Contents

0 Introduction

During the last few years, our understanding of string theory has undergone a dramatic change. The key to this development is the discovery of duality symmetries, which relate the strong and weak coupling limits of apparently different string theories. These symmetries not only relate apparently different string theories, but give us a way to compute certain strong coupling results in one string theory by mapping it to a weak coupling result in a dual string theory. In this review I shall try to give an introduction to this exciting subject. However, instead of surveying all the important developments in this subject I shall try to explain the basic ideas with the help of a few simple examples. I apologise for the inherent bias in the choice of examples and the topics, this is solely due to the varied degree of familiarity that I have with this vast subject. I have also not attempted to give a complete list of references. Instead I have only included those references whose results have been directly used or mentioned in this article. A complete list of references may be obtained by looking at the citations to some of the original papers in spires. There are also many other reviews in this subject where more references can be found[1]-[24]. I hope that this review will serve the limited purpose of initiating a person with a knowledge of perturbative string theory into this area. (For an introduction to perturbative string theory, see [25]).

The review will be divided into ten main sections as described below.

1. A brief review of perturbative string theory: In this section I shall very briefly recollect some of the results of perturbative string theory which will be useful to us in the rest of this article. This will in no way constitute an introduction to this subject; at best it will serve as a reminder to a reader who is already familiar with this subject.

2. Notion of duality symmetry: In this section I shall describe the notion of duality symmetry in string theory, a few examples of duality conjectures in string theory, and the general procedure for testing these duality conjectures.

3. Analysis of the low energy effective action: In this section I shall describe how one arrives at various duality conjectures by analyzing the low energy effective action of string theory.

4. Precision test of duality based on the spectrum of BPS states: In this section I shall discuss how one can device precision tests of various duality conjectures based on the analysis of the spectrum of a certain class of supersymmetric states in string theory.

5. Interrelation between various dualities: In this section I shall try to relate the various duality conjectures introduced in the sections 2 - 4 by 'deriving' them from a basic set of duality conjectures. I shall also discuss what we mean by relating different dualities and try to formulate the rules that must be followed during such a derivation.

6. Duality in theories with $<$ 16 supersymmetries: The discussion in sections 3-5 is focussed on string theories with at least 16 supersymmetry generators. In this section I consider theories with less number of supersymmetries. Specifically we shall focus our attention on theories with eight supercharges, which correspond to N=2 supersymmetry in four dimensions.

7. M-theory: In this section I discuss the emergence of a new theory in eleven dimensions −now known as M-theory − from the strong coupling limit of type IIA string theory. I also discuss how compactification of M-theory gives rise to new theories that cannot be regarded as perturbative compactification of a string theory.

8. F-theory: In this section I shall discuss yet another novel way of generating non-perturbative compactification of string theory based on a construction known as F-theory. This class of compactification is non-perturbative in the sense that the string coupling necessarily becomes strong in some regions of the internal compact manifold, unlike conventional compactification where the string coupling can be kept small everywhere on the internal manifold.

9. Microscopic derivation of the black hole entropy: In this section I shall discuss how many of the techniques and ideas that were used to test various duality conjectures in string theory can be used to give a microscopic derivation of the Bekenstein-Hawking entropy and Hawking radiation from black holes.

10. Matrix theory: In this final section I shall discuss a proposal for a non-perturbative

definition of M-theory and various other string theories in terms of quantum mechanics of $N \times N$ matrices in the large N limit.

Throughout this article I shall work in units where $\hbar = 1$ and $c = 1$.

1 A Brief Review of Perturbative String Theory

String theory is based on the simple idea that elementary particles, which appear as point-like objects to the present day experimentalists, are actually different vibrational modes of strings. The energy per unit length of the string, known as string tension, is parametrized as $(2\pi\alpha')^{-1}$, where α' has the dimension of $(\text{length})^2$. As we shall describe later, this theory automatically contains gravitational interaction between elmentary particles, but in order to correctly reproduce the strength of this interaction, we need to choose $\sqrt{\alpha'}$ to be of the order of $10^{-33}cm$. Since $\sqrt{\alpha'}$ is the only length parameter in the theory, the typical size of a string is of the order of $\sqrt{\alpha'} \sim 10^{-33}cm$ – a distance that cannot be resolved by present day experiments. Thus there is no direct way of testing string theory, and its appeal lies in its theoretical consistency.

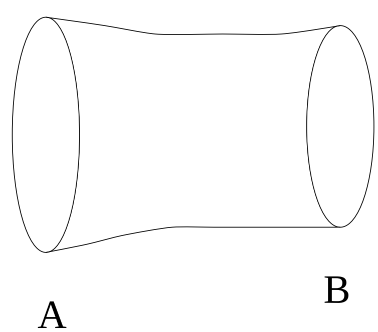


Figure 1: Propagation of a closed string.

The basic principle behind constructing a quantum theory of relativistic string is quite simple. Consider propagation of a string from a space-time configuration A to a space-time configuration B. During this motion the string sweeps out a two dimensional surface in space-time, known as the string world-sheet (see Fig.1). The amplitude for the propagation of the string from the space-time position A to space-time position B is given by the weighted sum over all world-sheet bounded by the initial and the final locations of

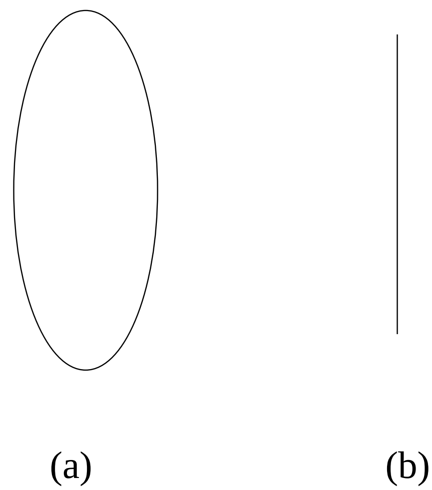


Figure 2: (a) A closed string, and (b) an open string.

the string. The weight factor is given by e^{-S} where S is the product of the string tension and the area of the world-sheet. It turns out that this procedure by itself does not give rise to a fully consistent string theory. In order to get a fully consistent string theory we need to add some internal fermionic degrees of freedom to the string and generalize the notion of area by adding new terms involving these fermionic degrees of freedom. The leads to five (apparently) different consistent string theories in (9+1) dimensional space-time, as we shall describe.

In the first quantized formalism, the dynamics of a point particle is described by quantum mechanics. Generalizing this we see that the first quantized description of a string will involve a (1+1) dimensional quantum field theory. However unlike a conventional quantum field theory where the spatial directions have infinite extent, here the spatial direction, which labels the coordinate on the string, has finite extent. It represents a compact circle if the string is closed (Fig.2(a)) and a finite line interval if the string is open (Fig.2(b)). This (1+1) dimensional field theory is known as the world-sheet theory. The fields in this (1+1) dimensional quantum field theory and the boundary conditions on these fields vary in different string theories. Since the spatial direction of the world-sheet theory has finite extent, each world-sheet field can be regarded as a collection of infinite number of harmonic oscillators labelled by the quantized momentum along this spatial direction. Different states of the string are obtained by acting on the Fock vacuum by these oscillators. This gives an infinite tower of states. Typically each string theory contains a set of massless states and an infinite tower of massive states. The massive string states typically have mass of the order of $(10^{-33}cm)^{-1} \sim 10^{19}GeV$ and are far beyond the reach of the present day accelerators. Thus the interesting part of the theory is the one involving the massless states. We shall now briefly describe the spectrum and interaction in various string theories and their compactifications.

1.1 The spectrum

There are five known fully consistent string theories in ten dimensions. They are known as type IIA, type IIB, type I, $E_8 \times E_8$ heterotic and SO(32) heterotic string theories respectively. Here we give a brief description of the degrees of freedom and the spectrum of massless states in each of these theories. We shall give the description in the so called light-cone gauge which has the advantage that all states in the spectrum are physical states.

1. Type II string theories: In this case the world-sheet theory is a free field theory containing eight scalar fields and eight Majorana fermions. These eight scalar fields are in fact common to all five string theories, and represent the eight transverse coordinates of a string moving in a nine dimensional space. It is useful to regard the eight Majorana fermions as sixteen Majorana-Weyl fermions, eight of them having left-handed chirality and the other eight having right-handed chirality. We shall refer to these as left- and right-moving fermions respectively. Both the type II string theories contain only closed strings; hence the spatial component of the world-sheet is a circle. The eight scalar fields satisfy periodic boundary condition as we go around the circle. The fermions have a choice of having periodic or anti-periodic boundary conditions. It is customary to refer to periodic boundary condition as Ramond (R) boundary condition[181] and anti-periodic boundary condition as Neveu-Schwarz (NS) boundary condition[182]. It turns out that in order to get a consistent string theory we need to include in our theory different classes of string states, some of which have periodic and some of which have anti-periodic boundary condition on the fermions. In all there are four classes of states which need to be included in the spectrum:

   - NS-NS where we put anti-periodic boundary conditions on both the left- and the right-moving fermions,

- NS-R where we put anti-periodic boundary condition on the left-moving fermions and periodic boundary condition on the right-moving fermions,
- R-NS where we put periodic boundary condition on the left-moving fermions and anti-periodic boundary condition on the right-moving fermions,
- R-R where we put anti-periodic boundary conditions on both the left- and the right-moving fermions.

Finally, we keep only about (1/4)th of the states in each sector by keeping only those states in the spectrum which have in them only even number of left-moving fermions and even number of right-moving fermions. This is known as the GSO projection[183]. The procedure has some ambiguity since in each of the four sectors we have the choice of assigning to the ground state either even or odd fermion number. Consistency of string theory rules out most of these possibilities, but at the end two possibilities remain. These differ from each other in the following way. In one possibility, the assignment of the left- and the right-moving fermion number to the left- and the right-moving Ramond ground states are carried out in an identical manner. This gives type IIB string theory. In the second possibility the GSO projections in the left- and the right-moving sector differ from each other. This theory is known as type IIA string theory.

Typically states from the Ramond sector are in the spinor representation of the SO(9,1) Lorentz algebra, whereas those from the NS sector are in the tensor representation. Since the product of two spinor representation gives us back a tensor representation, the states from the NS-NS and the RR sectors are bosonic, and those from the NS-R and R-NS sectors are fermionic. It will be useful to list the massless bosonic states in these two string theories. Since the two theories differ only in their R-sector, the NS sector bosonic states are the same in the two theories. They constitute a symmetric rank two tensor field, an anti-symmetric rank two tensor field, and a scalar field known as the dilaton.[3] The RR sector massless states of type IIA string theory consist of a vector, and a rank three anti-symmetric tensor. On the

[3]Although from string theory we get the spectrum of states, it is useful to organise the spectrum in terms of fields. In other words the spectrum of massless fields in string theory is identical to that of a free field theory with these fields.

other hand, the massless states from the RR sector of type IIB string theory consist of a scalar, a rank two anti-symmetric tensor field, and a rank four anti-symmetric tensor gauge field satisfying the constraint that its field strength is self-dual.

The spectrum of both these theories are invariant under space-time supersymmetry transformations which transform fermionic states to bosonic states and vice-versa. The supersymmetry algebra for type IIB theory is known as the chiral N=2 superalgebra and that of type IIA theory is known as the non-chiral N=2 superalgebra. Both superalgebras consist of 32 supersymmetry generators.

Often it is convenient to organise the infinite tower of states in string theory by their oscillator level defined as follows. As has already been pointed out before, the world-sheet degrees of freedom of the string can be regarded as a collection of infinite number of harmonic oscillators. For the creation operator associated with each oscillator we define the level as the absolute value of the number of units of world-sheet momentum that it creates while acting on the vacuum. The total oscillator level of a state is then the sum of the levels of all the oscillators that act on the Fock vacuum to create this state. (The Fock vacuum, in turn, is characterized by several quantum numbers, which are the momenta conjugate to the zero modes of various fields − modes carrying zero world-sheet momentum.) We can also separately define left- (right-) moving oscillator level as the contribution to the oscillator level from the left- (right-) moving bosonic and fermionic fields. Finally, if E and P denote respectively the world-sheet energy and momentum[4] then we define $L_0 = (E+P)/2$ and $\bar{L}_0 = (E-P)/2$. L_0 and $\bar{L}_0$ include contribution from the oscillators as well as from the Fock vacuum. Thus for example the total contribution to L_0 will be given by the sum of the right-moving oscillator level and the contribution to L_0 from the Fock vacuum.

2. Heterotic string theories: The world-sheet theory of the heterotic string theories consists of eight scalar fields, eight right-moving Majorana-Weyl fermions and thirty two left-moving Majorana-Weyl fermions. We have as before NS and R boundary conditions as well as GSO projection involving the right-moving fermions. Also as in the case of type II string theories, the NS sector states transform in the tensor

[4] We should distinguish between world-sheet momentum, and the momenta of the (9+1) dimensional theory. The latter are the the momenta conjugate to the zero modes of various bosonic fields in the world-sheet theory.

representation and the R sector states transform in the spinor representation of the SO(9,1) Lorentz algebra. However, unlike in the case of type II string theories, in this case the boundary condition on the left-moving fermions do not affect the Lorentz transformation properties of the state. Thus bosonic states come from states with NS boundary condition on the right-moving fermions and fermionic states come from states with R boundary condition on the right-moving fermions.

There are two possible boundary conditions on the left-moving fermions which give rise to fully consistent string theories. They are:

- SO(32) heterotic string theory: In this case we have two possible boundary conditions on the left-moving fermions: either all of them have periodic boundary condition, or all of them have anti-periodic boundary condition. In each sector we also have a GSO projection that keeps only those states in the spectrum which contain even number of left-moving fermions. The massless bosonic states in this theory consist of a symmetric rank two field, an anti-symmetric rank two field, a scalar field known as the dilaton and a set of 496 gauge fields filling up the adjoint representation of the gauge group SO(32).
- $E_8 \times E_8$ heterotic string theory: In this case we divide the thirty two left-moving fermions into two groups of sixteen each and use four possible boundary conditions, 1) all the left-moving fermions have periodic boundary condition 2) all the left-moving fermions have anti-periodic boundary condition, 3) all the left-moving fermions in group 1 have periodic boundary conditions and all the left-moving fermions in group 2 have anti-periodic boundary conditions, 4) all the left-moving fermions in group 1 have anti-periodic boundary conditions and all the left-moving fermions from group 2 have periodic boundary conditions. In each sector we also have a GSO projection that keeps only those states in the spectrum which contain even number of left-moving fermions from the first group, and also even number of left-moving fermions from the second group. The massless bosonic states in this theory consist of a symmetric rank two field, an anti-symmetric rank two field, a scalar field known as the dilaton and a set of 496 gauge fields filling up the adjoint representation of the gauge group $E_8 \times E_8$.

The spectrum of states in both the heterotic string theories are invariant under a set of space-time supersymmetry transformations. The relevant superalgebra is known as the chiral N=1 supersymmetry algebra, and has sixteen real generators.

Using the bose-fermi equivalence in (1+1) dimensions, we can reformulate both the heterotic string theories by replacing the thirty two left-moving fermions by sixteen left-moving bosons. In order to get a consistent string theory the momenta conjugate to these bosons must take discrete values. It turns out that there are only two consistent ways of quantizing the momenta, giving us back the two heterotic string theories.

3. Type I string theory: The world-sheet theory of type I theory is identical to that of type IIB string theory, with the following two crucial difference.

   - Type IIB string theory has a symmetry that exchanges the left- and the right-moving sectors in the world-sheet theory. This transformation is known as the world-sheet parity transformation. (This symmetry is not present in type IIA theory since the GSO projection in the two sectors are different). In constructing type I string theory we keep only those states in the spectrum which are invariant under this world-sheet parity transformation.
   - In type I string theory we also include open string states in the spectrum. The world-sheet degrees of freedom are identical to those in the closed string sector. Specifying the theory requires us to specify the boundary conditions on the various fields. We put Neumann boundary condition on the eight scalars, and appropriate boundary conditions on the fermions.

   The spectrum of massless bosonic states in this theory consists of a symmetric rank two tensor and a scalar dilaton from the closed string NS sector, an anti-symmetric rank two tensor from the closed string RR sector, and 496 gauge fields in the adjoint representation of SO(32) from the open string sector. This spectrum is also invariant under the chiral N=1 supersymmetry algebra with sixteen real supersymmetry generators.

1.2 Interactions

So far we have discussed the spectrum of string theory, but in order to fully describe the theory we must also describe the interaction between various particles in the spectrum.

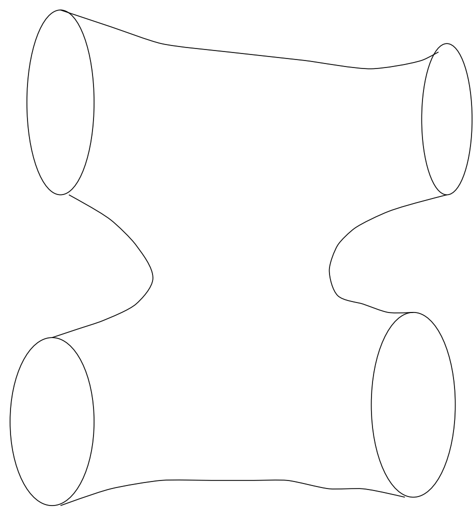


Figure 3: A string world-sheet bounded by four external strings.

In particular, we would like to know how to compute a scattering amplitude involving various string states. It turns out that there is a unique way of introducing interaction in string theory. Consider for example a scattering involving four external strings, situated along some specific curves in space-time. The prescription for computing the scattering amplitude is to compute the weighted sum over all possible string world-sheet bounded by the four strings with weight factor e^{-S}, S being the string tension multiplied by the generalized area of this surface (taking into account the fermionic degrees of freedom of the world-sheet). One such surface is shown in Fig.3. If we imagine the time axis running from left to right, then this diagram represents two strings joining into one string and then splitting into two strings, $-$ the analog of a tree diagram in field theory. A more complicated surface is shown in Fig.4. This represents two strings joining into one string, which then splits into two and joins again, and finally splits into two strings. This is the analog of a one loop diagram in field theory. The relative normalization between the contributions from these two diagrams is not determined by any consistency requirement. This introduces an arbitrary parameter in string theory, known as the string coupling constant. However, once the relative normalization between these two diagrams is fixed, the relative normalization between all other diagrams is fixed due to various consistency requirement. Thus besides the dimensionful parameter α', string theory has a single dimensionless coupling constant. As we shall see later, both these parameters can be absorbed into definitions of various fields in the theory.

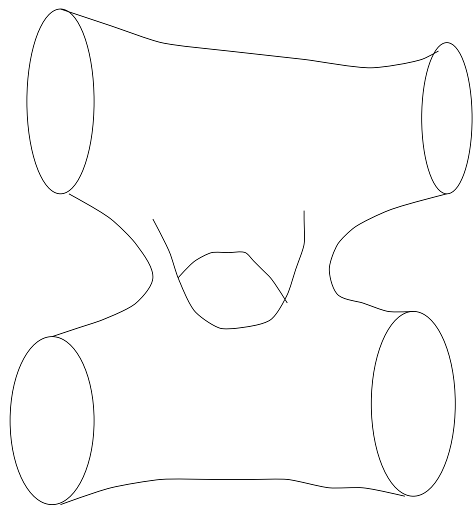


Figure 4: A more complicated string world-sheet.

What we have described so far is the computation of the scattering amplitude with fixed locations of the external strings in space-time. The more relevant quantity is the scattering amplitude where the external strings are in the eigenstates of the energy and momenta operators conjugate to the coordinates of the (9+1) dimensional space-time. This is done by simply taking the convolution of the above scattering amplitude with the wave-functions of the strings corresponding to the external states. In practice there is an extremely efficient method of doing this computation using the so called vertex operators. It turns out that unlike in quantum field theory, all of these scattering amplitudes in string theory are ultraviolet finite. This is one of the major achievements of string theory.

