, Paper 27, 23.

96. Ruben Minasian and Gregory Moore, *K-theory and Ramond- Ramond charge*, J. High Energy Phys. (1997), no. 11, Paper 2, 7 pp. (electronic).
97. Nicolas Moeller, Barton Zwiebach, and Ashoke Sen, *D-branes as tachyon lumps in string field theory*, J. High Energy Phys. (2000), no. 8, Paper 39, 31.
98. Robert C. Myers, *Dielectric-branes*, J. High Energy Phys. (1999), no. 12, Paper 22, 41.
99. Robert C. Myers and Øyvind Tafjord, *Superstars and giant gravitons*, J. High Energy Phys. (2001), no. 11, Paper 9, 19.
100. Yuji Okawa and Hirosi Ooguri, *How noncommutative gauge theories couple to gravity*, Nuclear Phys. B **599** (2001), no. 1-2, 55–82.
101. ———, *Energy-momentum tensors in matrix theory and in noncommutative gauge theories*, Strings, 2001 (Mumbai), Clay Math. Proc., vol. 1, Amer. Math. Soc., Providence, RI, 2002, pp. 263–277.
102. Jacek Pawelczyk, SU(2) *WZW D-branes and their noncommutative geometry from DBI action*, J. High Energy Phys. (2000), no. 8, Paper 6, 7.
103. Jacek Pawelczyk and Harold Steinacker, *Matrix description of D-branes on 3-spheres*, J. High Energy Phys. (2001), no. 12, Paper 18, 14.
104. ———, *A quantum algebraic description of D-branes on group manifolds*, Nuclear Phys. B **638** (2002), no. 3, 433–458.
105. Amanda W. Peet, *Black holes in string theory*, TASI 99: Strings, Branes, And Gravity (S. Kachru J. Harvey and E. Silverstein, eds.), World Scientific, 2001.
106. J. Polchinski and M.J. Strassler, *The string dual of a confining four-dimensional gauge theory*, Tech. Report hep-th/0003136, March 2000.
107. Joseph Polchinski, *Dirichlet branes and Ramond-Ramond charges*, Phys. Rev. Lett. **75** (1995), no. 26, 4724–4727.
108. Sanjaye Ramgoolam, *On spherical harmonics for fuzzy spheres in diverse dimensions*, Nuclear Phys. B **610** (2001), no. 3, 461–488.
109. A. Recknagel and V. Schomerus, *D-branes in Gepner models*, Nuclear Phys. B **531** (1998), no. 1-3, 185–225.
110. J. G. Russo and A. A. Tseytlin, *Waves, boosted branes and BPS states in M-theory*, Nuclear Phys. B **490** (1997), no. 1-2, 121–144.
111. Volker Schomerus, *Lectures on branes in curved backgrounds*, Classical Quantum Gravity **19** (2002), 5781–5847.
112. Peter Schupp, Paul Watts, and Bruno Zumino, *Differential geometry on linear quantum groups*, Lett. Math. Phys. **25** (1992), no. 2, 139–147.
113. Claudio A. Scrucca and Marco Serone, *A note on the torsion dependence of D-brane RR couplings*, Phys. Lett. B **504** (2001), no. 1-2, 47–54.
114. Nathan Seiberg and Edward Witten, *String theory and noncommutative geometry*, J. High Energy Phys. (1999), no. 9, Paper 32, 93 pp. (electronic).
115. Ashoke Sen, *Non-BPS states and branes in string theory*, Progress in string theory and M-theory (Dordrecht, The Netherlands) (Marco Picco, Laurent Baulieu, Michael Green and Paul Windey, eds.), Kluwer, 2001 187-234.
116. Alexander Sevrin, Jan Troost, and Walter Troost, *The non-Abelian Born-Infeld action at order* F^6, Nuclear Phys. B **603** (2001), no. 1-2, 389–412.
117. Eric Sharpe, *D-branes, derived categories, and Grothendieck groups*, Nuclear Phys. B **561** (1999), no. 3, 433–450.
118. H.S. Snyder, *The Electromagnetic Field In Quantized Space-Time*, Phys. Rev. **72** (1947), 68.
119. ———, *Quantized Space-Time*, Phys. Rev. **71** (1947), 38.
120. Washington Taylor, *M(atrix) theory: matrix quantum mechanics as a fundamental theory*, Rev. Modern Phys. **73** (2001), no. 2, 419–461.
121. Washington Taylor and Mark Van Raamsdonk, *Multiple* D0*-branes in weakly curved backgrounds*, Nuclear Phys. B **558** (1999), no. 1-2, 63–95.
122. ———, *Supergravity currents and linearized interactions for Matrix theory configurations with fermionic backgrounds*, J. High Energy Phys. (1999), no. 4, Paper 13, 76 pp. (electronic).
123. ———, *Multiple* Dp*-branes in weak background fields*, Nuclear Phys. B **573** (2000), no. 3, 703–734.
124. Paul Townsend, *P-brane democracy*, Particles, strings and cosmology (J. Bagger, ed.), World Scientific, 1996.