Out main interest will be in the scattering involving the external massless states. The most convenient way to summarize the result of this computation in any string theory is to specify the effective action. By definition this effective action is such that if we compute the *tree level* scattering amplitude using this action, we should reproduce the S-matrix elements involving the massless states of string theory. In general such an action will have to contain infinite number of terms, but we can organise these terms by examining the number of space-time derivatives that appear in a given term in the action. Terms with the lowest number of derivatives constitute the *low energy effective action*, – so called because this gives the dominant contribution if we want to evaluate the scattering amplitude when all the external particles have small energy and momenta.

The low energy effective action for all five string theories have been found. The actions for the type IIA and type IIB string theories correspond to those of two well known supergravity theories in ten space-time dimensions, called type IIA and type IIB supergravity theories respectively. On the other hand the actions for the three heterotic string theories correspond to another set of well-known supersymmetric theories in ten dimensions, $-$ $N=1$ supergravity coupled to N=1 super Yang-Mills theory. For type I and the SO(32) heterotic string theories the Yang-Mills gauge group is SO(32) whereas for the $E_8 \times E_8$ heterotic string theory the gauge group is $E_8 \times E_8$. The emergence of gravity in all the five string theories is the most striking result in string theory. Its origin can be traced to the existence of the symmetric rank two tensor state (the graviton) in all these theories. This, combined with the result on finiteness of scattering amplitudes, shows that string theory gives us a finite quantum theory of gravity. We shall explicitly write down the low energy effective action of some of the string theories in section 3.

The effective action of all five string theories are invariant under the transformation

$$\Phi \to \Phi - 2C, \qquad g_S \to e^C g_S, \tag{1.1}$$

together with possible rescaling of other fields. Here Φ denotes the dilaton field, g_S denotes the string coupling, and C is an arbitrary constant. Using this scaling property, g_S can be absorbed in Φ. Put another way, the dimensionless coupling constant in string theory is related to the vacuum expectation value $\langle\Phi\rangle$ of Φ. The perturbative effective action does not have any potential for Φ, and hence $\langle\Phi\rangle$ can take arbitrary value. One expects that in a realistic string theory where supersymmetry is spontaneouly broken, there will be a potential for Φ, and hence $\langle\Phi\rangle$ will be determined uniquely.

In a similar vain one can argue that in string theory even the string tension, or equivalently the parameter α', has no physical significance. Since α' has the dimension of (length)^2 and is the only dimensionful parameter in the theory, the effective action will have an invariance under the simultaneous rescaling of α' and the metric $g_{\mu\nu}$:

$$\alpha' \to \lambda\alpha', \qquad g_{\mu\nu} \to \lambda g_{\mu\nu}\,, \tag{1.2}$$

together with possible rescaling of other fields. Using this scaling symmetry α' can be absorbed into the definition of $g_{\mu\nu}$. We shall discuss these two rescalings in detail in section 3.1.

1.3 Compactification

So far we have described five different string theories, but they all live in ten space-time dimensions. Since our world is (3+1) dimensional, these are not realistic string theories. However one can construct string theories in lower dimensions using the idea of compactification. The idea is to take the (9+1) dimensional space-time as the product of a $(9-d)$ dimensional compact manifold $\mathcal{M}$ with euclidean signature and a $(d+1)$ dimensional Minkowski space $R^{d,1}$. Then, in the limit when the size of the compact manifold is sufficiently small so that the present day experiments cannot resolve this distance, the world will effectively appear to be $(d+1)$ dimensional. Choosing $d=3$ will give us a (3+1) dimensional theory. Of course we cannot choose any arbitrary manifold $\mathcal{M}$ for this purpose; it must satisfy the equations of motion of the effective field theory that comes out of string theory. One also normally considers only those manifolds which preserve part of the space-time supersymmetry of the original ten dimensional theory, since this guarantees vanishing of the cosmological constant, and hence consistency of the corresponding string theory order by order in perturbation theory. There are many known examples of manifolds satisfying these restrictions *e.g.* tori of different dimensions, K3, Calabi-Yau manifolds etc. Instead of going via the effective action, one can also directly describe these compactified theories as string theories. For this one needs to modify the string world-sheet action in such a way that it describes string propagation in the new manifold $\mathcal{M}\times R^{d,1}$, instead of in flat ten dimensional space-time. This modifies the world-sheet theory to an interacting non-linear σ-model instead of a free field theory. Consistency of string theory puts restriction on the kind of manifold on which the string can propagate. At the end both approaches yield identical results.

The simplest class of compact manifolds, on which we shall focus much of our attention in the rest of this article, are tori − product of circles. The effect of this compactification is to periodically identify some of the bosonic fields in the string world-sheet field theory − the fields which represent coordinates tangential to the compact circles. One effect of this is that the momentum carried by any string state along any of these circles is quantized in units of $1/R$ where R is the radius of the circle. But that is another novel effect: we now have new states that correspond to strings wrapped around a compact circle. For such a states, as we go once around the string, we also go once around the compact circle. These states are known as winding states and play a crucial role in the analysis of duality symmetries.

2 Notion of Duality Symmetries in String Theory

In this section I shall elaborate the notion of duality symmetries, the difficulties in testing them, and the way of avoiding these difficulties. We begin by introducing the notion of duality in string theory.

2.1 Duality symmetries: Definition and examples

As was described in the last section, there are five consistent string theories in ten space-time dimensions. We also saw that we can get many different string theories in lower dimensions by compactifying these five theories on appropriate manifold $\mathcal{M}$. Each of these theories is parametrized by a set of parameters known as moduli[5] *e.g.*

- String coupling constant (related to the vacuum expectation value of the dilaton field),
- Shape and size of $\mathcal{M}$ (information contained in the metric),
- various other background fields.

Inside the moduli space of the theory there is a certain region where the string coupling is weak and perturbation theory is valid. Elsewhere the theory is strongly coupled. This situation has been illustrated in fig.5.

String duality provides us with an equivalence map between two different string theories. In general this equivalence relation maps the weak coupling region of one theory to the strong coupling region of the second theory and vice versa. This situation is illustrated in fig.6.

Before we proceed, let us give a few examples of dual pairs:

- Type I and SO(32) heterotic string theories in D=10 are conjectured to be dual to each other[26, 27, 28, 29].
- Type IIA string theory compactified on K3 and heterotic string theory compactified on a four dimensional torus T^4 are conjectured to be dual to each other[30, 31, 26, 32, 33].[6]

[5]In string theory these moduli are related to vacuum expectation values of various dynamical fields and are expected to take definite values when supersymmetry is broken.

[6]Throughout this article a string theory on $\mathcal{M}$ will mean string theory in the background $\mathcal{M} \times R^{9-n,1}$ where n is the real dimension of $\mathcal{M}$, and $R^{9-n,1}$ denotes $(10-n)$ dimensional Minkowski space.

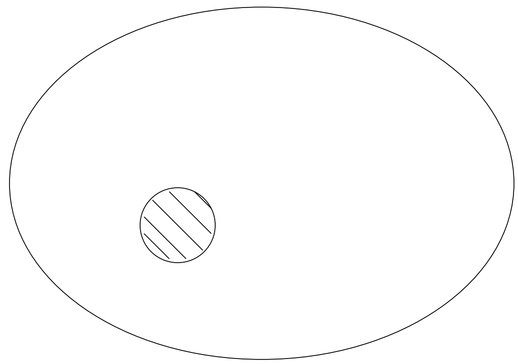


Figure 5: A schematic representation of the moduli space of a string theory. The shaded region denotes the weak coupling region, whereas the white region denotes the strong coupling region.

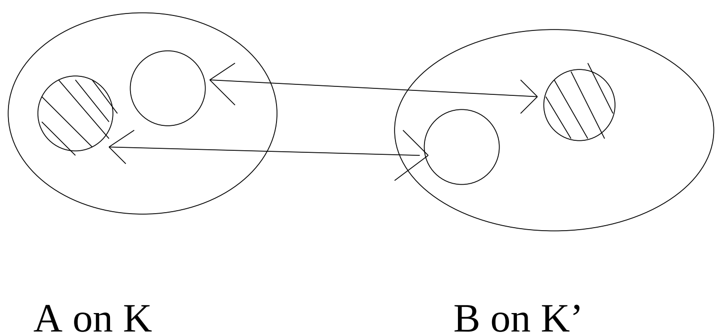


Figure 6: A schematic representation of the duality map between the moduli spaces of two different string theories, A on K and B on K', where A and B are two of the five string theories in ten dimensions, and K, K' are two compact manifolds. Under this duality the weak coupling region of the first theory (denoted by the shaded region) gets mapped to the strong coupling region of the second theory and vice versa.

Under duality, typically perturbation expansions get mixed up. Thus for example, tree level results in one theory might include perturbative and non-perturbative corrections in the dual theory. Also under duality, many of the elementary string states in one theory get mapped to solitons and their bound states in the dual theory.

Although duality in general relates the weak coupling limit of one theory to the strong coupling limit of another theory, there are special cases where the situation is a bit different. For example, we can have:

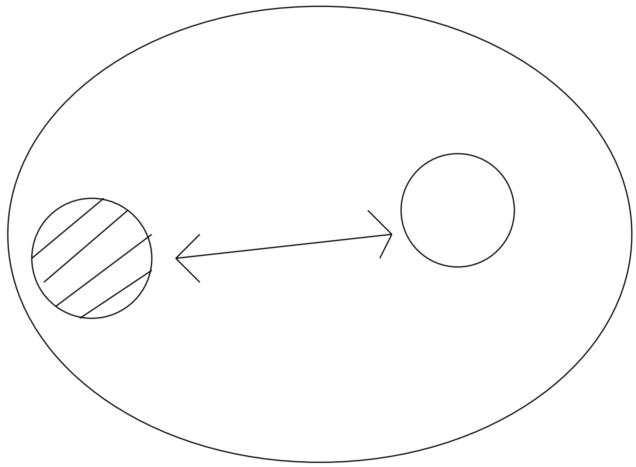


Figure 7: Schematic representation of the moduli space of a self-dual theory. Duality relates weak and strong coupling regions of the same theory.

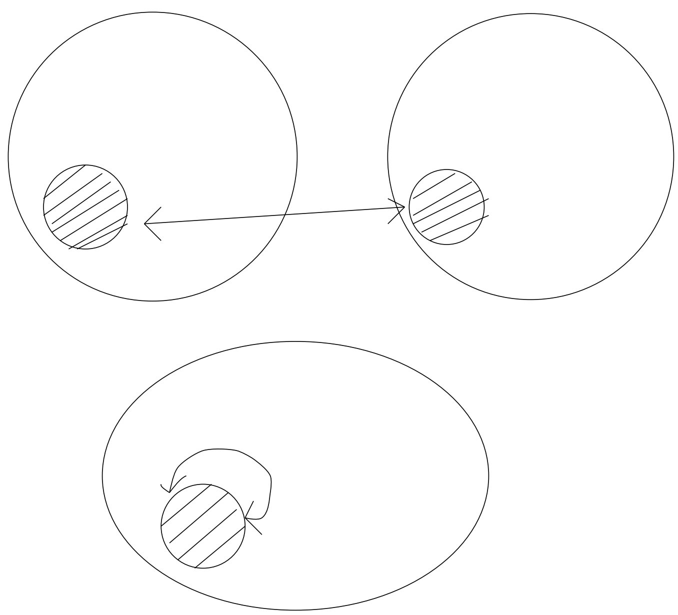


Figure 8: Examples of T-duality relating a weakly coupled theory to a different or the same weakly coupled theory.

- Self-duality: Here duality gives an equivalence relation between different regions of the moduli space of the same theory, as illustrated in fig.7. In this case, duality

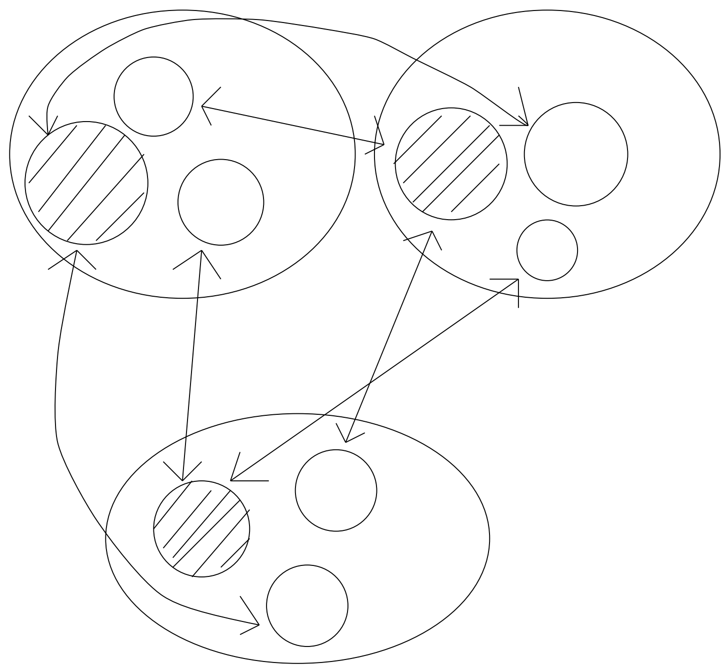


Figure 9: A schematic representation of the moduli spaces of a chain of theories related by duality. In each case the shaded region denotes weak coupling region as usual.

transformations form a symmetry group that acts on the moduli space of the theory. For example, type IIB string theory in D=10 is conjectured to have an SL(2,Z) self-duality group[30].

- T-duality: In this case duality transformation maps the weak coupling region of one theory to the weak coupling region of another theory or the same theory as illustrated in fig.8. For example, type IIA string theory compactified on a circle of radius R is dual to IIB string theory compactified on a circle of radius R^{-1} at the same value of the string coupling. Also, either of the two heterotic string theories compactified on a circle of radius R is dual to the same theory compactified on a circle of radius R^{-1} at the same value of the coupling constant. As a result the duality map does not mix up the perturbation expansions in the two theories. (For a review of this subject, see [34].)

In a generic situation duality can relate not just two theories, but a whole chain of theories, as illustrated in fig.9. Thus for example, type IIA string theory compactified on K3 is related to heterotic string theory compactified on T^4. On the other hand, due to the equivalence of the SO(32) heterotic and type I string theory in ten dimensions, SO(32) heterotic string theory compactified on T^4 is related to type I string theory compactified on T^4. Thus these three theories are related by a chain of duality transformations.

From this discussion we see that the presence of duality in string theory has two important consequences. First of all, it reduces the degree of non-uniqueness of string theory, by relating various apparently unrelated (compactified) string theories. Furthermore, it allows us to study a strongly coupled string theory by mapping it to a weakly coupled dual theory whenever such a dual theory exists.

2.2 Testing duality conjectures

Let us now turn to the question of testing duality. As we have already emphasized, duality typically relates a weakly coupled string theory to a strongly coupled string theory. Thus in order to prove / test duality we must be able to analyze at least one of the theories at strong coupling. But in string theory we only know how to define the theory perturbatively at weak coupling. Thus it would seem impossible to prove or test any duality conjecture in string theory.[7] This is where supersymmetry comes to our rescue. Supersymmetry gives rise to certain non-renormalization theorems in string theory, due to which some of the weak coupling calculations can be trusted even at strong coupling. Thus we can focus our attention on such 'non-renormalized' quantities and ask if they are invariant under the proposed duality transformations. Testing duality invariance of these quantities provides us with various tests of various duality conjectures, and is in fact the basis of all duality conjectures.

The precise content of these non-renormalization theorems depends on the number of supersymmetries present in the theory. The maximum number of supersymmetry generators that can be present in a string theory is 32. This gives N=2 supersymmetry in ten dimensions, and N=8 supersymmetry in four dimensions. Examples of such theories are type IIA or type IIB string theories compactified on n dimensional tori T^n. The next interesting class of theories are those with 16 supersymmetry generators. This corresponds to N=1 supersymmetry in ten dimensions and N=4 supersymmetry in four dimensions.

[7] Note that this problem is absent for T-duality transformations which relates two weakly coupled string theories, and hence can be tested using string perturbation theory. All T-duality symmetries in string theory can be 'proved' this way, at least to all orders in perturbation theory.

Examples of such theories are type IIA or type IIB string theories compactified on $K3 \times T^n$, heterotic string theory compactified on T^n, etc. Another class of theories that we shall discuss are those with eight supersymmetry generators, *e.g.* heterotic string theory on $K3 \times T^n$, type IIA or IIB string theory on six dimensional Calabi-Yau manifolds, etc. For theories with 16 or more SUSY generators the non-renormalization theorems are particularly powerful. In particular,

- Form of the low energy effective action involving the massless states of the theory is completely fixed by the requirement of supersymmetry (and the spectrum)[35]. Thus this effective action cannot get renormalized by string loop corrections. As a result, any valid symmetry of the theory must be a symmetry of this effective field theory.

- These theories contain special class of states which are invariant under part of the supersymmetry transformations. They are known as BPS states, named after Bogomol'nyi, Prasad and Sommerfeld. The mass of a BPS state is completely determined in terms of its charge as a consequence of the supersymmetry algebra. Since this relation is derived purely from an analysis of the supersymmetry algebra, it is not modified by quantum corrections. Furthermore it can be argued that the degeneracy of BPS states of a given charge does not change as we move in the moduli space even from weak to strong coupling region[36]. Thus the spectrum of BPS states can be calculated from weak coupling analysis and the result can be continued to the strong coupling region. Since any valid symmetry of the theory must be a symmetry of the spectrum of BPS states, we can use this to design non-trivial tests of duality[1].

For theories with eight supersymmetries the non-renormalization theorems are less powerful. However, even in this case one can design non-trivial tests of various duality conjectures. We shall discuss these in section 6.

3 Analysis of Low Energy Effective Field Theory

In this section I shall discuss tests of various dualities in string theories with ≥ 16 supersymmetries based on the analysis of their low energy effective action. As has been emphasized in the previous section, the form of this low energy effective action is determined completely by the requirement of supersymmetry and the spectrum of massless states in the theory. Thus it does not receive any quantum corrections, and if a given duality transformation is to be a symmetry of a string theory, it must be a symmetry of the corresponding low energy effective action. Actually, since the low energy *effective action* is to be used only for deriving the equations of motion from this action, and/or computing the tree level S-matrix elements using this action, but not to perform a full-fledged path integral, it is enough that only the equations of motion derived from this action are invariant under duality transformations. (This also guarantees that the tree level S-matrix elements computed from this effective action are invariant under the duality transformations.) It is not necessary for the action itself to be invariant.

Throughout this article we shall denote by $G_{\mu\nu}$ the string metric $-$ the metric that is used in computing the area of the string world-sheet embedded in space time for calculating string scattering amplitudes. For a string theory compactified on a $(9-d)$ dimensional manifold $\mathcal{M}$, we shall denote by Φ the shifted dilaton, related to the dilaton $\Phi^{(10)}$ of the ten dimensional string theory as

$$\Phi = \Phi^{(10)} - \ln V \, , \tag{3.1}$$

where $(2\pi)^{9-d}V$ is the volume of $\mathcal{M}$ measured in the ten dimensional string metric. The dilaton is normalized in such a way that $e^{\langle \Phi^{(10)} \rangle}$ corresponds to the square of the closed string coupling constant in ten dimensions.[8] $g_{\mu\nu}$ will denote the canonical Einstein metric which is related to the string metric by an appropriate conformal rescaling involving the dilaton field,

$$g_{\mu\nu} = e^{-\frac{2}{d-1}\Phi} G_{\mu\nu} \, . \tag{3.2}$$

We shall always use this metric to raise and lower indices. The signature of space-time will be taken as $(-, +, \cdots +)$. Finally, all fields will be made dimensionless by absorbing appropriate powers of α' in them.

We shall now consider several examples. The discussion will closely follow refs.[1, 30, 26]. For a detailed review of the material covered in this section, see ref.[15].

[8] Φ is related to the more commonly normalized dilaton ϕ by a factor of two: $\Phi = 2\phi$.

3.1 Type I - SO(32) heterotic duality in D=10

In SO(32) heterotic string theory, the massless bosonic states come from the NS sector of the closed heterotic string, and contains the metric $g^{(H)}_{\mu\nu}$, the dilaton $\Phi^{(H)}$, the rank two anti-symmetric tensor field $B^{(H)}_{\mu\nu}$, and gauge fields $A^{(H)a}_\mu$ ($1 \le a \le 496$) in the adjoint representation of SO(32). The low energy dynamics involving these massless bosonic fields is described by the N=1 supergravity coupled to SO(32) super Yang-Mills theory in ten dimensions[147]. The action is given by[115]:

$$\begin{aligned} S^{(H)} &= \frac{1}{(2\pi)^7(\alpha'_H)^4 g_H^2}\int d^{10}x\sqrt{-g^{(H)}}\Big[R^{(H)} - \frac{1}{8}g^{(H)\mu\nu}\partial_\mu\Phi^{(H)}\partial_\nu\Phi^{(H)} \\ &\quad -\frac{1}{4}g^{(H)\mu\mu'}g^{(H)\nu\nu'}e^{-\Phi^{(H)}/4}Tr(F^{(H)}_{\mu\nu}F^{(H)}_{\mu'\nu'}) \\ &\quad -\frac{1}{12}g^{(H)\mu\mu'}g^{(H)\nu\nu'}g^{(H)\rho\rho'}e^{-\Phi^{(H)}/2}H^{(H)}_{\mu\nu\rho}H^{(H)}_{\mu'\nu'\rho'}\Big]\,, \end{aligned} \tag{3.3}$$

where $R^{(H)}$ is the Ricci scalar, $F^{(H)}_{\mu\nu}$ denotes the non-abelian gauge field strength,

$$F^{(H)}_{\mu\nu} = \partial_\mu A^{(H)}_\nu - \partial_\nu A^{(H)}_\mu + \sqrt{\frac{2}{\alpha'_H}}\,[A^{(H)}_\mu, A^{(H)}_\nu]\,, \tag{3.4}$$

Tr denotes trace in the vector representation of SO(32), and $H^{(H)}_{\mu\nu\rho}$ is the field strength associated with the $B^{(H)}_{\mu\nu}$ field:

$$\begin{aligned} H^{(H)}_{\mu\nu\rho} &= \partial_\mu B^{(H)}_{\nu\rho} - \frac{1}{2}Tr(A^{(H)}_\mu F^{(H)}_{\nu\rho} - \frac{1}{3}\sqrt{\frac{2}{\alpha'_H}}\,A^{(H)}_\mu[A^{(H)}_\nu, A^{(H)}_\rho]) \\ &\quad + \text{cyclic permutations of } \mu,\nu,\rho\,. \end{aligned} \tag{3.5}$$

$2\pi\alpha'_H$ and g_H are respectively the inverse string tension and the coupling constant of the heterotic string theory. The rescalings (1.1), (1.2) take the following form acting on the complete set of fields:

$$\begin{aligned} &g_H \to e^C g_H, \qquad \Phi^{(H)} \to \Phi^{(H)} - 2C, \qquad g^{(H)}_{\mu\nu} \to e^{C/2}g^{(H)}_{\mu\nu} \\ &B^{(H)}_{\mu\nu} \to B^{(H)}_{\mu\nu}, \qquad A^{(H)a}_\mu \to A^{(H)a}_\mu\,, \end{aligned} \tag{3.6}$$

$$\begin{aligned} &\alpha'_H \to \lambda\alpha'_H, \qquad \Phi^{(H)} \to \Phi^{(H)}, \qquad g^{(H)}_{\mu\nu} \to \lambda g^{(H)}_{\mu\nu} \\ &B^{(H)}_{\mu\nu} \to \lambda B^{(H)}_{\mu\nu}, \qquad A^{(H)a}_\mu \to \lambda^{1/2}A^{(H)a}_\mu\,, \end{aligned} \tag{3.7}$$

Since g_H and α'_H can be changed by this rescaling, these parameters cannot have a universal significance. In particular, we can absorb g_H and α'_H into the various fields by setting $e^{-C} = g_H$ and $\lambda = (\alpha'_H)^{-1}$ in (3.6), (3.7). This is equivalent to setting $g_H = 1$ and $\alpha'_H = 1$. In this notation the physical coupling constant is given by the vacuum expectation value of $e^{\Phi^{(H)}/2}$, and the ADM mass per unit length of an infinitely long straight string, measured in the metric $e^{\langle\Phi^{(H)}\rangle/4} g^{(H)}_{\mu\nu}$ that approaches the string metric $G^{(H)}_{\mu\nu}$ far away from the string, is equal to $1/2\pi$. By changing $\langle\Phi^{(H)}\rangle$ we can get all possible values of string coupling, and using a metric that differs from the one used here by a constant multiplicative factor, we can get all possible values of the string tension.

For $\alpha'_H = 1$ and $g_H = 1$ eqs.(3.3)-(3.5) take the form:

$$\begin{aligned} S^{(H)} &= \frac{1}{(2\pi)^7}\int d^{10}x\sqrt{-g^{(H)}}\Big[R^{(H)} - \frac{1}{8} g^{(H)\mu\nu}\partial_\mu\Phi^{(H)}\partial_\nu\Phi^{(H)} \\ &\quad -\frac{1}{4} g^{(H)\mu\mu'} g^{(H)\nu\nu'} e^{-\Phi^{(H)}/4} Tr(F^{(H)}_{\mu\nu} F^{(H)}_{\mu'\nu'}) \\ &\quad -\frac{1}{12} g^{(H)\mu\mu'} g^{(H)\nu\nu'} g^{(H)\rho\rho'} e^{-\Phi^{(H)}/2} H^{(H)}_{\mu\nu\rho} H^{(H)}_{\mu'\nu'\rho'}\Big]\,, \end{aligned} \tag{3.8}$$

$$F^{(H)}_{\mu\nu} = \partial_\mu A^{(H)}_\nu - \partial_\nu A^{(H)}_\mu + \sqrt{2}[A^{(H)}_\mu, A^{(H)}_\nu]\,, \tag{3.9}$$

$$\begin{aligned} H^{(H)}_{\mu\nu\rho} &= \partial_\mu B^{(H)}_{\nu\rho} - \frac{1}{2} Tr\Big(A^{(H)}_\mu F^{(H)}_{\nu\rho} - \frac{\sqrt{2}}{3} A^{(H)}_\mu [A^{(H)}_\nu, A^{(H)}_\rho]\Big) \\ &\quad + \text{cyclic permutations of } \mu, \nu, \rho\,. \end{aligned} \tag{3.10}$$

Let us now turn to the type I string theory. The massless bosonic states in type I theory come from three different sectors. The closed string Neveu-Schwarz $-$ Neveu-Schwarz (NS) sector gives the metric $g^{(I)}_{\mu\nu}$ and the dilaton $\Phi^{(I)}$. The closed string Ramond-Ramond (RR) sector gives an anti-symmetric tensor field $B^{(I)}_{\mu\nu}$. Besides these, there are bosonic fields coming from the NS sector of the open string. This sector gives rise to gauge fields $A^{(I)a}_\mu$ ($a = 1, \ldots 496$) in the adjoint representation of the group SO(32). (The superscript (I) refers to the fact that these are the fields in the type I string theory.) The low energy dynamics is again described by the N=1 supergravity theory coupled to SO(32) super Yang-Mills theory[148]. But it is instructive to rewrite the effective action in terms of the type I variables. For suitable choice of the string tension and the coupling constant, this is given by[115]

$$S^{(I)} = \frac{1}{(2\pi)^7}\int d^{10}x\sqrt{-g^{(I)}}\Big[R^{(I)} - \frac{1}{8} g^{(I)\mu\nu}\partial_\mu\Phi^{(I)}\partial_\nu\Phi^{(I)}$$

$$
\begin{aligned}
&-\frac{1}{4}g^{(I)\mu\mu'}g^{(I)\nu\nu'}e^{\Phi^{(I)}/4}Tr(F^{(I)}_{\mu\nu}F^{(I)}_{\mu'\nu'}) \\
&-\frac{1}{12}g^{(I)\mu\mu'}g^{(I)\nu\nu'}g^{(I)\rho\rho'}e^{\Phi^{(I)}/2}H^{(I)}_{\mu\nu\rho}H^{(I)}_{\mu'\nu'\rho'}\Big] ,
\end{aligned}
\tag{3.11}
$$

where $R^{(I)}$ is the Ricci scalar, $F^{(I)}_{\mu\nu}$ denotes the non-abelian gauge field strength,

$$
F^{(I)}_{\mu\nu} = \partial_\mu A^{(I)}_\nu - \partial_\nu A^{(I)}_\mu + \sqrt{2}[A^{(I)}_\mu, A^{(I)}_\nu] , \tag{3.12}
$$

and $H^{(I)}_{\mu\nu\rho}$ is the field strength associated with the $B^{(I)}_{\mu\nu}$ field:

$$
\begin{aligned}
H^{(I)}_{\mu\nu\rho} \;\; = \;\; & \partial_\mu B^{(I)}_{\nu\rho} - \frac{1}{2}Tr\Big(A^{(I)}_\mu F^{(I)}_{\nu\rho} - \frac{\sqrt{2}}{3}A^{(I)}_\mu[A^{(I)}_\nu, A^{(I)}_\rho]\Big) \\
& + \text{cyclic permutations of } \mu, \nu, \rho \, .
\end{aligned}
\tag{3.13}
$$

For both, the type I and the SO(32) heterotic string theory, the low energy effective action is derived from the string tree level analysis. However, to this order in the derivatives, the form of the effective action is determined completely by the requirement of supersymmetry for a given gauge group. Thus neither action can receive any quantum corrections.

It is straightforward to see that the actions (3.8) and (3.11) are identical provided we make the identification:

$$
\begin{aligned}
\Phi^{(H)} = -\Phi^{(I)}, \qquad & g^{(H)}_{\mu\nu} = g^{(I)}_{\mu\nu} \\
B^{(H)}_{\mu\nu} = B^{(I)}_{\mu\nu}, \qquad & A^{(H)a}_\mu = A^{(I)a}_\mu \, .
\end{aligned}
\tag{3.14}
$$

This led to the hypothesis that the type I and the SO(32) heterotic string theories in ten dimensions are equivalent[26]. One can find stronger evidence for this hypothesis by analysing the spectrum of supersymmetris states, but the equivalence of the two effective actions was the reason for proposing this duality in the first place.

Note the $-$ sign in the relation between $\Phi^{(H)}$ and $\Phi^{(I)}$ in eq.(3.14). Recalling that $e^{\langle\Phi\rangle/2}$ is the string coupling, we see that the strong coupling limit of one theory is related to the weak coupling limit of the other theory and vice versa.

From now on I shall use the unit $\alpha' = 1$ for writing down the effective action of all string theories. Physically this would mean that the ADM mass per unit length of a test string, measured in the metric $e^{2\langle\Phi\rangle/(d-1)}g_{\mu\nu}$ that agrees with the string metric $G_{\mu\nu}$

defined in (3.2) far away from the test string, is given by $1/2\pi$. In future we shall refer to the ADM mass of a particle measured in this metric as the mass measured in the string metric.

3.2 Self-duality of heterotic string theory on T^6

In the previous subsection we have described the massless bosonic field content of the ten dimensional SO(32) heterotic string theory. When we compactify it on a six dimensional torus, we can get many other massless scalar fields from the internal components of the metric, the anti-symmetric tensor field and the gauge fields in the Cartan subalgebra of the gauge group.[9] This gives a total of (21+15+96=132) scalar fields. It turns out that these scalars can be represented by a 28×28 matrix valued field M satisfying[10]

$$MLM^T = L, \qquad M^T = M\,, \tag{3.15}$$

where

$$L = \begin{pmatrix} & I_6 & \\ I_6 & & \\ & & -I_{16} \end{pmatrix}. \tag{3.16}$$

I_n denotes an $n \times n$ identity matrix. We shall choose a convention in which $M = I_{28}$ corresponds to a compactification on $(S^1)^6$ with each S^1 having radius $\sqrt{\alpha'} = 1$ measured in the string metric, and without any background gauge or antisymmetric tensor fields. We can get another scalar field a by dualizing the gauge invariant field strength H of the antisymmetrix tensor field through the relation:

$$H^{\mu\nu\rho} = -(\sqrt{-g})^{-1} e^{2\Phi} \epsilon^{\mu\nu\rho\sigma} \partial_\sigma a\,, \tag{3.17}$$

where Φ denotes the four dimensional dilaton and $g_{\mu\nu}$ denotes the (3+1) dimensional canonical metric defined in eqs.(3.1), (3.2) respectively. It is convenient to combine the dilaton Φ and the axion field a into a single complex scalar λ:

$$\lambda = a + ie^{-\Phi} \equiv \lambda_1 + i\lambda_2\,. \tag{3.18}$$

At a generic point in the moduli space, where the scalars M take arbitrary vacuum expectation values, the non-abelian gauge symmetry of the ten dimensional theory is

[9] Only the sixteen gauge fields in the Cartan subalgebra of the gauge group can develop vacuum expectation value since such vacuum expectation values do not generate any field strength, and hence do not generate energy density.

[10] For a review of this construction, see [1].

broken to its abelian subgroup $U(1)^{16}$. Besides these sixteen U(1) gauge fields we get twelve other U(1) gauge fields from components $G_{m\mu}$, $B_{m\mu}$ ($4 \le m \le 9$, $0 \le \mu \le 3$) of the metric and the anti-symmetric tensor field respectively. Let us denote these 28 U(1) gauge fields (after suitable normalization) by A^a_μ ($1 \le a \le 28$). In terms of these fields, the low energy effective action of the theory is given by[37, 38, 39, 41, 1],[11]

$$\begin{aligned} S &= \frac{1}{2\pi}\int d^4x\sqrt{-g}\Big[R - g^{\mu\nu}\frac{\partial_\mu\lambda\partial_\nu\bar{\lambda}}{2(\lambda_2)^2} + \frac{1}{8}g^{\mu\nu}Tr(\partial_\mu ML\partial_\nu ML) \\ &\quad -\frac{1}{4}\lambda_2 g^{\mu\mu'}g^{\nu\nu'}F^a_{\mu\nu}(LML)_{ab}F^b_{\mu'\nu'} + \frac{1}{4}\lambda_1 g^{\mu\rho}g^{\nu\sigma}F^a_{\mu\nu}L_{ab}\widetilde{F}^b_{\rho\sigma}\Big]\,, \end{aligned} \tag{3.19}$$

where $F^a_{\mu\nu}$ is the field strength associated with A^a_μ, R is the Ricci scalar. and

$$\widetilde{F}^{a\mu\nu} = \frac{1}{2}(\sqrt{-g})^{-1}\epsilon^{\mu\nu\rho\sigma}F^a_{\rho\sigma}\,. \tag{3.20}$$

This action is invariant under an O(6,22) transformation:[12]

$$M \to \Omega M\Omega^T, \qquad A^a_\mu \to \Omega_{ab}A^b_\mu, \qquad g_{\mu\nu} \to g_{\mu\nu}, \qquad \lambda \to \lambda\,, \tag{3.21}$$

where Ω satisfies:

$$\Omega L\Omega^T = L\,. \tag{3.22}$$

An O(6,22;Z) subgroup of this can be shown to be a T-duality symmetry of the full string theory[34]. This O(6,22;Z) subgroup can be described as follows. Let Λ_{28} denote a twenty eight dimensional lattice obtained by taking the direct sum of the twelve dimensional lattice of integers, and the sixteen dimensional root lattice of SO(32).[13] O(6,22;Z) is defined to be the subset of O(6,22) transformations which leave Λ_{28} invariant, *i.e.* acting on any vector in Λ_{28}, produces another vector in Λ_{28}. It will be useful for our future reference to undertstand why only an O(6,22;Z) subgroup of the full O(6,22) group is a symmetry of the full string theory. Since O(6,22;Z) is a T-duality symmetry, this question can be answered within the context of perturbative string theory. The point is

[11] The normalization of the gauge fields used here differ from that in ref.[1] by a factor of two. Also there we used $\alpha' = 16$ whereas here we are using $\alpha' = 1$.

[12] $O(p,q)$ denotes the group of Lorentz transformations in p space-like and q time-like dimensions. (These have nothing to do with physical space-time, which always has only one time-like direction.) $O(p,q;Z)$ denotes a discrete subgroup of $O(p,q)$.

[13] More precisely we have to take the root lattice of $Spin(32)/Z_2$ which is obtained by adding to the SO(32) root lattice the weight vectors of the spinor representations of SO(32) with a definite chirality.

that although at a generic point in the moduli space the massless string states do not carry any charge, there are massive charged states in the spectrum of full string theory. Since there are 28 charges associated with the 28 U(1) gauge fields, a state can be characterized by a 28 dimensional charge vector. With appropriate normalization, this charge vector can be shown to lie in the lattice Λ_{28}, *i.e.* the charge vector of any state in the spectrum can be shown to be an element of the lattice Λ_{28}. Since the O(6,22) transformation acts linearly on the U(1) gauge fields, it also acts linearly on the charge vectors. As a result only those O(6,22) elements can be genuine symmetries of string theory which preserve the lattice Λ_{28}. Any other O(6,22) element, acting on a physical state in the spectrum, will take it to a state with charge vector outside the lattice Λ_{28}. Since such a state does not exist in the spectrum, such an O(6,22) transformation cannot be a symmetry of the full string theory.

In order to see a specific example of a T-duality transformation, let us consider heterotic string theory compactified on $(S^1)^6$ with one of the circles having radius R measured in the string metric, and the rest having unit radius. Let us also assume that there is no background gauge or anti-symmetric tensor fields. Using the convention of ref.[1] one can show that for this background

$$M^{(H)} = \begin{pmatrix} R^{-2} & & & & \\ & I_5 & & & \\ & & R^2 & & \\ & & & I_5 & \\ & & & & I_{16} \end{pmatrix} . \qquad (3.23)$$

Consider now the O(6,22;Z) transformation with the matrix:

$$\Omega = \begin{pmatrix} 0 & & 1 & \\ & I_5 & & \\ 1 & & 0 & \\ & & & I_{21} \end{pmatrix} . \qquad (3.24)$$

Using eq.(3.21) we see that this transforms $M^{(H)}$ to

$$M^{(H)} = \begin{pmatrix} R^{2} & & & & \\ & I_5 & & & \\ & & R^{-2} & & \\ & & & I_5 & \\ & & & & I_{16} \end{pmatrix} . \qquad (3.25)$$

Thus the net effect of this transformation is $R \to R^{-1}$. It says that the heterotic string theory compactified on a circle of radius R is equivalent to the same theory compactified on a circle of radius R^{-1}. For this reason $R = 1$ (*i.e.* $R = \sqrt{\alpha'}$) is known as the self-dual radius. Other O(6,22;Z) transformations acting on (3.23) will give rise to more complicated $M^{(H)}$ corresponding to a configuration with background gauge and / or anti-symmetric tensor fields.

Besides this symmetry, the equations of motion derived from this action can be shown to be invariant under an $SL(2,R)$ transformation of the form[37, 42, 43]

$$F^a_{\mu\nu} \to (r\lambda_1 + s)F^a_{\mu\nu} + r\lambda_2(ML)_{ab}\widetilde{F}^b_{\mu\nu}, \qquad \lambda \to \frac{p\lambda + q}{r\lambda + s},$$
$$g_{\mu\nu} \to g_{\mu\nu}, \qquad M \to M\,, \tag{3.26}$$

where p, q, r, s are real numbers satisfying $ps - qr = 1$. The existence of such symmetries (known as hidden non-compact symmetries) in this and in other supergravity theories were discovered in early days of supergravity theories and in fact played a crucial role in the construction of these theories in the first place[146, 37]. Since this SL(2,R) transformation mixes the gauge field strength with its Poincare dual, it is an electric-magnetic duality transformation. This leads to the conjecture that a subgroup of this continuous symmetry group is an exact symmetry of string theory[44, 45, 43, 46, 47, 48, 49, 1]. One might wonder why the conjecture refers to only a discrete subgroup of SL(2,R) instead of the full SL(2,R) group as the genuine symmetry group. This follows from the same logic that was responsible for breaking O(6,22) to O(6,22;Z); however since the SL(2,R) transformation mixes electric field with magnetic field, we now need to take into account the quantization of magnetic charges. We have already described the quantization condition on the electric charges. Using the usual Dirac-Schwinger-Zwanziger rules one can show that in appropriate normalization, the 28 dimensional magnetic charge vectors also lie in the same lattice Λ_{28}. Also with this normalization convention the electric and magnetic charge vectors transform as doublet under the SL(2,R) transformation; thus it is clear that the subgroup of SL(2,R) that respects the charge quantization condition is SL(2,Z). An arbitrary SL(2,R) transformation acting on the quantized electric and magnetic charges will not give rise to electric and magnetic charges consistent with the quantization law. This is the reason behind the conjectured SL(2,Z) symmetry of heterotic string theory on T^6. Note that since this duality acts non-trivially on the dilaton and hence the string coupling, this is a non-perturbative symmetry, and cannot be verified order by order in perturbation theory. Historically, this is the first example of a concrete duality conjecture in string theory. Later we shall review other tests of this duality conjecture.