125. Sandip P. Trivedi and Sachindeo Vaidya, *Fuzzy cosets and their gravity duals*, J. High Energy Phys. (2000), no. 9, Paper 41, 33.
126. A. A. Tseytlin, *Vector field effective action in the open superstring theory*, Nuclear Phys. B **276** (1986), no. 2, 391–428.
127. ———, *On non-Abelian generalisation of the Born-Infeld action in string theory*, Nuclear Phys. B **501** (1997), no. 1, 41–52.
128. ———, *Born-Infeld action, supersymmetry and string theory*, The many faces of the superworld, World Sci. Publishing, River Edge, NJ, 2000, pp. 417–452.
129. Mark Van Raamsdonk, *Open dielectric branes*, J. High Energy Phys. (2002), no. 2, No. 1, 55.
130. Julius Wess and Bruno Zumino, *Covariant differential calculus on the quantum hyperplane*, Nuclear Phys. B Proc. Suppl. **18B** (1990), 302–312 (1991).
131. Edward Witten, *Bound states of strings and p-branes*, Nuclear Phys. B **460** (1996), no. 2, 335–350.
132. ———, *D-branes and K-theory*, J. High Energy Phys. (1998), no. 12, Paper 19, 41 pp. (electronic).
133. ———, *Overview of K-theory applied to strings*, Internat. J. Modern Phys. A **16** (2001), no. 5, 693–706.
134. Niclas Wyllard, *Derivative corrections to D-brane actions with constant background fields*, Nuclear Phys. B **598** (2001), no. 1-2, 247–275.

Perimeter Institute for Theoretical Physics, Waterloo, Ontario N2J 2W9 Canada

Department of Physics, McGill University, Montréal, Québec H3A 2T8 Canada
E-mail address: rmyers@perimeterinstitute.ca

List of Participants

Bobby Acharya
Rutgers University
Piscataway, NJ, USA
bacharya@physics.rutgers.edu

Cecilia Albertsson
Stockholm Centre for Physics
Stockholm University, Sweden
cecilia@physto.se

Luis Alvarez-Consul
University of Bath
Bath, U.K.
lalvarez@math.uiuc.edu

Costas Bachas
Ecole Normale Supérieure
Paris, France
bachas@physique.ens.fr

Serguei Barannikov
DMA Ecole Normale Superieure
Paris, France
serguei.barannikov@ens.fr

Victor Batyrev
Universitaet Tuebingen
Tuebingen, Germany
victor.batyrev@uni-tuebingen.de

Ingrid Bauer
Universitaet Bayreuth
Bayreuth, Germany
ingrid.bauer@uni-bayreuth.de

Marcin Bobienski
Warsaw University
Warsaw, Poland
bobi@fuw.edu.pl

Volker Braun
Humboldt University
Berlin, Germany
volker.braun@physik.hu-berlin.de