3.3 Duality between heterotic on T^4 and type IIA on K3

The massless bosonic field content of heterotic string theory compactified on T^4 can be found in a manner identical to that in heterotic string theory on T^6. Besides the dilaton $\Phi^{(H)}$, we get many other massless scalar fields from the internal components of the metric, the anti-symmetric tensor field and the gauge fields. In this case these scalars can be represented by a 24×24 matrix valued field $M^{(H)}$ satisfying

$$M^{(H)} L M^{(H)T} = L, \qquad M^{(H)T} = M^{(H)}\,, \tag{3.27}$$

where

$$L = \begin{pmatrix} & I_4 & \\ I_4 & & \\ & & -I_{16} \end{pmatrix}\,. \tag{3.28}$$

We again use the convention that $M^{(H)} = I_{24}$ corresponds to compactification on $(S^1)^4$ with each S^1 having self-dual radius ($\sqrt{\alpha'} = 1$), without any background gauge field or anti-symmetric tensor field. At a generic point in the moduli space, where the scalars $M^{(H)}$ take arbitrary vacuum expectation values, we get a $U(1)^{24}$ gauge group, with 16 gauge fields coming from the Cartan subalgebra of the original gauge group in ten dimensions, and eight other gauge fields from components $G_{m\mu}$, $B_{m\mu}$ ($6 \le m \le 9$, $0 \le \mu \le 5$) of the metric and the anti-symmetric tensor field respectively. Here x^m denote the compact directions, and x^μ denote the non-compact directions. Let us denote these 24 U(1) gauge fields by $A^{(H)a}_\mu$ ($1 \le a \le 24$). Finally, let $g^{(H)}_{\mu\nu}$ and $B^{(H)}_{\mu\nu}$ denote the canonical metric and the anti-symmetric tensor field respectively. In terms of these fields, the low energy effective action of the theory is given by,

$$\begin{aligned} S_H \;=\; & \frac{1}{(2\pi)^3} \int d^6x \sqrt{-g^{(H)}} \Big[R^{(H)} - \frac{1}{2} g^{(H)\mu\nu} \partial_\mu \Phi^{(H)} \partial_\nu \Phi^{(H)} \\ & + \frac{1}{8} g^{\mu\nu} Tr(\partial_\mu M^{(H)} L \partial_\nu M^{(H)} L) \\ & - \frac{1}{4} e^{-\Phi^{(H)}/2} g^{(H)\mu\mu'} g^{(H)\nu\nu'} F^{(H)a}_{\mu\nu} (L M^{(H)} L)_{ab} F^{(H)b}_{\mu'\nu'} \\ & - \frac{1}{12} e^{-\Phi^{(H)}} g^{(H)\mu\mu'} g^{(H)\nu\nu'} g^{(H)\rho\rho'} H^{(H)}_{\mu\nu\rho} H^{(H)}_{\mu'\nu'\rho'} \Big]\,, \end{aligned} \tag{3.29}$$

where $F^{(H)a}_{\mu\nu}$ is the field strength associated with $A^{(H)a}_\mu$, $R^{(H)}$ is the Ricci scalar, and $H^{(H)}_{\mu\nu\rho}$ is the field strength associated with $B^{(H)}_{\mu\nu}$:

$$H^{(H)}_{\mu\nu\rho} = (\partial_\mu B^{(H)}_{\nu\rho} + \frac{1}{2} A^{(H)a}_\mu L_{ab} F^{(H)b}_{\nu\rho}) + (\text{cyclic permutations of } \mu, \nu, \rho). \tag{3.30}$$

This action is invariant under an O(4,20) transformation:

$$M^{(H)} \to \Omega M^{(H)} \Omega^T, \qquad A_\mu^{(H)a} \to \Omega_{ab} A_\mu^{(H)b}, \qquad g_{\mu\nu}^{(H)} \to g_{\mu\nu}^{(H)},$$
$$B_{\mu\nu}^{(H)} \to B_{\mu\nu}^{(H)}, \qquad \Phi^{(H)} \to \Phi^{(H)}\,, \tag{3.31}$$

where Ω satisfies:

$$\Omega L \Omega^T = L\,. \tag{3.32}$$

Again as in the case of T^6 compactification, only an O(4,20;Z) subgroup of this which preserves the charge lattice Λ_{24} is an exact T-duality symmetry of this theory. The lattice Λ_{24} is obtained by taking the direct sum of the 8 dimensional lattice of integers and the root lattice of $Spin(32)/Z_2$.

Let us now turn to the spectrum of massless bosonic fields in type IIA string theory on K3. In ten dimensions the massless bosonic fields in type IIA string theory are the metric g_{MN}, the rank two anti-symmetric tensor B_{MN} and the scalar dilation Φ coming from the NS sector, and a gauge field A_M and a rank three antisymmetric tensor field C_{MNP} coming from the RR sector. The low energy effective action of this theory involving the massless bosonic fields is given by[97]

$$\begin{aligned} S_{IIA} \;\; = \;\; & \frac{1}{(2\pi)^7}\int d^{10}x\sqrt{-g}\Big[R - \frac{1}{8}g^{\mu\nu}\partial_\mu\Phi\partial_\nu\Phi \\ & -\frac{1}{12}e^{-\Phi/2}g^{\mu\mu'}g^{\nu\nu'}g^{\rho\rho'}H_{\mu\nu\rho}H_{\mu'\nu'\rho'} - \frac{1}{4}e^{3\Phi/4}g^{\mu\mu'}g^{\nu\nu'}F_{\mu\nu}F_{\mu'\nu'} \\ & -\frac{1}{48}e^{\Phi/4}g^{\mu\mu'}g^{\nu\nu'}g^{\rho\rho'}g^{\sigma\sigma'}G_{\mu\nu\rho\sigma}G_{\mu'\nu'\rho'\sigma'} \\ & -\frac{1}{(48)^2}(\sqrt{-g})^{-1}\varepsilon^{\mu_0\cdots\mu_9}B_{\mu_0\mu_1}G_{\mu_2\cdots\mu_5}G_{\mu_6\cdots\mu_9}\Big]\,, \end{aligned} \tag{3.33}$$

where R is the Ricci scalar, and

$$\begin{aligned} F_{\mu\nu} &= \partial_\mu A_\nu - \partial_\nu A_\mu, \\ H_{\mu\nu\rho} &= \partial_\mu B_{\nu\rho} + \text{cyclic permutations of } \mu,\, \nu,\, \rho\,, \\ G_{\mu\nu\rho} &= \partial_\mu C_{\nu\rho\sigma} + A_\mu H_{\nu\rho\sigma} + (-1)^P \cdot \text{cyclic permutations}\,, \end{aligned} \tag{3.34}$$

are the field strengths associated with A_μ, $B_{\mu\nu}$ and $C_{\mu\nu\rho}$ respectively. Upon compactification on $K3$ we get a new set of scalar fields from the Kahler and complex structure moduli of K3. These can be regarded as deformations of the metric and give a total of

58 real scalar fields. We get 22 more scalar fields $\phi^{(p)}$ by decomposing the antisymmetric tensor field B_{MN} along the twenty two harmonic two forms $\omega^{(p)}_{mn}$ in K3:

$$B_{mn}(x,y) \sim \sum_{p=1}^{22} \phi_p(x)\omega^{(p)}_{mn}(y) + \cdots . \tag{3.35}$$

Here $\{x^\mu\}$ and $\{y^m\}$ denote coordinates along the non-compact and K3 directions respectively. These eighty scalar fields together parametrize a coset $O(4,20)/O(4)\times O(20)$ and can be described by a matrix $M^{(A)}$ satisfying properties identical to those of $M^{(H)}$ described in (3.27). This theory also has twenty four U(1) gauge fields. 22 of the gauge fields arise from the components of the three form field C_{MNP}:

$$C_{mn\mu}(x,y) = \sum_{p=1}^{22} \omega^{(p)}_{mn}(y)\mathcal{A}^{(p)}_\mu(x) + \dots . \tag{3.36}$$

$\mathcal{A}^{(p)}_\mu$ defined in (3.36) behaves as gauge fields in six dimensions. One more gauge field comes from the original RR gauge field A_μ. The last one $\mathcal{A}_\mu$ comes from dualizing $C_{\mu\nu\rho}$:

$$G \sim \ ^*(d\mathcal{A}) , \tag{3.37}$$

where * denotes Poincare dual in six dimensions. Together we shall denote these gauge fields by $A^{(A)a}_\mu$ for $1 \le a \le 24$. Besides these fields, the theory contains the canonical metric and the anti-symmetric tensor field which we shall denote by $g^{(A)}_{\mu\nu}$ and $B^{(A)}_{\mu\nu}$ respectively. The action involving these fields is given by,

$$\begin{aligned} S_A \; = \; & \frac{1}{(2\pi)^3}\int d^6x\sqrt{-g^{(A)}}\Big[R^{(A)} - \frac{1}{2}g^{(A)\mu\nu}\partial_\mu\Phi^{(A)}\partial_\nu\Phi^{(A)} \\ & +\frac{1}{8}g^{\mu\nu}Tr(\partial_\mu M^{(A)}L\partial_\nu M^{(A)}L) \\ & -\frac{1}{4}e^{\Phi^{(A)}/2}g^{(A)\mu\mu'}g^{(A)\nu\nu'}F^{(A)a}_{\mu\nu}(LM^{(A)}L)_{ab}F^{(A)b}_{\mu'\nu'} \\ & -\frac{1}{12}e^{-\Phi^{(A)}}g^{(A)\mu\mu'}g^{(A)\nu\nu'}g^{(A)\rho\rho'}H^{(A)}_{\mu\nu\rho}H^{(A)}_{\mu'\nu'\rho'} \\ & -\frac{1}{16}\varepsilon^{\mu\nu\rho\delta\epsilon\eta}(\sqrt{-g^{(A)}})^{-1}B^{(A)}_{\mu\nu}F^{(A)a}_{\rho\delta}L_{ab}F^{(A)b}_{\epsilon\eta}\Big] , \end{aligned} \tag{3.38}$$

where $F^{(A)a}_{\mu\nu}$ is the field strength associated with $A^{(A)a}_\mu$, $R^{(A)}$ is the Ricci scalar, and $H^{(A)}_{\mu\nu\rho}$ is the field strength associated with $B^{(A)}_{\mu\nu}$:

$$H^{(A)}_{\mu\nu\rho} = \partial_\mu B^{(A)}_{\nu\rho} + (\text{cyclic permutations of } \mu,\nu,\rho) . \tag{3.39}$$

In writing down the above action we have used the convention that $M^{(A)} = I_{24}$ corresponds to compactification on a specific reference K3, possibly with specific background B_{mn} fields. This action has an O(4,20) symmetry of the form:

$$
\begin{aligned}
& M^{(A)} \to \Omega M^{(A)} \Omega^T, \qquad A^{(A)a}_\mu \to \Omega_{ab} A^{(A)b}_\mu, \qquad g^{(A)}_{\mu\nu} \to g^{(A)}_{\mu\nu}, \\
& B^{(A)}_{\mu\nu} \to B^{(A)}_{\mu\nu}, \qquad \Phi^{(A)} \to \Phi^{(A)}\,,
\end{aligned} \tag{3.40}
$$

where Ω satisfies:

$$
\Omega L \Omega^T = L\,. \tag{3.41}
$$

An O(4,20;Z) subgroup of this can be shown to be an exact T-duality symmetry of string theory[184]. The lattice Λ'_{24} which is preserved by this O(4,20;Z) subgroup of O(4,20) is not the lattice Λ_{24} defined earlier, but is in general an O(4,20) rotation of that lattice:

$$
\Lambda'_{24} = \Omega_0 \Lambda_{24}\,. \tag{3.42}
$$

Ω_0 depends on the choice of the special reference K3 mentioned earlier.

It is now a straightforward exercise to show that the equations of motion and the Bianchi identities derived from (3.29) and (3.38) are identical if we use the following map between the heterotic and the type II variables[50, 30]:

$$
\begin{aligned}
& g^{(H)}_{\mu\nu} = g^{(A)}_{\mu\nu}, \qquad M^{(H)} = \widetilde\Omega M^{(A)} \widetilde\Omega^T, \\
& \Phi^{(H)} = -\Phi^{(A)}, \qquad A^{(H)a}_\mu = \widetilde\Omega_{ab} A^{(A)a}_\mu, \\
& \sqrt{-g^{(H)}} \exp(-\Phi^{(H)}) H^{(H)\mu\nu\rho} = \frac{1}{6} \varepsilon^{\mu\nu\rho\delta\epsilon\eta} H^{(A)}_{\delta\epsilon\eta}\,.
\end{aligned} \tag{3.43}
$$

where $\widetilde\Omega$ is an arbitrary O(4,20) matrix. This leads to the conjectured equivalence between heterotic string theory compactified on T^4 and type IIA string theory compactified on $K3$[30]. But clearly the two theories cannot be equivalent for all $\widetilde\Omega$ since in the individual theories the O(4,20) symmetry is broken down to O(4,20;Z). $\widetilde\Omega$ can be found (up to an O(4,20;Z) transformation) by comparing the T-duality symmetry transformations in the two theories. To do this let us note that according to eq.(3.43) a transformation $M^{(H)} \to \Omega M^{(H)} \Omega^T$ will induce a transformation

$$
M^{(A)} \to (\widetilde\Omega^{-1} \Omega \widetilde\Omega) M^{(A)} (\widetilde\Omega^{-1} \Omega \widetilde\Omega)^T\,. \tag{3.44}
$$

Thus if Ω preserves the lattice Λ_{24}, $\widetilde\Omega^{-1}\Omega\widetilde\Omega$ should preserve the lattice $\Lambda'_{24} = \Omega_0 \Lambda_{24}$. This happens if we choose:

$$
\widetilde\Omega = \Omega_0^{-1}\,. \tag{3.45}
$$

Note again that there is a relative minus sign that relates $\Phi^{(H)}$ and $\Phi^{(A)}$, showing that the strong coupling limit of one theory corresponds to the weak coupling limit of the other theory.

3.4 SL(2,Z) self-duality of Type IIB in D=10

As described in section 1.1, the massless bosonic fields in type IIB string theory come from two sectors, – Neveu-Schwarz–Neveu-Schwarz (NS) and Ramond-Ramond (RR). The NS sector gives the graviton described by the metric $g_{\mu\nu}$, an anti-symmetric tensor field $B_{\mu\nu}$, and a scalar field Φ known as the dilaton. The RR sector contributes a scalar field a sometimes called the axion, another rank two anti-symmetric tensor field$B'_{\mu\nu}$, and a rank four anti-symmetric tensor field$D_{\mu\nu\rho\sigma}$ whose field strength is self-dual.

It is often convenient to combine the axion and the dilaton into a complex scalar field λ as follows:[14]

$$\lambda = a + ie^{-\Phi/2} \equiv \lambda_1 + i\lambda_2 \,. \tag{3.46}$$

The low energy effective action in this theory can be determined either from the requirement of supersymmetry, or by explicit computation in string theory. Actually it turns out that there is no simple covariant action for this low energy theory, but there are covariant field equations[51], which are in fact just the equations of motion of type IIB supergravity. Although in string theory this low energy theory is derived from the tree level analysis, non-renormalization theorems tell us that this is exact to this order in the space-time derivatives. Basically supersymmetry determines the form of the equations of motion to this order in the derivatives completely, and so there is no scope for the quantum corrections to change the form of the action.

For the sake of brevity, we shall not explicitly write down the equations of motion. The main point is that these equations of motion are covariant (in the sense that they transform into each other) under an SL(2,R) transformation[51]:

$$\lambda \to \frac{p\lambda + q}{r\lambda + s}, \qquad \begin{pmatrix} B_{\mu\nu} \\ B'_{\mu\nu} \end{pmatrix} \to \begin{pmatrix} p & q \\ r & s \end{pmatrix} \begin{pmatrix} B_{\mu\nu} \\ B'_{\mu\nu} \end{pmatrix},$$
$$g_{\mu\nu} \to g_{\mu\nu}, \quad D_{\mu\nu\rho\sigma} \to D_{\mu\nu\rho\sigma} \,, \tag{3.47}$$

[14]Note that this field λ has no relation to the field λ defined in section 3.2 for heterotic string theory on T^6, although both transform as modulus under the respective SL(2,Z) duality transformations in the two theories.

where p, q, r, s are real numbers satisfying,

$$ps - qr = 1\,. \tag{3.48}$$

The existence of this SL(2,R) symmetry in the type IIB supergravity theory led to the conjecture that an SL(2,Z) subgroup of this SL(2,R), obtained by restricting p, q, r, s to be integers instead of arbitrary real numbers, is a symmetry of the full string theory[30]. The breaking of SL(2,R) to SL(2,Z) can be seen as follows. An elementary string is known to carry $B_{\mu\nu}$ charge. In suitable normalization convention, it carries exactly one unit of $B_{\mu\nu}$ charge. This means that the $B_{\mu\nu}$ charge must be quantized in integer units, as the spectrum of string theory does not contain fractional strings carrying a fraction of the charge carried by the elementary string. From (3.47) we see that acting on an elementary string state carrying one unit of $B_{\mu\nu}$ charge, the SL(2,R) transformation gives a state with p units of $B_{\mu\nu}$ charge and r units of $B'_{\mu\nu}$ charge. Thus p must be an integer. It is easy to see that the maximal subgroup of SL(2,R) for which p is always an integer consists of matrices of the form

$$\begin{pmatrix} p & \alpha q \\ \alpha^{-1} r & s \end{pmatrix}\,, \tag{3.49}$$

with p, q, r, s integers satisfying $(ps - qr) = 1$, and α a fixed constant. Absorbing α into a redefinition of $B'_{\mu\nu}$ we see that the subgroup of SL(2,R) matrices consistent with charge quantization are the SL(2,Z) matrices $\begin{pmatrix} p & q \\ r & s \end{pmatrix}$ with p, q, r, s integers satisfying $ps - qr = 1$.

Note that this argument only shows that SL(2,Z) is the maximal possible subgroup of SL(2,R) that *can be a symmetry of the full string theory*, but does not prove that SL(2,Z) is a symmetry of string theory. In particular, since SL(2,Z) acts non-trivially on the dilaton, whose vacuum expectation value represents the string coupling constant, it cannot be verified order by order in string perturbation theory. We shall see later how one can find non-trivial evidence for this symmetry.

Besides this non-perturbative SL(2,Z) transformation, type IIB theory has two perturbatively verifiable discrete Z_2 symmetries. They are as follows:

- $(-1)^{F_L}$: It changes the sign of all the Ramond sector states on the left moving sector of the world-sheet. In particular, acting on the massless bosonic sector fields, it changes the sign of a, $B'_{\mu\nu}$ and $D_{\mu\nu\rho\sigma}$, but leaves $g_{\mu\nu}$, $B_{\mu\nu}$ and Φ invariant.

- Ω: This is the world-sheet parity transformation mentioned in section 1.1 that exchanges the left- and the right-moving sectors of the world-sheet. Acting on the massless bosonic sector fields, it changes the sign of $B_{\mu\nu}$, a and $D_{\mu\nu\rho\sigma}$, leaving the other fields invariant.

From this description, we see that the effect of $(-1)^{F_L} \cdot \Omega$ is to change of sign of $B_{\mu\nu}$ and $B'_{\mu\nu}$, leaving the other massless bosonic fields invariant. Comparing this with the action of the SL(2,Z) transformation laws of the massless bosonic sector fields, we see that $(-1)^{F_L} \cdot \Omega$ can be identified with the SL(2,Z) transformation:

$$\begin{pmatrix} -1 & \\ & -1 \end{pmatrix} . \tag{3.50}$$

This information will be useful to us later.

Theories obtained by modding out (compactified) type IIB string theory by a discrete symmetry group, where some of the elements of the group involve Ω, are known as orientifolds[113, 114]. The simplest example of an orientifold is type IIB string theory modded out by Ω. This corresponds to type I string theory. The closed string sector of type I theory consists of the Ω invariant states of type IIB string theory. The open string states of type I string theory are the analogs of twisted sector states in an orbifold, which must be added to the theory in order to maintain finiteness.

3.5 Other examples

Following the same procedure, namely, studying symmetries of the effective action together with charge quantization rules, we are led to many other duality conjectures in theories with 16 or more supersymmetry generators. Here we shall list the main series of such duality conjectures. We begin with the self duality groups of type II string theories compactified on tori of different dimensions. As mentioned earlier, there is a T-duality that relates type IIA on a circle to type IIB on a circle of inverse radius. Thus for $n \geq 1$, the self-duality groups of type IIA and type IIB theories compactified on an n-dimensional torus T^n will be identical. We now list the conjectured self-duality groups of type IIA/IIB string theory compactified on T^n for different values of n[30]:

$D = (10-n)$	Full Duality Group	T-duality Group
9	$SL(2,Z)$	$-$
8	$SL(2,Z) \times SL(3,Z)$	$SL(2,Z) \times SL(2,Z)$
7	$SL(5,Z)$	$SO(3,3;Z)$
6	$SO(5,5;Z)$	$SO(4,4;Z)$
5	$E_{6(6)}(Z)$	$SO(5,5;Z)$
4	$E_{7(7)}(Z)$	$SO(6,6;Z)$
3	$E_{8(8)}(Z)$	$SO(7,7;Z)$
2	$\widehat{E_{8(8)}}(Z)$	$SO(8,8;Z)$

Note that besides the full duality group, we have also displayed the T-duality group of each theory which can be verified order by order in string perturbation theory. $E_{n(n)}$ denotes a non-compact version of the exceptional group E_n for $n = 6,7,8$, and $E_{n(n)}(Z)$ denotes a discrete subgroup of $E_{n(n)}$. $\widehat{G}$ for any group G denotes the loop group of G based on the corresponding affine algebra and $\widehat{G}(Z)$ denotes a discrete subgroup of this loop group. Note that we have stopped at $D = 2$. We could in principle continue this all the way to $D = 1$ where all space-like directions are compactified. In this case one expects a very large duality symmetry group based on hyperbolic Lie algebra[116], which is not well understood to this date.

In each of the cases mentioned, the low energy effective field theory is invariant under the full continuous group[52], but charge quantization breaks this symmetry to its discrete subgroup. As noted before, these symmetries were discovered in the early days of supergravity theories, and were known as hidden non-compact symmetries.

Next we turn to the self-duality conjectures involving compactified heterotic string theories. Although there are two distinct heterotic string theories in ten dimensions, upon compactification on a circle, the two heterotic string theories can be shown to be related by a T-duality transformation. As a result, upon compactification on T^n, both of them will have the same self-duality group. We now display this self-duality group in various dimensions:

$D = (10-n)$	Full Duality Group	T-duality Group

9	$O(1,17,Z)$	$O(1,17;Z)$
8	$O(2,18,Z)$	$O(2,18;Z)$
7	$O(3,19,Z)$	$O(3,19;Z)$
6	$O(4,20,Z)$	$O(4,20;Z)$
5	$O(5,21,Z)$	$O(5,21;Z)$
4	$O(6,22,Z)\times SL(2,Z)$	$O(6,22;Z)$
3	$O(8,24,Z)$	$O(7,23;Z)$
2	$O(\widehat{8,24},Z)$	$O(8,24;Z)$

Since type I and SO(32) heterotic string theories are conjectured to be dual to each other in ten dimensions, the second column of the above table also represents the duality symmetry group of type I string theory on T^n. However, in the case of type I string theory, there is no perturbatively realised self-duality group (except trivial transformations which are part of the SO(32) gauge group and the group of global diffeomorphisms of T^n).

The effective action of type IIB string theory compactified on $K3$ has an $SO(5,21)$ symmetry[50], which leads to the conjecture that an SO(5,21;Z) subgroup of this is an exact self-duality symmetry of the type IIB string theory on K3. The conjectured duality between type IIA string theory compactified on $K3$ and heterotic string theory compactified on T^4 has already been discussed before. Due to the equivalence of type IIB on S^1 and type IIA on S^1, type IIA on $K3\times T^n$ is equivalent to type IIB on $K3\times T^n$. Finally, due to the conjectured duality between type IIA on K3 and heterotic on T^4, type IIA/IIB on $K3\times T^n$ are dual to heterotic string theory on T^{n+4} for $n\geq 1$. Thus the self-duality symmetry groups in these theories can be read out from the second column of the previous table displaying the self-duality groups of heterotic string theory on T^n.

Besides the theories discussed here, there are other theories with 16 or more supercharges obtained from non-geometric compactification of heterotic/type II string theories[53, 54, 55]. The duality symmetry groups of these theories can again be guessed from an analysis of the low energy effective field theory and the charge quantization conditions. Later we shall also describe a more systematic way of 'deriving' various duality conjectures from some basic set of dualities.

Although in this section I have focussed on duality symmetries of the low energy effective action which satisfy a non-renormalization theorem as a consequence of space-time supersymmetry, this is not the only part of the full effective action which satisfy such a non-renormalization theorem. Quite often the effective action contains another set of terms satisfying non-renormalization theorems. They are required for anomaly cancellation, and are known as Green-Schwarz terms. Adler-Bardeen theorem guarantees that they are not renormalized beyond one loop. These terms have also been used effectively for testing various duality conjectures[185], but I shall not discuss it in this article.

4 Precision Test of Duality: Spectrum of BPS States

Analysis of the low energy effective action, as discussed in the last section, provides us with only a crude test of duality. Its value lies in its simplicity. Indeed, most of the duality conjectures in string theory were arrived at by analysing the symmetries of the low energy effective action.

But once we have arrived at a duality conjecture based on the analysis of the low energy effective action, we can perform a much more precise test by analysing the spectrum of BPS states in the theories. BPS states are states which are invariant under part of the supersymmetry transformation, and are characterized by two important properties:

- They belong to a supermultiplet which has typically less dimension than a non-BPS state. This has an analog in the theory of representations of the Lorentz group, where massless states form a shorter representation of the algebra than massive states. Thus for example a photon has only two polarizations but a massive vector particle has three polarizations.

- The mass of a BPS state is completely determined by its charge as a consequence of the supersymmetry algebra. This relation between the mass and the charge is known as the BPS mass formula. This statement also has an analog in the theory of representations of the Lorentz algebra, *e.g.* a spin 1 representation of the Lorentz algebra containing only two states must be necessarily massless.

We shall now explain the origin of these two properties[36]. Suppose the theory has N real supersymmetry generators Q_α ($1 \leq \alpha \leq N$). Acting on a single particle state *at rest*, the supersymmetry algebra takes the form:

$$\{Q_\alpha, Q_\beta\} = f_{\alpha\beta}(m, \vec{Q}, \{y\}) \,, \tag{4.1}$$

where $f_{\alpha\beta}$ is a real symmetric matrix which is a function of its arguments m, $\vec{Q}$ and $\{y\}$. Here m denotes the rest mass of the particle, $\vec{Q}$ denotes various gauge charges carried by the particle, and $\{y\}$ denotes the coordinates labelling the moduli space of the theory.[15] We shall now consider the following distinct cases:

1. $f_{\alpha\beta}$ has no zero eigenvalue. In this case by taking appropriate linear combinations of Q_α we can diagonalize f. By a further appropriate rescaling of Q_α, we can bring f into the identity matrix. Thus in this basis the supersymmetry algebra has the form:

$$\{Q_\alpha, Q_\beta\} = \delta_{\alpha\beta}. \tag{4.2}$$

   This is the N dimensional Clifford algebra. Thus the single particle states under consideration form a representation of this Clifford algebra, which is $2^{N/2}$ dimensional. (We are considering the case where N is even.) Such states would correspond to non-BPS states.

2. f has $(N - M)$ zero eigenvalues for some $M < N$. In this case, by taking linear combinations of the Q_α we can bring the algebra into the form:

$$\begin{aligned} \{Q_\alpha, Q_\beta\} &= \delta_{\alpha\beta}, \quad \text{for} \quad 1 \le \alpha, \beta \le M\,, \\ &= 0 \quad \text{for} \quad \alpha \text{ or } \beta > M\,. \end{aligned} \tag{4.3}$$

   We can form an irreducible representation of this algebra by taking all states to be annihilated by Q_α for $\alpha > M$. In that case the states will form a representation of an M dimensional Clifford algebra generated by Q_α for $1 \le \alpha \le M$. This representation is $2^{M/2}$ dimensional for M even. Since $M < N$, we see that these are lower dimensional representations compared to that of a generic non-BPS state. Furthermore, these states are invariant under part of the supersymmetry algebra generated by Q_α for $\alpha > M$. These are known as BPS states. We can get different kinds of BPS states depending on the value of M, *i.e.* depending on the number of supersymmetry generators that leave the state invariant.

From this discussion it is clear that in order to get a BPS state, the matrix f must have some zero eigenvalues. This in turn, gives a constraint involving mass m, charges $\vec{Q}$ and the moduli $\{y\}$, and is the origin of the BPS formula relating the mass and the charge of the particle.

[15] Only specific combinations of $\vec{Q}$ and $\{y\}$, known as central charges, appear in the algebra.

Before we proceed, let us illustrate the preceeding discussion in the context of a string theory. Consider Type IIB string theory compactified on a circle S^1. The total number of supersymmetry generators in this theory is 32. Thus a generic non-BPS supermultiplet is $2^{16} = (256)^2$ dimensional. These are known as long multiplets. This theory also has BPS states breaking half the space-time supersymmetry. For these states $M = 16$ and hence we have $2^8 = 256$ dimensional representation of the supersymmetry algebra. These states are known as ultra-short multiplets. We can also have BPS states breaking 3/4 of the space-time supersymmetry ($M = 24$). These will form a $2^{12} = 256 \times 16$ dimensional representation, and are known as short multiplets. In each case there is a specific relation between the mass and the various charges carried by the state. We shall discuss this relation as well as the origin of these BPS states in more detail later.

As another example, consider heterotic string theory compactified on an n-dimensional torus T^n. The original theory has 16 supercharges. Thus a generic non-BPS state will belong to a $2^8 = 256$ dimensional representation of the supersymmetry algebra. But if we consider states that are invariant under half of the supercharges, then they belong to a $2^4 = 16$ dimensional representation of the supersymmetry algebra. This is known as the short representation of this superalgebra. We can also have states that break 3/4 of the supersymmetries.[16] These belong to a 64 dimensional representation of the supersymmetry algebra known as intermediate states.

BPS states are further characterized by the property that the degeneracy of BPS states with a given set of charge quantum numbers is independent of the value of the moduli fields $\{y\}$. Since string coupling is also one of the moduli of the theory, this implies that the degeneracy at any value of the string coupling is the same as that at weak coupling. This is the key property of the BPS states that makes them so useful in testing duality, so let us review the argument leading to this property[36]. We shall discuss this in the context of the specific example of type IIB string theory compactified on S^1, but it can be applied to any other theory. Suppose the theory has an ultra-short multiplet at some point in the moduli space. Now let us change the moduli. The question that we shall be asking is: can the ultra-short multiplet become a long (or any other) multiplet as we change the moduli? If we assume that the total number of states does not change discontinuously, then this is clearly not possible since other multiplets have different number of states. Thus as long

[16]It turns out that these states can exist only for $n \geq 5$. This constraint arises due to the fact that the unbroken supersymmetry generators must form a representation of the little group $SO(9-n)$ of a massive particle in $(10-n)$ dimensional space-time.

as the spectrum varies smoothly with the moduli (which we shall assume), an ultra-short multiplet stays ultra-short as we move in the moduli space[95]. Furthermore, as long as it stays ultra-short, its mass is determined by the BPS formula. Thus we see that the degeneracy of ultra-short multiplets cannot change as we change the moduli of the theory. A similar argument can be given for other multiplets as well. Note that for this argument to be strictly valid, we require that the mass of the BPS state should stay away from the continuum, since otherwise the counting of states is not a well defined procedure. This requires that the mass of a BPS state should be strictly less than the total mass of any set of two or more particles carrying the same total charge as the BPS state.

Given this result, we can now adapt the following strategy to carry out tests of various duality conjectures using the spectrum of BPS states in the theory:

1. Identify BPS states in the spectrum of elementary string states. The spectrum of these BPS states can be trusted at all values of the coupling even though it is calculated at weak coupling.

2. Make a conjectured duality transformation. This typically takes a BPS state in the spectrum of elementary string states to another BPS state, but with quantum numbers that are not present in the spectrum of elementary string states. Thus these states must arise as solitons /composite states.

3. Try to explicitly verify the existence of these solitonic states with degeneracy as predicted by duality. This will provide a non-trivial test of the corresponding duality conjecture.

We shall now illustrate this procedure with the help of specific examples. We shall mainly follow [58, 72, 69].

4.1 SL(2,Z) S-duality in heterotic on T^6 and multi-monopole moduli spaces

As discussed in section 3.2, heterotic string theory compactified on T^6 is conjectured to have an SL(2,Z) duality symmetry. In this subsection we shall see how one can test this conjecture by examining the spectrum of BPS states.

Since the BPS spectrum does not change as we change the moduli, we can analyse the spectrum near some particular point in the moduli space. As discussed in section 3.2, at a generic point in the moduli space the unbroken gauge group is $U(1)^{28}$. But there are special points in this moduli space where we get enhanced non-abelian gauge group[120]. Thus for example, if we set the internal components of the original ten dimensional gauge fields to zero, we get unbroken $E_8 \times E_8$ or SO(32) gauge symmetry. Let us consider a special point in the moduli space where an SU(2) gauge symmetry is restored. This can be done for example by taking a particular S^1 in T^6 to be orthogional to all other circles, taking the components of the gauge fields along this S^1 to be zero, and taking the radius of this S^1 to be the self-dual radius. In that case the effective field theory at energies much below the string scale will be described by an N=4 supersymmetric SU(2) gauge theory, together with a set of decoupled N=4 supersymmetric U(1) gauge theories and N=4 supergravity. The conjectured SL(2,Z) duality of the heterotic string theory will require the N=4 supersymmetric SU(2) gauge theory to have this SL(2,Z) symmetry.[17] Thus by testing the duality invariance of the spectrum of this N=4 supersymmetric SU(2) gauge theory we can test the conjectured SL(2,Z) symmetry of heterotic string theory.

The N=4 supersymmetric SU(2) gauge theory has a vector, six massless scalars and four massless Majorana fermions in the adjoint representation of SU(2)[57]. The form of the lagrangian is fixed completely by the requirement of $N = 4$ supersymmetry up to two independent parameters − the coupling constant g that determines the strength of all interactions (gauge, Yukawa, scalar self-interaction etc.), and the vacuum angle θ that multiplies the topological term $Tr(F\widetilde{F})$ involving the gauge field. With the choice of suitable normalization convention, g and θ are related to the vacuum expectation value of the field λ defined in (3.18) through the relation:

$$\langle\lambda\rangle = \frac{\theta}{2\pi} + i\frac{4\pi}{g^2}\,. \tag{4.4}$$

The potential involving the six adjoint representation scalar fields ϕ^α_m $(1 \le \alpha \le 3, 1 \le m \le 6)$ is proportional to

$$\sum_{m<n}\sum_{\alpha}(\epsilon^{\alpha\beta\gamma}\phi^\beta_m\phi^\gamma_n)^2\,. \tag{4.5}$$

This vanishes for

$$\phi^\alpha_m = a_m\delta_{\alpha 3}\,. \tag{4.6}$$

Vacuum expectation values of ϕ^α_m of the form (4.6) does not break supersymmetry, but breaks the gauge group $SU(2)$ to $U(1)$. The parameters $\{a_m\}$ correspond to the vacuum

[17] Independently of string theory, the existence of a strong-weak coupling duality in this theory was conjectured earlier[56, 57].

expectation values of a subset of the scalar moduli fields M in the full string theory. We shall work in a region in the moduli space where $a_m \neq 0$ for some m, but the scale of breaking of SU(2) is small compared to the string scale ($|a_m| << (\sqrt{\alpha'})^{-1}$ for all m), so that gravity is still decoupled from this gauge theory. The BPS states in the spectrum of elementary particles in this theory are the heavy charged bosons $W^{\pm}$ and their superpartners. These break half of the 16 space-time supersymmetry generators and hence form a $2^{8/2} = 16$ dimensional representation of the supersymmetry algebra. These states can be found explicitly in the spectrum of elementary string states from the sector containing strings with one unit of winding and one unit of momentum along the special S^1 that is responsible for the enhanced SU(2) gauge symmetry. As we approach the point in the moduli space where this special S^1 has self-dual radius, these states become massless and form part of the SU(2) gauge multiplet.

When SU(2) is broken to U(1) by the vacuum expectation value of ϕ_m, the spectrum of solitons in this theory is characterized by two quantum numbers, the electric charge quantum number n_e and the magnetic charge quantum number n_m, normalized so that n_e and n_m are both integers. We shall denote such a state by $\begin{pmatrix} n_e \\ n_m \end{pmatrix}$. In this notation the elementary W^+ boson corresponds to a $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ state. By studying the action of the SL(2,Z) transformation (3.26) on the gauge fields, we can easily work out its action on the charge quantum numbers $\begin{pmatrix} n_e \\ n_m \end{pmatrix}$[1]. The answer is

$$\begin{pmatrix} n_e \\ n_m \end{pmatrix} \to \begin{pmatrix} p & q \\ r & s \end{pmatrix} \begin{pmatrix} n_e \\ n_m \end{pmatrix} , \tag{4.7}$$

for appropriate choice of sign convention for n_e and n_m. Thus acting on an $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ state it produces a $\begin{pmatrix} p \\ r \end{pmatrix}$ state. From the relation $ps - qr = 1$ satisfied by an SL(2,Z) matrix, we can easily see that p and r are relatively prime. Furthermore for every p and r relatively prime, we can find integers q and s satisfying $ps - qr = 1$. Thus SL(2,Z) duality predicts that *for every p and r relatively prime, the theory must contain a unique short multiplet with charge quantum numbers* $\begin{pmatrix} p \\ r \end{pmatrix}$ *[58].*

We can now directly examine the solitonic sector of the theory to check this prediction. The theory contains classical monopole solutions which break half of the supersymmetries of the original theory. These solutions are non-singular everywhere, and in fact, for a given r, there is a $4r$ parameter non-singular solution with r units of total magnetic charge[117, 59]. These $4r$ parameters correspond to the bosonic collective excitations of this system[118]. In order to study the spectrum of BPS solitons, we need to quantize these collective excitations and look for supersymmetric ground states of the corresponding quantum mechanical system. Each solution also has infinite number of vibrational modes with non-zero frequency, but excitations of these modes are not relevant for finding supersymmetric ground states.

States with $r = 1$ come from one monopole solution. This has four bosonic collective coordinates, three of which correspond to the physical position of the monopole in the three dimensional space, and the fourth one is an angular variable describing the U(1) phase of the monopole. The momenta conjugate to the first three coordinates correspond to the components of the physical momentum of the particle. These can be set to zero by working in the rest frame of the monopole. The fourth coordinate is periodically identified and hence its conjugate momentum is quantized in integer units. This integer p corresponds to the electric charge quantum number n_e. Thus the states obtained by quantizing the bosonic sector of the theory has charge quantum numbers $\begin{pmatrix} p \\ 1 \end{pmatrix}$ for all integer p.