Gavin Brown
University of Sydney
Sydney, Australia
gavin@maths.usyd.edu.au

Nero Budur
University of Illinois
Chicago, USA
nero@math.uic.edu

Igor Burban
Universitat Kaiserlautern
Kaiserlautern, Germany
burban@mathematik.uni-kl.de

Fabrizio Catanese
Universtaet Bayreuth
Bayreuth, Germany
fabrizio.catanese@uni-bayreuth.de

Richard Cleyton
University of Southern Denmark
Odense, Denmark
cleyton@imada.sdu.dk

Alessio Corti
University of Cambridge
Cambridge, UK
a.corti@dpmms.cam.ac.uk

Alastair Craw
University of Utah
Salt Lake City, USA
craw@math.utah.edu

Andrew Dancer
University of Oxford
Oxford, UK
dancer@maths.ox.ac.uk

Sarah Dean
Harvard University
Cambridge, USA
sarah@math.harvard.edu

Nihat Sadik Deger
Feza Gursey Institute Rasathane Yolu
Istanbul, Turkey
deger@gursey.gov.tr

Robert Henricus Dijkgraaf
University of Amsterdam
Amsterdam, Netherlands
rhd@wins.uva.nl

Igor Dolgachev
University of Michigan
Ann Arbor, USA
idolga@math.lsa.umich.edu

Ron Donagi
University of Pennsylvania
Philadelphia, USA
donagi@math.upenn.edu

Simon Donaldson
Imperial College
London, UK
s.donaldson@ic.ac.uk

Nick Dorey
University of Wales Swansea
Swansea, UK
n.dorey@swansea.ac.uk

Michael Douglas
Rutgers University
Piscataway, USA
mrd@physics.rutgers.edu

Christian Van Enckevort
Johannes Gutenberg-Universitaet
Mainz, Germany
enckevor@mathematik.uni-mainz.de

Has-Georg Freiermuth
Columbia University
New York, USA
freiermuth@math.columbia.edu

Jerome Gauntlett
Queen Mary, University of London
London, UK
j.p.gauntlett@qmw.ac.uk

Jos Gheerardyn
University of Leuven
Leuven, Belgium
jos.gheerardyn@fys.kuleuven.ac.be

Gary Gibbons
University of Cambridge
Cambridge, UK
g.w.gibbons@damtp.cam.ac.uk

Mark Gross
University of Warwick
Coventry, UK
mgross@maths.warwick.ac.uk

Daniel Grunberg
University of Amsterdam
Amsterdam, Netherlands
grunberg@science.uva.nl

Marco Gualtieri
University of Oxford
Oxford, UK
marco.gualtieri@sjc.ox.ac.uk

Stefano Guerra
University of Pennsylvania
Philadelphia, USA
guerras@math.upenn.edu

Sergei Gukov
Harvard University
Cambridge, USA
gukov@gauss.harvard.edu

Sean Hartnoll
University of Cambridge
Cambridge, UK
s.a.hartnoll@damtp.cam.ac.uk

Yang-Hui He
MIT
Cambridge, USA
yhe@ctp.mit.edu

Nigel Hitchin
University of Oxford
Oxford, UK
hitchin@maths.ox.ac.uk

Petr Horava
University of California
Berkeley, USA
horava@socrates.berkeley.edu

Chris Hull
Queen Mary, University of London
London, UK
c.m.hull@qmw.ac.uk

Tasneem Zehra Husain
University of Texas
Austin, USA
tasneem@zippy.ph.utexas.edu

Akira Ishii
Kyoto University
Kyoto, Japan
akira@kusm.kyoto-u.ac.jp

Yukari Ito
Tokyo Metropolitan University
Tokyo, Japan
yukari@comp.metro-u.ac.jp

Arthur Jaffe
Clay Mathematics Institute
Cambridge, USA
USA

Clifford Johnson
University of Durham
Durham, UK
c.v.johnson@durham.ac.uk

Dominic Joyce
Oxford University
Oxford, UK
dominic.joyce@lincoln.ox.ac.uk

Bernard Julia
Ecole Normale Superieure
Paris, France
bernard.julia@ens.fr

Dmitry Kaledin
Independent University of Moscow
Moscow, Russia
kaledin@mccme.ru

Anton Kapustin
California Institute of Technology
Pasadena, USA
kapustin@theory.caltech.edu