The degeneracy comes from quantizing the fermionic sector. There are eight fermionic zero modes, which describe the result of applying the eight broken supersymmetry generators on the monopole solution. These form an eight dimensional Clifford algebra. Thus the ground state has $2^4 = 16$-fold degeneracy, exactly as predicted by SL(2,Z)[57].

Let us now turn to the analysis of states with $r > 1$[58]. As has already been said, this system has $4r$ bosonic collective coordinates, which, when the monopoles are far away from each other, correspond to the spatial location and the U(1) phase of each of the r monopoles. The total number of fermionic collective coordinates can be computed from an index theorem and is equal to $8r$[119]. We can divide this set into the 'center of mass' coordinates containing four bosonic and eight fermionic coordinates, and the 'relative coordinates' containing $4(r-1)$ bosonic and $8(r-1)$ fermionic coordinates. The quantization of the center of mass system gives states carrying charge quantum numbers $\begin{pmatrix} p \\ r \end{pmatrix}$ with 16-fold degeneracy, p being the momentum conjugate to the overall U(1) phase. This shows that the degeneracy is always a multiple of 16, consistent with the fact that a short multiplet is 16-fold degenerate. At this stage p can be any integer, not necessarily prime relative to r. However, since the total wave-function is a product of the wave-function of the center of mass system and the relative system, in order to determine the number of short multiplets for a given value of p, we need to turn to the quantum mechanics of the relative coordinates.

It turns out that the bosonic coordinates in the relative coordinate system describe a non-trivial $4(r-1)$ dimensional manifold, known as the relative moduli space of r monopoles[118, 59, 60]. The quantum mechanics of the bosonic and fermionic relative coordinates can be regarded as that of a supersymmetric particle moving in this moduli space. There are several subtleties with this system. They are listed below:

- First of all, the center of mass and the relative coordinates do not completely decouple, although they decouple locally. The full moduli space has the structure[59]:

$$(R^3 \times S^1 \times \mathcal{M}_r)/Z_r\,, \tag{4.8}$$

  where R^3 is parametrized by the center of mass location, S^1 by the overall U(1) phase, and $\mathcal{M}_r$ by the relative coordinates. There is an identification of points in the product space $R^3 \times S^1 \times \mathcal{M}_r$ by a Z_r transformation that acts as a shift by $2\pi/r$ on S^1 and as a diffeomorphism on $\mathcal{M}_r$ without any fixed point[59, 60]. Due to this identification, the total wave-function must be invariant under this Z_r transformation. Since the part of the wave-function involving the coordinate of S^1 picks up a phase $\exp(2\pi ip/r)$ under this Z_r, we see that the wave-function involving the relative coordinates must pick up a phase of $\exp(-2\pi ip/r)$ under this Z_r transformation.

- Normally the part of the wave-function involving the relative coordinates will be a function on $\mathcal{M}_r$. But it turns out that the effect of the $8(r-1)$ fermionic degrees of freedom in the quantum mechanical system makes the wave-function a differential form of arbitrary rank on $\mathcal{M}_r$[61, 62].

- Finally, among all the possible states, the ones saturating Bogomol'nyi bound correspond to harmonic differential forms on $\mathcal{M}_r$. This can be understood as follows. It can be shown that the Hamiltonian of the relative coordinates correspond to the Laplacian on $\mathcal{M}$. Also it turns out that the BPS mass formula is saturated by contribution from the center of mass coordinates. Hence in order to get a BPS state, the part of the wave-function involving the relative coordinates must be an eigenstate of the corresponding Hamiltonian with zero eigenvalue *i.e.* it must be a harmonic form on $\mathcal{M}_r$. Thus for every harmonic differential form we get a short multiplet, since the fermionic degrees of freedom associated with the center of mass coordinates supply the necessary 16-fold degeneracy.

Thus the existence of a short multiplet of charge quantum numbers $\begin{pmatrix} p \\ r \end{pmatrix}$ would require the existence of a harmonic form on $\mathcal{M}_r$ that picks up a phase of $\exp(2\pi i p/r)$ under the action of Z_r. According to the prediction of SL(2,Z) *such a harmonic form should exist only for p and r relatively prime, and not for other values of p[58].*

For $r = 2$ the relevant harmonic form can be constructed explicitly[63, 64, 58], therby verifying the existence of the states predicted by SL(2,Z) duality. For $r > 2$ the analysis is more complicated since the metric in the multimonopole moduli space is not known. However general arguments showing the existence of the necessary harmonic forms has been given[65, 66].

Besides the BPS states discussed here, the spectrum of elementary string states in the heterotic string theory on T^6 contains many other BPS states. In the world-sheet theory, a generic state is created by applying oscillators from the left- and the right-moving sector on the Fock vacuum. The Fock vacuum, in turn, is characterized by a pair of vectors $(\vec{k}_L, \vec{k}_R)$ specifying the charges (momenta) associated with the six right-handed and twenty two left-handed currents on the world-sheet. From the viewpoint of the space-time theory, these 28 components of $(\vec{k}_L, \vec{k}_R)$ are just appropriate linear combinations of the charges carried by the state under the 28 U(1) gauge fields. The tree level mass formula for an elementary string state in the NS sector is given by,[18]

$$m^2 = \frac{4}{\lambda_2}\Big[\frac{\vec{k}_R^2}{2} + N_R - \frac{1}{2}\Big] = \frac{4}{\lambda_2}\Big[\frac{\vec{k}_L^2}{2} + N_L - 1\Big]\,, \tag{4.9}$$

where N_R and N_L denote respectively the oscillator levels of the state in the right- and the left-moving sectors of the world-sheet. In the above equation the terms in the square bracket denote the total contribution to L_0 and $\bar{L}_0$ from the oscillators, the internal momenta, and the vacua in the right- and the left-moving sectors respectively. Normally we do not have the factor of λ_2^{-1} in the mass formula since the formula refers to the ADM mass measured in the string metric $G_{\mu\nu} = \lambda_2^{-1} g_{\mu\nu}$. But here (and in the rest of the article) we quote the ADM mass measured in the canonical metric $g_{\mu\nu}$. This is more convenient for discussing duality invariance of the spectrum, since it is $g_{\mu\nu}$ and not

[18]In this and all subsequent mass formula λ_2 should really be interpreted as the vacuum expectation value of λ_2.

$G_{\mu\nu}$ that remains invariant under a duality transformation. The additive factor of $-1/2$ and -1 can be interpreted as the contributions to L_0 and $\bar{L}_0$ from the vacuum. (In the covariant formulation these can be traced to the contributions from the world-sheet ghost fields).

It turns out that of the full set of elementary string states, only those states which satisfy the constraint[71]

$$N_R = \frac{1}{2}\,, \tag{4.10}$$

correspond to BPS states (short multiplets). From eqs.(4.9) we see that for these states

$$N_L = \frac{1}{2}(\vec{k}_R^2 - \vec{k}_L^2) + 1\,. \tag{4.11}$$

The degeneracy $d(N_L)$ of short multiplets for a given set of $\vec{k}_L$, $\vec{k}_R$ is determined by the number of ways a level N_L state can be created out of the Fock vacuum by the 24 left-moving bosonic oscillators (in the light-cone gauge) − 8 from the transverse bosonic coordinates of the string and 16 from the bosonization of the 32 left-moving fermions on the world-sheet − and is given by the formula:

$$\sum_{N_L=0}^{\infty} d(N_L) q^{N_L} = \prod_{n=1}^{\infty} \frac{1}{(1-q^n)^{24}}\,. \tag{4.12}$$

The BPS states discussed earlier − the ones which can be regarded as the massive gauge bosons of a spontaneously broken non-abelian gauge theory − correspond to the $N_L = 0$ states in this classification. From eq.(4.12) we see that we have only one short multiplet for states with this quantum number; this is consistent with their description as heavy gauge bosons in an N=4 supersymmetric gauge theory. The next interesting class of states are the ones with $N_L = 1$. From (4.12) we see that they have degeneracy 24.[19] An SL(2,Z) transformation relates these states to appropriate magnetically charged states with r units of magnetic charge and p units of electric charge for p and r relatively prime. Thus the SL(2,Z) self-duality symmetry of the heterotic string theory predicts the existence of 24-fold degenerate solitonic states with these charge quantum numbers.

Verifying the existence of these solitonic states turns out to be quite difficult[67]. The main problem is that unlike the $N_L = 0$ states, the solitonic states (known as H-monopoles) which are related to the $N_L = 1$ states by SL(2,Z) duality turn out to be

[19] In counting degeneracy we are only counting the number of short multiplets, and ignoring the trivial factor of 16 that represents the degeneracy within each short multiplet.

singular objects, and hence we cannot unambiguously determine the dynamics of collective coordinates of these solitons just from the low energy effective field theory. Nevertheless, the problem has now been solved for $r = 1$[79, 80, 77, 81], and one finds that these solitons have exactly the correct degeneracy 24.

Similar analysis based on soliton solutions of low energy supergravity theory has been used to test many other duality conjectures[30, 32, 33, 27, 28, 68]. One of the main problems with this approach has been that unlike the example discussed in this section, most of these other solutions are either singular, or has strong curvature at the core where the low energy approximation breaks down. As a result, analysis based on these solutions has been of limited use. The situation changed after the advent of D-branes, to which we now turn.

4.2 SL(2,Z) duality in type IIB on S^1 and D-branes

As discussed earlier, type IIB string theory in ten dimensions has a conjectured SL(2,Z) duality symmetry group. In this section I shall discuss the consequence of this conjectured symmetry for the spectrum of BPS states in type IIB string theory compactified on a circle S^1. For details, see [69, 70].

The spectrum of elementary string states in this theory are characterized by two charges k_L and k_R defined as:

$$k_L = (k\lambda_2^{1/4}/R - wR/\lambda_2^{1/4})/\sqrt{2}, \qquad k_R = (k\lambda_2^{1/4}/R + wR/\lambda_2^{1/4})/\sqrt{2}\,, \tag{4.13}$$

where R denotes the radius of S^1 measured in the ten dimensional canonical metric, k/R denotes the momentum along S^1 with k being an integer, and w, also an integer, denotes the number of times the elementary string is wound along S^1. As usual we have set $\alpha' = 1$. In the world-sheet theory describing first quantized string theory, k_L and k_R denote the left and the right-moving momenta respectively. There are infinite tower of states with this quantum number, obtained by applied appropriate oscillators, both from the left- and the right-moving sector of the world-sheet, on the Fock vacuum of the world-sheet theory carrying these quantum numbers. The mass formula for any state in this tower, measured in the ten dimensional canonical metric, is given by:

$$m^2 = \frac{2}{\sqrt{\lambda_2}}(k_L^2 + 2N_L) = \frac{2}{\sqrt{\lambda_2}}(k_R^2 + 2N_R)\,, \tag{4.14}$$

where N_L, N_R denote oscillator levels on the left- and the right- moving sector of the world-sheet respectively.[20] In normal convention, one does not have the factors of (λ_2) in the mass formula, but here it comes due to the fact that we are using the *ten dimensional* canonical metric instead of the string metric to define the mass of a state. (Note that if we had used the nine dimensional canonical metric as defined in eqs.(3.1), (3.2), there will be an additional multiplicative factor of $R^{-2/9}$ in the expression for m^2.)

Most of these states are not BPS states as they are not invariant under any part of the supersymmetry transformation. It turns out that in order to be invariant under half of the space-time supersymmetry coming from the left- (right-) moving sector of the world-sheet, N_L (N_R) must vanish[71]. Thus a state with $N_L = N_R = 0$ will preserve half of the total number of supersymmetries and will correspond to ultra-short multiplets. From eq.(4.14) we see that mass formula for these states takes the form:

$$m^2 = \frac{2k_L^2}{\sqrt{\lambda_2}} = \frac{2k_R^2}{\sqrt{\lambda_2}} \,. \tag{4.15}$$

This is the BPS mass formula for these ultra-short multiplets. This requires $k_L = \pm k_R$ or, equivalently, $k = 0$ or $w = 0$. On the other hand, a state with either $N_L = 0$ or $N_R = 0$ will break (3/4)th of the total number of supersymmetries in the theory, and will correspond to short multiplets. If, for definiteness, we consider states with $N_R = 0$, then the BPS mass formula takes the form:

$$m^2 = \frac{2k_R^2}{\sqrt{\lambda_2}} \,. \tag{4.16}$$

N_L is determined in terms of k_L and k_R through the relation:

$$N_L = \frac{1}{2}(k_R^2 - k_L^2) = wk \,. \tag{4.17}$$

There is no further constraint on w and k. Although we have derived these mass formulae by directly analysing the spectrum of elementary string states, they can also be derived by analyzing the supersymmetry algebra, as indicated earlier.

One can easily calculate the degeneracy of these states by analyzing the spectrum of elementary string states in detail. For example, for the states with $N_L = N_R = 0$, there is a 16-fold degeneracy of states in each (left- and right-) sector of the world-sheet, − 8

[20] We have stated the formula in the RR sector, but due to space-time supersymmetry we get identical spectrum from the NS and the R sectors.

from the NS sector and 8 from the R sector. Thus the net degeneracy of such a state is $16 \times 16 = 256$, showing that there is a unique ultra-short multiplet carrying given charges (k_L, k_R). The degeneracy of short multiplets can be found in a similar manner. Consider for example states with $N_R = 0$, $N_L = 1$. In this case there is a 16-fold degeneracy coming from the right-moving sector of the world-sheet. There is an 8-fold degeneracy from the Ramond sector Fock vacuum of the left-moving sector. There is also an extra degeneracy factor in the left-moving Ramond sector due to the fact that there are many oscillators that can act on the Fock vacuum of the world-sheet theory to give a state at oscillator level $N_L = 1$. For example we get eight states by acting with the transverse bosonic oscillators α^i_{-1} $(1 \le i \le 8)$, and eight states by acting with the transverse fermionic oscillators ψ^i_{-1}.[21] This gives total degeneracy factor of 8×16 in the left-moving Ramond sector. Due to supersymmetry, we get an identical factor from the left-moving NS sector as well. Thus we get a state with total degeneracy $16 \times 16 \times 16$, $-$ 16 from the right moving sector, and 16×16 from the left-moving sector $-$ which is the correct degeneracy of a single short multiplet. Similar counting can be done for higher values of N_L as well. It turns out that the total number of short multiplets $d(N_L)$ with $N_R = 0$ for some given value of $N_L \ge 1$ is given by the formula:

$$\sum_{N_L} d(N_L) q^{N_L} = \frac{1}{16} \prod_{n=1}^{\infty} \Big(\frac{1+q^n}{1-q^n} \Big)^8 . \tag{4.18}$$

The $(1+q^n)^8$ and $(1-q^n)^8$ factors in the numerator and the denominator are related respectively to the fact that in the light-cone gauge there are 8 left-moving fermionic fields and 8 left-moving bosonic fields on the world-sheet. The overall factor of (1/16) is due to the fact that the lowest level state is only 256-fold degenerate but a single short multiplet requires 16×256 states.

Let us first consider the ultra-short multiplet with $k = 0$, $w = 1$. These states have mass

$$m^2 = \frac{R^2}{\lambda_2} . \tag{4.19}$$

It is well known that an elementary string acts as a source of the $B_{\mu\nu}$ field (see *e.g.* ref.[71]). Thus in the (8+1) dimensional theory obtained by compactifying type IIB on

[21] Since ψ^i_{-1} has fermion number one, it has to act on the Fock vacua with odd fermion number in order that the states obtained after acting with ψ^i_{-1} on the vacua satisfy GSO projection.

S^1, the $w=1$ state will carry one unit of $B_{9\mu}$ gauge field charge. Now, under SL(2,Z)

$$\begin{pmatrix} B_{9\mu} \\ B'_{9\mu} \end{pmatrix} \to \begin{pmatrix} p & q \\ r & s \end{pmatrix} \begin{pmatrix} B_{9\mu} \\ B'_{9\mu} \end{pmatrix} . \tag{4.20}$$

This converts the $w=1$ state, which we shall denote by $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ reflecting the $\begin{pmatrix} B_{9\mu} \\ B'_{9\mu} \end{pmatrix}$ charge carried by the state, to a $\begin{pmatrix} p \\ r \end{pmatrix}$ state, *i.e.* a state carrying p units of $B_{9\mu}$ charge and r units of $B'_{9\mu}$ charge. The condition $ps - qr = 1$ implies that the pair of integers (p, r) are relatively prime. Thus SL(2,Z) duality of type IIB string theory predicts that $\forall (p, r)$ relatively prime, the theory must have a unique ultra-short multiplet with p units of $B_{9\mu}$ charge and r units of $B'_{9\mu}$ charge[68]. The BPS mass formula for these states can be derived by analysing the supersymmetry algebra, as indicated earlier, and is given by,

$$m^2 = \frac{R^2}{\lambda_2} |r\lambda - p|^2 . \tag{4.21}$$

Note that this formula is invariant under the SL(2,Z) transformation:

$$\lambda \to \frac{a\lambda + b}{c\lambda + d}, \qquad \begin{pmatrix} p \\ r \end{pmatrix} \to \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} p \\ r \end{pmatrix} , \tag{4.22}$$

where $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is an SL(2,Z) matrix.

A similar prediction for the spectrum of BPS states can be made for short multiplets as well. In this case the state is characterized by three integers p, r and k reflecting the $B_{9\mu}$, $B'_{9\mu}$ and $G_{9\mu}$ charge (momentum along S^1) respectively. Let us denote by $d(k, p, r)$ the degeneracy of such short multiplets. For (p, r) relatively prime, an SL(2,Z) transformation relates these to elementary string states with one unit of winding and k units of momentum along S^1. Such states have degeneracy $d(k)$ given in eq.(4.18). Then by following the same logic as before, we see that the SL(2,Z) duality predicts that for (p, r) relatively prime, $d(k, p, r)$ is independent of p and r and depends on k according to the relation:

$$\sum_k d(k, p, r) q^k = \frac{1}{16} \prod_{n=1}^{\infty} \Big(\frac{1+q^n}{1-q^n}\Big)^8 . \tag{4.23}$$

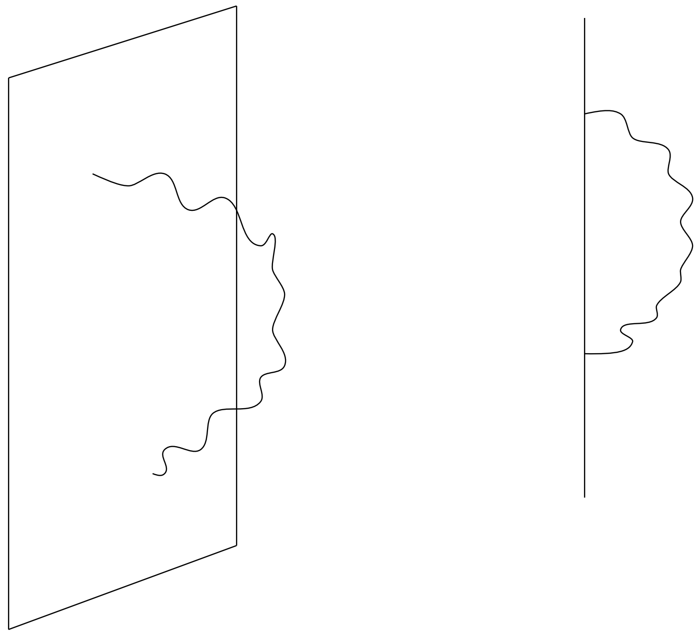
In other words, there should be a Hagedorn spectrum of short multiplets with charge $\begin{pmatrix} p \\ r \end{pmatrix}$.

A test of SL(2,Z) symmetry involves explicitly verifying the existence of these states. To see what such a test involves, recall that $B'_{\mu\nu}$ arises in the RR sector of string theory.

In type II theory, all elementary string states are neutral under RR gauge fields as can be seen by computing a three point function involving any two elementary string states and an RR sector gauge field. Thus a state carrying $B'_{9\mu}$ charge must arise as a soliton. The naive approach will involve constructing such a soliton solution as a solution to the low energy supergravity equations of motion, quantizing its zero modes, and seeing if we recover the correct spectrum of BPS states. However, in actual practice, when one constructs the solution carrying $B'_{\mu\nu}$ charge, it turns out to be singular. Due to this fact it is difficult to proceed further along this line, as identifying the zero modes of a singular solution is not a well defined procedure. In particular we need to determine what boundary condition the modes must satisfy at the singularity. Fortunately, in this theory, there is a novel way of constructing a soliton solution that avoids this problem. This construction uses Dirichlet (D-) branes[72, 73]. In order to compute the degeneracy of these solitonic states, we must understand the definition and some of the the properties of these D-branes. This is the subject to which we now turn.

Normally type IIA/IIB string theory contains closed string states only. But we can postulate existence of solitonic extended objects in these theories such that in the presence of these solitons, there can be open string states whose ends lie on these extended objects (see Fig.10). This can in fact be taken to be the defining relation for these solitons, with the open string states with ends lying on the soliton corresponding to the (infinite number of) vibrational modes of the soliton. Of course, one needs to ensure that the soliton defined this way satisfy all the properties expected of a soliton solution in this theory *e.g.* partially unbroken supersymmetry, existence of static multi-soliton solutions etc. Since open strings satisfy Dirichlet boundary condition in directions transverse to these solitons, these solitons are called D-branes. In particular, we shall call a D-brane with Neumann boundary condition in $(p+1)$ directions (including time) and Dirichlet boundary condition in $(9-p)$ directions a Dirichlet p-brane, since it can be regarded as a soliton extending along p space-like directions in which we have put Neumann boundary condition. (Thus a 0-brane represents a particle like object, a 1-brane a string like object, and a 2-brane a membrane like object.) To be more explicit, let us consider the following boundary condition on the open string:

$$\begin{aligned} X^m(\sigma=0,\pi) &= x_0^m \quad \text{for} \quad (p+1)\le m\le 9\,, \\ \partial_\sigma X^\mu(\sigma=0,\pi) &= 0 \quad \text{for} \quad 0\le\mu\le p\,, \end{aligned} \tag{4.24}$$



(a) (b)

Figure 10: Open string states with ends attached to a (a) Dirichlet membrane, (b) Dirichlet string.

where σ denotes the spatial direction on the string world-sheet. The boundary conditions on the world-sheet fermion fields are determined from (4.24) using various consistency requirements including world-sheet supersymmetry that relates the world-sheet bosons and fermions. Note that these boundary conditions break translational invariance along x^m. Since we want the full theory to be translationally invariant, the only possible interpretation of such a boundary condition is that there is a p dimensional extended object situated at $x^m = x_0^m$ that is responsible for breaking this translational invariance. We call this a Dirichlet p-brane located at $x^m = x_0^m$ $(p+1 \le m \le 9)$, and extended along $x^1, \ldots x^p$.

Let us now summarize some of the important properties of D-branes that will be relevant for understanding the test of SL(2,Z) duality in type IIB string theory:

- The Dirichlet p-brane in IIB is invariant under half of the space-time supersymmetry transformations for odd p. To see how this property arises, let us denote by ϵ_L and ϵ_R the space-time supersymmetry transformation parameters in type IIB string theory, originating in the left- and the right-moving sector of the world-sheet theory respectively. ϵ_L and ϵ_R satisfy the chirality constraint:

$$\Gamma^0 \cdots \Gamma^9 \epsilon_L = \epsilon_L, \qquad \Gamma^0 \cdots \Gamma^9 \epsilon_R = \epsilon_R \,, \tag{4.25}$$

  where Γ^μ are the ten dimensional gamma matrices. The open string boundary conditions (4.24) together with the corresponding boundary conditions on the world-sheet fermions give further restriction on ϵ_L and ϵ_R of the form[72]:

$$\epsilon_L = \Gamma^{p+1} \dots \Gamma^9 \epsilon_R \,. \tag{4.26}$$

  It is easy to see that the two equations (4.25) and (4.26) are compatible only for odd p. Thus in type IIB string theory Dirichlet p-branes are invariant under half of the space-time supersymmetry transformations for odd p. An identical argument shows that in type IIA string theory we have supersymmetric Dirichlet p-branes only for even p since in this theory eq.(4.25) is replaced by,

$$\Gamma^0 \cdots \Gamma^9 \epsilon_L = \epsilon_L, \qquad \Gamma^0 \cdots \Gamma^9 \epsilon_R = -\epsilon_R \,. \tag{4.27}$$

- Type IIB (IIA) string theory contains a p-form gauge field for even (odd) p. For example, in type IIB string theory these p-form gauge fields correspond to the scalar a, the rank two anti-symmetric tensor field $B'_{\mu\nu}$ and the rank four anti-symmetric tensor field $D_{\mu\nu\rho\sigma}$. It can be shown that a Dirichlet p-brane carries one unit of charge under the RR $(p+1)$-form gauge field[72]. More precisely, if we denote by $C_{\mu_1 \cdots \mu_q}$ the q-form gauge potential, then a Dirichlet p-brane extending along $1 \cdots p$ direction acts as a source of $C_{01 \cdots p}$. (For $p = 5$ and 7 these correspond to magnetic dual potentials of $B'_{\mu\nu}$ and a respectively.) This result can be obtained by computing the one point function of the vertex operator for the field C in the presence of a D-brane. The relevant string world-sheet diagram has been indicated in Fig.11. We shall not discuss the details of this computation here.

From this discussion it follows that a Dirichlet 1-brane (D-string) in type IIB theory carries one unit of charge under the RR 2-form field $B'_{\mu\nu}$. This means that in type IIB

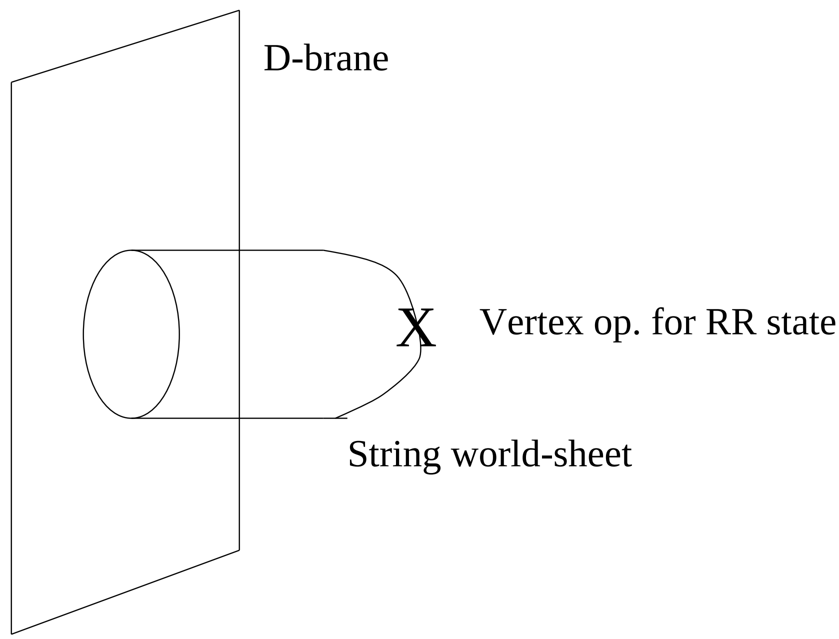


Figure 11: The string world-sheet diagram relevant for computing the coupling of the RR gauge field to the D-brane. It corresponds to a surface of the topology of a hemisphere with its boundary glued to the D-brane. The vertex operator of the RR-field is inserted at a point on the hemisphere.

on S^1 (labelled by the coordinate x^9) a D-string wrapped around the S^1 describes a particle charged under $B'_{9\mu}$. This then is a candidate soliton carrying charge quantum numbers $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ that is related to the $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ state via SL(2,Z) duality. As we had seen earlier, SL(2,Z) duality predicts that there should be a unique ultra-short multiplet carrying charge quantum numbers $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$. Thus our task now is as follows:

- Quantize the collective coordinates of this soliton.
- Verify if we get an ultra-short multiplet in this quantum theory.

Since the D-string is a one dimensional object, the dynamics of its collective coordinates should be described by a (1+1) dimensional field theory. As we had discussed earlier, all the vibrational modes of the D-string are given by the open string states with ends attached to the D-string. In particular, the zero frequency modes (collective modes) of the D-string that are relevant for analyzing the spectrum of BPS states correspond to *massless* open string states propagating on the D-string. By analyzing the spectrum of these open string states one finds that the collective coordinates in this case correspond to

- 8 bosonic fields y^m denoting the location of this string in eight transverse directions.
- A U(1) gauge field.
- 8 Majorana fermions.

It can be shown that the dynamics of these collective coordinates is described by a (1+1) dimensional supersymmetric quantum field theory which is the dimensional reduction of the N=1 supersymmetric U(1) gauge theory from (9+1) to (1+1) dimensions. Normally in (1+1) dimension gauge fields have no dynamics. But here since the space direction is compact, $y \equiv \oint A_1 dl$ is a physical variable. Furthermore, the compactness of U(1) makes y to be periodically identified (($y \equiv y + a$) for some a). Thus the momentum p_y conjugate to y is quantized ($p_y = 2\pi k/a$ with k integer.) It can be shown that[69] this momentum, which represents electric flux along the D-string, is actually a source of $B_{9\mu}$ charge! Thus if we restrict to the $p_y = 0$ sector then these states carry $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ charge quantum numbers as discussed earlier, but by taking $p_y = 2\pi k/a$, we can get states carrying charge quantum numbers $\begin{pmatrix} k \\ 1 \end{pmatrix}$ as well.

Due to the compactness of the space direction, we can actually regard this as a quantum mechanical system instead of a (1+1) dimensional quantum field theory. It turns out that in looking for ultra-short multiplets, we can ignore all modes carrying momentum along S^1. This corresponds to dimensionally reducing the theory to (0+1) dimensions. The degrees of freedom of this quantum mechanical system are:

- 8 bosonic coordinates y^m,
- 1 compact bosonic coordinate y,
- 16 fermionic coordinates.

A quantum state is labelled by the momenta conjugate to y^m (ordinary momenta) and an integer labelling momentum conjugate to y which can be identified with the quantum number p labelling $B_{9\mu}$ charge. The fermionic coordinates satisfy the sixteen dimensional Clifford algebra. Thus quantization of the fermionic coordinates gives $2^8 = 256$ -fold degeneracy, which is precisely the correct degeneracy for a ultra-short multiplet. This establishes the existence of all the required states of charge $\begin{pmatrix} p \\ 1 \end{pmatrix}$ predicted by SL(2,Z) symmetry.

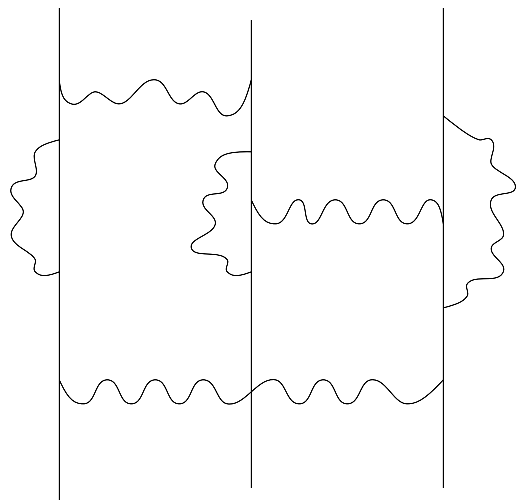


Figure 12: Possible open string states in the presence of three parallel D-strings.

What about $\begin{pmatrix} p \\ r \end{pmatrix}$ states with $r > 1$? These carry r units of $B'_{9\mu}$ charge and hence must arise as a bound state of r D-strings wrapped along S^1. Thus the first question we need to ask is: what is the (1+1) dimensional quantum field theory governing the dynamics of this system? In order to answer this question we need to study the dynamics of r D-strings. This system can be described as easily as a single D-string: instead of allowing open strings to end on a single D-string, we allow it to end on any of the r D-strings situated at

$$x^m = x^m_{(i)}\,, \qquad 2 \le m \le 9, \qquad 1 \le i \le r\,, \tag{4.28}$$

where $\vec{x}_{(i)}$ denotes the location of the i-th D-string. The situation is illustrated in Fig.12. Thus the dynamics of this system will now be described not only by the open strings starting and ending on the same D-string, but also by open strings whose two ends lie on two different D-strings.

For studying the spectrum of BPS states we need to focus our attention on the massless open string states. First of all, for each of the r D-strings we get a U(1) gauge field, eight scalar fields and eight Majorana fermions from open strings with both ends lying on that

D-string. But we can get extra massless states from open strings whose two ends lie on two different D-strings when these two D-strings coincide. It turns out that for r coincident D-strings the dynamics of massless strings on the D-string world-sheet is given by the dimensional reduction to (1+1) dimension of N=1 supersymmetric $U(r)$ gauge theory in ten dimensions, or equivalently, N=4 supersymmetric $U(r)$ gauge theory in four dimensions[69]. Following a logic similar to that in the case of a single D-string, one can show that the problem of computing the degeneracy of $\begin{pmatrix} p \\ r \end{pmatrix}$ states reduces to the computation of certain Witten index in this quantum theory. We shall not go through the details of this analysis, but just state the final result. It turns out that *there is a unique ultra-short multiplet for every pair of integers* (p, r) *which are relatively prime, precisely as predicted by* $SL(2, Z)$[69]!

A similar analysis can be carried out for the short multiplets that carry momentum k along S^1 besides carrying the B and B' charges p and r[69, 70]. In order to get these states from the D-brane spectrum, we can no longer dimensionally reduce the (1+1) dimensional theory to (0+1) dimensions. Instead we need to take into account the modes of the various fields of the (1+1) dimensional field theory carrying momentum along the internal S^1. The BPS states come from configurations where only the left- (or right-) moving modes on S^1 are excited. The calculation of the degeneracy $d(k, p, r)$ of BPS states carrying given charge quantum numbers (p, r, k) is done by determining in how many ways the total momentum k can be divided among the various left-moving bosonic and fermionic modes. This counting problem turns out to be identical to the one used to get the Hagedorn spectrum of BPS states in the elementary string spectrum, except that the elementary string is replaced here by the solitonic D-string. Naturally, we get back the Hegedorn spectrum for $d(k, p, r)$ as well. Thus the answer agrees exactly with that predicted by SL(2,Z) duality. This provides us with a test of the conjectured SL(2,Z) symmetry of type IIB on S^1.

The method of using D-branes to derive the dynamics of collective coordinates has been used to verify the predictions of other duality conjectures involving various string compactifications. Among them are self-duality of type II string theory on T^4[74, 75, 76, 77], the duality between heterotic on T^4 and type IIA on K3[78], the duality between type I and SO(32) heterotic string theory[29], etc.

4.3 Massless solitons and tensionless strings

An interesting aspect of the conjectured duality between the heterotic string theory on T^4 and type IIA string theory on K3 is that at special points in the moduli space the heterotic string theory has enhanced non-abelian gauge symmetry *e.g.* $E_8 \times E_8$ or SO(32) in the absence of vacuum expectation value of the internal components of the gauge fields, SU(2) at the self-dual radius etc. Perturbative type IIA string theory on K3 does not have any such gauge symmetry enhancement, since the spectrum of elementary string states does not contain any state charged under the U(1) gauge fields arising in the RR sector. Thus, for example, we do not have the $W^{\pm}$ bosons that are required for enhancing a U(1) gauge group to SU(2). At first sight this seems to lead to a contradiction. However upon closer examination one realises that this cannot really be a problem[150]. To see this let us consider a point in the moduli space of heterotic string theory on T^4 where the non-abelian gauge symmetry is broken. At this point the would be massless gauge bosons of the non-abelian gauge theory acquire mass by Higgs mechanism, and appear as BPS states in the abelian theory. As we approach the point of enhanced gauge symmetry, the masses of these states vanish. Since the masses of BPS states are determined by the BPS formula, the vanishing of the masses must be a consequence of the BPS formula. Thus if we are able to find the images of these BPS states on the type IIA side as appropriate D-brane states, then the masses of these D-brane states must also vanish as we approach the point in the moduli space where the heterotic theory has enhanced gauge symmetry. These massless D-brane solitons will then provide the states necessary for enhancing the gauge symmetry.

To see this more explicitly, let us examine the BPS formula. It can be shown that in the variables defined in section 3.3 the BPS formula is given by,

$$m^2 = e^{-\Phi^{(A)}/2} \alpha^T (L M^{(A)} L + L) \alpha \,, \tag{4.29}$$

where α is a 24 dimensional vector belonging to the lattice Λ'_{24}, and represents the U(1) charges carried by this particular state. For each $\vec{\alpha}$ we can assign an occupation number $n(\vec{\alpha})$ which gives the number of BPS multiplets carrying this specific set of charges. Since $M^{(A)}$ is a symmetric O(4,20) matrix, we can express this as $\Omega^{(A)T}\Omega^{(A)}$ for some O(4,20) matrix $\Omega^{(A)}$, and rewrite eq.(4.29) as

$$m^2 = e^{-\Phi^{(A)}/2} \alpha^T L \Omega^{(A)T} (I_{24} + L) \Omega^{(A)} L \alpha \,. \tag{4.30}$$

As can be seen from eq.(3.28), $(I_{24}+L)$ has 20 zero eigenvalues. As we vary $M^{(A)}$ and hence $\Omega^{(A)}$, the vector $\Omega^{(A)}L\alpha$ rotates in the twenty four dimensional space. If for some $\Omega^{(A)}$ it is aligned along one of the eigenvectors of $(I_{24}+L)$ with zero eigenvalue, we shall get massless solitons provided the occupation number $n(\vec{\alpha})$ for this specific $\vec{\alpha}$ is non-zero.

Although this argument resolves the problem at an abstract level, one would like to understand this mechanism directly by analysing the type IIA string theory, since, after all, we do not encounter massless solitons very often in physics. This has been possible through the work of [26, 151, 78]. For simplicity let us focus on the case of enhanced SU(2) gauge symmetry. First of all, one finds that at a generic point in the moduli space where SU(2) is broken, the images of the $W^{\pm}$ bosons in the type IIA theory are given by a D-2 brane wrapped around a certain 2-cycle (topologically non-trivial two dimensional surface) inside K3, the $+$ and the $-$ sign of the charge being obtained from two different orientations of the D-2 brane. Since the two tangential directions on the D-2 brane are directed along the two internal directions of K3 tangential to the 2-cycle, this object has no extension in any of the five non-compact spatial directions, and hence behaves like a particle.[22]. It turns out that as we approach the point in the moduli space where the theory on the heterotic side develops enhanced SU(2) gauge symmetry, the K3 on which type IIA theory is compactified becomes singular. At this singularity the area of the topologicaly non-trivial 2-cycle mentioned above goes to zero. As a result, the mass of the wrapped D-2 brane, obtained by multiplying the tension of the D-2 brane by the area of the two cycle, vanishes. This gives us the massless solitons that are required for the gauge symmetry enhancement. A similar mechanism works for getting other gauge groups as well. In fact it turns out that there is a one to one correspondence between the enhanced gauge groups, which are classified by A-D-E dynkin diagram, and the singularity type of K3, which are also classified by the A-D-E dynkin diagram[26]. This establishes an explicit physical relationship between A-D-E singularities and A-D-E lie algebras.