Yujiro Kawamata
University of Tokyo
Tokyo, Japan
kawamata@ms.u-tokyo.ac.jp

Masayuki Kawakita
University of Cambridge
Cambridge, UK
kawakita@ms.u-tokyo.ac.jp

Oskar Kedzierski
UJ Krakow
Krakow, Poland
kedziers@in.uj.edu.pl

Alastair King
University of Bath
Bath, UK
a.d.king@maths.bath.ac.uk

Frances Kirwan
University of Oxford
Oxford, UK
kirwan@maths.ox.ac.uk

Masanori Kobayashi
Tokyo Metropolitan University
Tokyo, Japan
masanori@math.metro-u.ac.jp

Alexey Kokotov
197101 Kronverkskaya 8, apt. 26
St Petersburgh, Russia
alexey_kokotov@hotmail.com

Boris Kors
Utrecht University
Utrecht, Netherlands
kors@phys.uu.nl

Alexei Kovalev
University of Cambridge
Cambridge, UK
a.g.kovalev@dpmms.cam.ac.uk

S. Prem Kumar
University of Wales
Swansea, UK
s.p.kumar@swansea.ac.uk

Adrian Langer
University of Warwick
Coventry, UK
langer@maths.warwick.ac.uk

Frederic Leblond
Perimeter Institute
Waterloo, Canada
fleblond@hep.physics.mcgill.ca

Felipe Leitner
Humboldt University
Berlin, Germany
leitner@mathematik.hu-berlin.de

Wolfgang Lerche
CERN
Geneva, Switzerland
lerche@nxth04.cern.ch

Tim Logvinenko
University of Bath
Bath, UK
maptl@maths.bath.ac.uk

Juha Petteri Loikkanen
Stockholm University
Stockholm, Sweden
juha@theophys.kth.se

Jan Louis
Martin-Luther-Universitat
Halle, Germany
j.louis@physik.uni-halle.de

Fernando Marchesano
Universidad Autonoma de Madrid
Madrid, Spain
marchesa@delta.ft.uam.es

Antonio Mateos
University of Barcelona
Barcelona, Spain
tonim@ecm.ub.es

Tomoo Matsumura
Boston University
Boston, USA
mushmt@math.bu.edu

James McKernan
University of California
Santa Barbara, USA
mckernan@math.ucsb.edu

Ilarion Victor Melnikov
Duke University
Durham, USA
lmel@cgtp.duke.edu

Ruben Minasian
Ecole Polytechnique
Palaiseau, France
ruben@cpht.polytechnique.fr

Gabriele Mondello
Scuola Normale Superiore
Pisa, Italy
mondello@cibs.sns.it

Gregory Moore
Rutgers University
Piscataway, USA
gmoore@physics.rutgers.edu

Mircea Mustata
University of California
Berkeley, USA
mustata@math.berkeley.edu

Rob Myers
Perimeter Institute
Waterloo, Canada
rcm@hep.physics.mcgill.ca

Yasunari Nagai
University of Tokyo
Tokyo, Japan
nagai@ms.u-tokyo.ac.jp

Viacheslav Nikulin
University of Liverpool
Liverpool, UK
vnikulin@liv.ac.uk

Carlos Nunez
University of Cambridge
Cambridge, UK
cn239@damtp.cam.ac.uk

Aybike Ozer
Queen Mary, University of London
London, UK
a.c.ozer@qmul.ac.uk

David Page
University of Durham
Durham, UK
d.c.page@durham.ac.uk

Stathis Pakis
Queen Mary, University of London
London, UK
s.pakis@qmw.ac.uk

Ari Pankiewicz
Max-Planck-Institut
Golm, Germany
apankie@aei.mpg.de

George Papadopoulos
King's College London
London, UK
gpapas@mth.kcl.ac.uk

Louis Paulot
Ecole Normale Superieure
Paris, France
paulot@lpt.ens.fr

Simon Philip
University of Edinburgh
Edinburgh, UK
simonp@maths.ed.ac.uk

Sanjay Ramgoolam
Brown University
Providence, USA
ramgosk@het.brown.edu

Miles Reid
University of Warwick
Coventry, UK
miles@maths.warwick.ac.uk

Soo-Jong Rey
Seoul National University
Seoul, Korea
sjrey@phya.snu.ac.kr

Sven Rinke
Duke University
Durham, USA
rinke@phy.duke.edu

Daniel Roggenkamp
Universitaet Bonn
Bonn, Germany
roggenka@th.physik.uni-bonn.de

Gregory Sankaran
University of Bath
Bath, UK
G.K.Sankaran@bath.ac.uk

Hisham Sati
University of Michigan
Ann Arbor, USA
hsati@umich.edu

Natalia Saulina
Princeton University
Princeton, USA
saulina@princeton.edu

Justin Sawon
University of Oxford
Oxford, UK
sawon@maths.ox.ac.uk

Sakura Schafer-Nameki
University of Cambridge
Cambridge, UK
s.schafer-nameki@damtp.cam.ac.uk

Chad(mark) Schoen
Max-Planck-Institut
Bonn, Germany
schoen@mpim-bonn.mpg.de

Emmanuel Scheidegger
Institut fuer Theoretische Physick
Wien, Austria
esche@hep.itp.tuwien.ac.at

Savdeep Sethi
University of Chicago
Chicago, USA
sethi@theory.uchicago.edu

Annamaria Sinkovics
University of Wales
Swansea, UK
a.sinkovics@swansea.ac.uk

Aninda Sinha
University of Cambridge
Cambridge, UK
a.sinha@damtp.cam.ac.uk

Daniel Stevenson
University of Adelaide
Adelaide, Australia
dstevens@maths.adelaide.edu.au

Dr Balázs Szendrői
University of Warwick
Coventry, UK
balazs@maths.warwick.ac.uk

Hiromichi Takagi
Kyoto University
Kyoto, Japan
takagi@kurims.kyoto-u.ac.jp

Richard Thomas
Imperial College, London
London, UK
richard.thomas@ic.ac.uk

Andrea Tomatis
Imperial College, London
London, UK
andrea.tomatis@ic.ac.uk

Alessandro Tomasiello
Ecole Polytechnique
Palaiseau, France
tomasiel@cpht.polytechnique.fr

Burt Totaro
University of Cambridge
Cambridge, UK
b.totaro@dpmms.cam.ac.uk

Dimitrios Tsimpis
Chalmers Institute of Technology
Goteborg, Sweden
tsimpis@fy.chalmers.se

Andrei Tyurin
Steklov Math Institute
Moscow, Russia
Tyurin@newton.kias.re.kr

Nikolaos Tziolas
University of Warwick
Coventry, UK
tziolas@maths.warwick.ac.uk

Misha Verbitsky
11-1-83 Orekhovy Blvrd
Moscow, Russia
verbit@mccme.ru

Dan Waldram
Queen Mary, University of London
London, UK
d.j.waldram@qmw.ac.uk

Brian Wecht
Univerity of California
San Diego, USA
bwecht@physics.ucsd.edu

Robert Wendt
University of Vienna
Vienna, Austria
robert.wendt@univie.ac.at

Professor Pelham Wilson
University of Cambridge
Cambridge, UK
pmhw@dpmms.cam.ac.uk

Brookie Williams
University of California
Santa Barbara, USA
brook@physics.ucsb.edu

Takehiko Yasuda
University of Tokyo
Tokyo, Japan
takehiko@tkg.att.ne.jp

Alberto Zaffaroni
Universita di Milano-Bicocca
Milano, Italy
alberto.zaffaroni@mib.infn.it

Chen-Gang Zhou
Rutgers University
Piscataway, USA
czhou@physics.rutgers.edu

Francesco Zucconi
University of Cambridge
Cambridge, UK
f.zucconi@dpmms.cam.ac.uk