The appearance of enhanced gauge symmetry in type IIA on K3 poses another puzzle. Let us compactify this theory on one more circle. Since such a compactification does not destroy gauge symmetry, this theory also has enhanced gauge symmetry when the K3 becomes singular. But type IIA on $K3\times S^1$ is T-dual to type IIB on $K3\times S^1$; thus type IIB on $K3\times S^1$ must also develop enhanced gauge symmetry when K3 develops singularities. Does this imply that type IIB on K3 also develops enhanced gauge symmetry at these

[22]These states were analyzed in detail in [152]

special points in the K3 moduli space? This does not seem possible, since type IIB string theory does not have any D-2 brane solitons which can be wrapped around the collapsed two cycles of K3. It turns out that instead of acquiring enhanced gauge symmetry, type IIB string theory acquires tensionless strings at these special points in the K3 moduli space[153]. These arise from taking a D-3 brane of type IIB string theory, and wrapping it on a two cycle of K3. Thus two of the tangential directions of the three brane are directed along the internal directions of K3, and the third direction of the three brane is along one of the non-compact spatial directions. Thus from the point of view of the (5+1) dimensional theory such a configuration will appear as a string. The tension of this string is given by the product of the tension (energy per unit three volume) of the three brane and the area of the two cycle on which the three brane is wrapped. Thus as we approach the singular point on the K3 moduli space where the area of the two cycle vanishes, the tension of the string goes to zero. In other words, we get tensionless strings. Upon further compactification on a circle we get massless particles from configurations where this tensionless string is wound around the circle. These are precisely the massless gauge bosons required for the gauge symmetry enhancement in type IIB on $K3 \times S^1$.

5 Interrelation Between Different Duality Conjectures

In the last three sections we have seen many different duality conjectures and have learned how to test these conjectures. We shall now see that many of these conjectures are not independent, but can be 'derived' from each other. There are several different ways in which dualities can be related to each other. We shall discuss them one by one. The material covered in this section is taken mainly from [84, 85, 82].

5.1 Combining non-perturbative and T- dualities

Suppose a string theory A compactified on a manifold K_A has a conjectured duality symmetry group G. Now further compactify this theory on some manifold $\mathcal{M}$. Then the theory A on $K_A \times \mathcal{M}$ is expected to have the following set of duality symmetries:

- It inherits the original duality symmetry group G of A on K_A.
- It also has a perturbatively verifiable T-duality group. Let us call it H.

Quite often G and H do not commute and together generate a much bigger group[83, 30]. In that case, the existence of this bigger group of symmetries can be regarded as a consequence of the duality symmetry of A on K_A and T-dualities.

We shall illustrate this with a specific example[30]. We have seen that in ten dimensions type IIB string theory has a conjectured duality group SL(2,Z) that acts non-trivially on the coupling constant. From the table given in section 3.5 we see that type IIB on T^n also has a T-duality group $SO(n, n; Z)$, whose existence can be verified order by order in string perturbation theory. It turns out that typically these two duality groups do not commute, and in fact generate the full duality symmetry group of type IIB on T^n as given in the table of section 3.5. Thus we see that the existence of the full duality symmetry group of type IIB on T^n can be infered from the SL(2,Z) duality symmetry of the ten dimensional type IIB string theory, and the perturbatively verifiable T-duality symmetries of type IIB on T^n.

5.2 Duality of dualities

Suppose two theories are conjectured to be dual to each other, and each theory in turn has a conjectured self-duality group. Typically part of this self duality group is T-duality, and the rest involves non-trivial transformation of the coupling constant. But quite often the non-perturbative duality transformations in one theory correspond to T-duality in the dual theory and vice versa. As a result, the full self duality group in both theories follows from the conjectured duality between the two theories.

Again we shall illustrate this with an example[31, 26]. Let us start with the conjectured duality between heterotic on T^4 and type IIA on K3. Now let us compactify both theories further on a two dimensional torus T^2. This produces a dual pair of theories: type IIA on $K3 \times T^2$ and heterotic on T^6. Now, heterotic on T^6 has a T-duality group $O(6, 22; Z)$ that can be verified using heterotic perturbation theory. On the other hand , type IIA on $K3 \times T^2$ has a T-duality group $O(4, 20; Z) \times SL(2, Z) \times SL(2, Z)'$ that can be verified using type II perturbation theory. The full conjectured duality group in both theories is $O(6, 22, Z) \times SL(2, Z)$.

Now the question we would like to address is, how are the T-duality symmetry groups in the two theories embedded in the full conjectured $O(6, 22; Z) \times SL(2, Z)$ duality group? This has been illustrated in Fig.13. In particular we find that the SL(2,Z) factor of the full duality group is a subgroup of the T-duality group in type IIA on $K3 \times T^2$, and

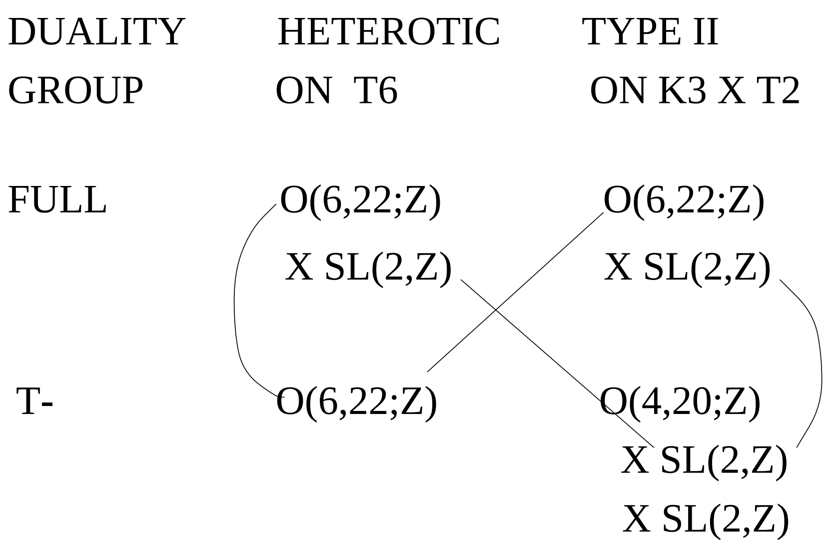


Figure 13: The embedding of the T-duality groups in the full duality group in heterotic on T^6 and type IIA on $K3 \times T^2$.

hence can be verified in this theory order by order in perturbation theory. On the other hand, the O(6,22;Z) factor of the duality group appears as a T-duality symmetry of the heterotic string theory, and hence can be verified order by order in perturbation theory in this theory. Thus assuming that T-duality in either theory is a valid symmetry, and the duality between the heterotic on T^4 and type IIA on K3, we can establish the existence of the self-duality group O(6,22;Z)×SL(2,Z) in heterotic on T^6 and type IIA on K3×T^2.

Using the results of this and the previous subsection, we see that so far among all the conjectured non-perturbative duality symmetries, the independent ones are:

1. SL(2,Z) of type IIB in D=10,

2. type I ↔ SO(32) heterotic in D=10, and

3. IIA on K3 ↔ heterotic on T^4.

We shall now show how to 'derive' 3) from 1) and 2).

5.3 Fiberwise duality transformation

In this subsection we shall describe the idea of constructing dual pairs of theories using fiberwise duality transformation[84]. Suppose (Theory A on $\mathcal{K}_A$) has been conjectured to

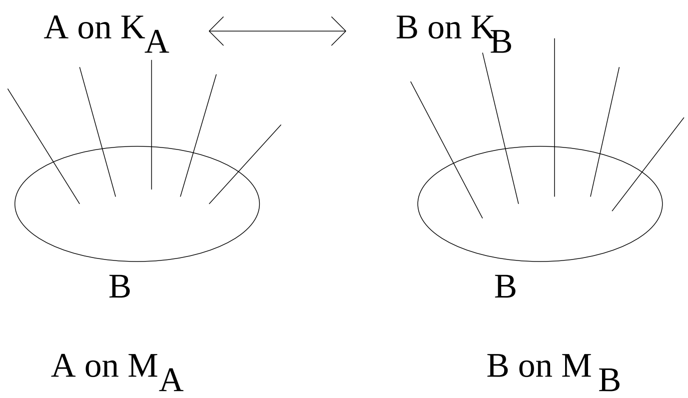


Figure 14: Application of fiberwise duality transformation. In each local neighbourhood of the base manifold $\mathcal{B}$, the two theories are equivalent due to the equivalence of the theories living on the fiber ($\times$ any manifold). Thus we would expect the theories A on $\mathcal{M}_A$ and B on $\mathcal{M}_B$ to be equivalent.

be dual to (Theory B on $\mathcal{K}_B$). Here A and B are two of the five different string theories in D=10, and $\mathcal{K}_A$, $\mathcal{K}_B$ are two different manifolds (in general). This duality involves a precise map between the moduli spaces of the two theories. Now construct a pair of new manifolds $\mathcal{M}_A$, $\mathcal{M}_B$ by starting from some other manifold $\mathcal{B}$, and erecting at every point on $\mathcal{B}$ a copy of $\mathcal{K}_A$, $\mathcal{K}_B$. The moduli of $\mathcal{K}_A$, $\mathcal{K}_B$ vary slowly over $\mathcal{B}$ and are related to each other via the duality map that relates (A on $\mathcal{K}_A$) to (B on $\mathcal{K}_B$). Then we would expect a duality

$$\text{Theory A on } \mathcal{M}_A \leftrightarrow \text{Theory B on } \mathcal{M}_B$$

by applying the duality transformation fiberwise. This then gives rise to a new duality conjecture. This situation has been illustrated in Fig.14.

Now suppose that at some isolated points (or subspaces of codimension ≥ 1) on $\mathcal{B}$ the fibers $\mathcal{K}_A$ and $\mathcal{K}_B$ degenerate. (We shall see some explicit examples of this later.) Is the duality between (A on $\mathcal{M}_A$) and (B on $\mathcal{M}_B$) still valid? We might expect that even in this case the duality between the two theories holds since the singularities occur on subspaces of 'measure zero'. Although there is no rigorous argument as to why this should be so, this appears to be the case in all known examples. Conversely, assuming that this is the case, we can derive the existence of many new duality symmetries from a given duality symmetry.

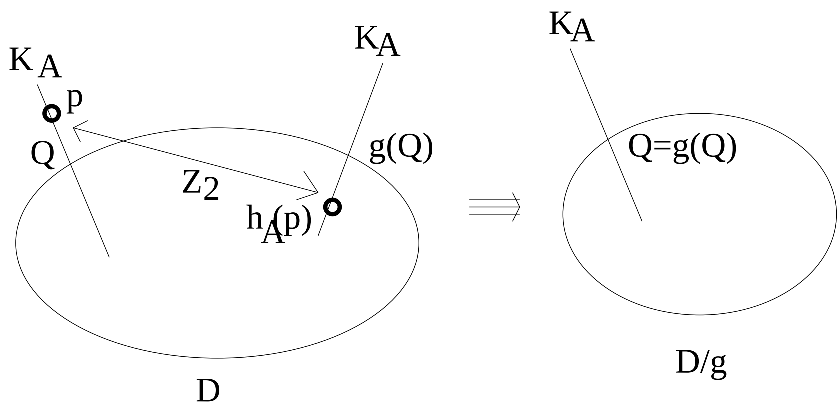


Figure 15: Representation of a Z_2 orbifold as a fibered space. The Z_2 transformation relates the point (Q, p) on $\mathcal{D} \times \mathcal{K}_A$ to the point $(g(Q), h_A(p))$.

A special case of this construction involves Z_2 orbifolds. Suppose we have a dual pair (A on $\mathcal{K}_A$) $\leftrightarrow$ (B on $\mathcal{K}_B$). Further suppose that (A on $\mathcal{K}_A$) has a Z_2 symmetry generated by h_A. Then the dual theory must also have a Z_2 symmetry generated by h_B. h_A and h_B are mapped to each other under duality. Now compactify both theories on another manifold $\mathcal{D}$ with a Z_2 isometry generated by g, and compare the two quotient theories

$$\text{(A on } K_A \times \mathcal{D}/h_A \cdot g) \text{ and (B on } K_B \times \mathcal{D}/h_B \cdot g)$$

$(\mathcal{K}_A \times \mathcal{D}/h_A \cdot g)$ is obtained from the product manifold $\mathcal{K}_A \times \mathcal{D}$ by identifying points that are related by the Z_2 transformation $h_A \cdot g$. This situation is illustrated in Fig. 15. As shown in this figure, $(\mathcal{K}_A \times \mathcal{D}/h_A \cdot g)$ admits a fibration with base $\mathcal{D}/g$ and fiber $\mathcal{K}_A$. In particular, note that since $h_A \cdot g$ takes a point $(p \in \mathcal{K}_A, Q \in \mathcal{D})$ to $(h_A(p), g(Q))$, if we focus our attention on a definite point Q on $\mathcal{D}$, then there is no identification of the points in the copy of $\mathcal{K}_A$ that is sitting at Q. This shows that the fiber is $\mathcal{K}_A$ and *not* $\mathcal{K}_A/h_A$. As we go from Q to $g(Q)$, which is a closed cycle on $\mathcal{D}/g$, the fiber gets twisted by the transformation h_A.

The second theory, B on $(\mathcal{K}_B \times \mathcal{D})/(h_B \cdot g)$ has an identical structure. Thus we can now apply fiberwise duality transformation to derive a new duality:

$$\text{(A on } K_A \times \mathcal{D}/h_A \cdot g) \leftrightarrow \text{(B on } K_B \times \mathcal{D}/h_B \cdot g)$$

Note that if $P_0 \in \mathcal{D}$ is a fixed point of g (i.e. if $g(P_0) = P_0$) then in $\mathcal{K}_A \times \mathcal{D}/h_A \cdot g$ there is an identification of points (p, P_0) and $(h_A(p), P_0)$. Similarly in $\mathcal{K}_B \times \mathcal{D}/h_B \cdot g$ there is an identification of points (p', P_0) and $(h_B(p'), P_0)$. Thus at P_0 the fibers degenerate to $\mathcal{K}_A/h_A$ and $\mathcal{K}_B/h_B$ respectively. At these points the argument in support of duality between the two theories breaks down. However, as we have discuused earlier, since these are points of 'measure zero' on $\mathcal{D}$, we would expect that the two quotient theories are still dual to each other[85]. We shall now illustrate this construction in the context of a specific example.

We start with type IIB string theory in ten dimensions. This has a conjectured SL(2,Z) symmetry. Let S denote the SL(2,Z) element

$$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}. \tag{5.1}$$

Recall that this theory also has two global discrete symmetries $(-1)^{F_L}$ and Ω. The action of S, $(-1)^{F_L}$ and Ω on the massless bosonic fields in this theory were described in section 3.4. From this one can explicitly compute the action of $S(-1)^{F_L}S^{-1}$ on these massless fields. This action turns out to be identical to that of Ω. A similar result holds for their action on the massless fermionic fields as well. Finally, since the action of S on the massive fields is not known, one can define this action in such a way that the actions of $S(-1)^{F_L}S^{-1}$ and Ω are identical on all states. This gives:

$$S(-1)^{F_L}S^{-1} = \Omega\,. \tag{5.2}$$

We are now ready to apply our formalism. We take (A on $\mathcal{K}_A$) to be type IIB in D=10, (B on $\mathcal{K}_B$) to be type IIB in D=10 transformed by S, h_A to be $(-1)^{F_L}$, h_B to be Ω, $\mathcal{D}$ to be T^4, and g to be the transformation $\mathcal{I}_4$ that changes the sign of all the coordinates on T^4. This gives the duality:

$$(\text{IIB on } T^4/(-1)^{F_L}\cdot\mathcal{I}_4) \leftrightarrow (\text{IIB on } T^4/\Omega\cdot\mathcal{I}_4)$$

Note that in this case the fibers $\mathcal{K}_A$ and $\mathcal{K}_B$ are points, but this does not prevent us from applying our method of constructing dual pairs. Also there are sixteen fixed points on T^4 under $\mathcal{I}_4$ where the application of fiberwise duality transformation breaks down, but as has been argued before, we still expect the duality to hold since these are points of measure zero on T^4.

We shall now bring this duality into a more familiar form via T-duality transformation. Let us make $R \to (1/R)$ duality transformation on one of the circles of T^4 in the theory on the left hand side. This converts type IIB theory to type IIA. This also transforms $(-1)^{F_L}\cdot$

$\mathcal{I}_4$ to $\mathcal{I}_4$, which can be checked by explicitly studying the action of these transformations on the various massless fields. Thus the theory on the left hand side is T-dual to type IIA on $T^4/\mathcal{I}_4$. This of course is just a special case of type IIA on K3.

Let us now take the theory on the right hand side and make $R \to (1/R)$ duality transformation on all four circles. This takes type IIB theory to type IIB theory. But this transforms $\Omega \cdot \mathcal{I}_4$ to Ω, which can again be seen by studying the action of these transformations on the massless fields. Thus the theory on the right is T-dual to type IIB on T^4/Ω. Since type I string theory can be regarded as type IIB string theory modded out by Ω, we see that the theory on the right hand side is type I on T^4. But by (heterotic - type I) duality in ten dimensions this is dual to heterotic on T^4. Thus we have 'derived' the duality

$$(\text{Type IIA on K3}) \leftrightarrow (\text{Heterotic on } T^4)$$

from other conjectured dualities in D=10. Although this way the duality has been established only at a particular point in the moduli space (the orbifold limit of K3), the argument can be generalized to establish this duality at a generic point in the moduli space as well[85].

There are many other applications of fiberwise duality transformation. Some of them will be discussed later in this review.

5.4 Recovering higher dimensional dualities from lower dimensional ones

So far we have discussed methods of deriving dualities involving compactified string theories by starting with the duality symmetries of string theories in higher dimensions. But we can also proceed in the reverse direction. Suppose a string theory compactified on a manifold $\mathcal{M}_1 \times \mathcal{M}_2$ has a self-duality symmetry group G. Now consider the limit when the size of $\mathcal{M}_2$ goes to infinity. A generic element of G, acting on this configuration, will convert this configuration to one where $\mathcal{M}_2$ has small or finite size. However, there may be a subgroup H of G that commutes with this limit, *i.e.* any element of this subgroup, acting on a configuration where $\mathcal{M}_2$ is big, gives us back a configuration where $\mathcal{M}_2$ is big. Thus we would expect that H is the duality symmetry group of the theory in the decompactification limit, *i.e.* of the original string theory compactified on $\mathcal{M}_1$. The same argument can be extended to the case of a pair of dual theories.

A priori this procedure does not appear to be very useful, since one normally likes to derive more complicated duality transformations of lower dimensional theories from the simpler ones in the higher dimensional theory. But we shall now show how this procedure can be used to derive the SL(2,Z) duality symmetry of type IIB string theory from the conjectured duality between type I and SO(32) heterotic string theories, and T-duality symmetries of the heterotic string theory. We shall describe the main steps in this argument, for details, see [82]. We start with the duality between type I on T^2 and heterotic on T^2 that follows from the duality between these theories in ten dimensions. Now heterotic string theory on T^2 has a T-duality group O(2,18;Z). We shall focus our attention on an $SL(2,Z) \times SL(2,Z)$ subgroup of this T-duality group. One of these two SL(2,Z) factors is associated with the global diffeomorphism of T^2, and the other one is associated with the $R \to (1/R)$ duality symmetries on the two circles. By the 'duality of dualities' argument, this must also be a symmetry of type I on T^2. Since type I string theory can be regarded as type IIB string theory modded out by the world-sheet parity transformation Ω discussed in section 3.4, we conclude that $SL(2,Z) \times SL(2,Z)$ is a subgroup of the self-duality group of type IIB on T^2/Ω. Let us now make an $R \to (1/R)$ duality transformation on both the circles of this T^2. This converts type IIB on T^2 to type IIB theory compactified on a dual T^2, and Ω to $(-1)^{F_L} \cdot \Omega \cdot \mathcal{I}_2$, where $\mathcal{I}_2$ denotes the reversal of orientation of both the circles of T^2. (This can be seen by studying the action of various transformations on the massless fields.) Geometrically, this model describes type IIB string theory compactified on the surface of a tetrahedron (which is geometrically $T^2/\mathcal{I}_2$), with an added twist of $(-1)^{F_L} \cdot \Omega$ as we go around any of the four vertices of the tetrahedron (the fixed points of $\mathcal{I}_2$). (This theory will be discussed in more detail in section 8). Thus we conclude that $SL(2,Z) \times SL(2,Z)$ is a subgroup of the self-duality group of type IIB on a tetrahedron. Now take the limit where the size of the tetrahedron goes to infinity. It turns out that both the SL(2,Z) factors commute with this limit. One of these SL(2,Z) groups becomes part of the diffeomorphism group of type IIB string theory and does not correspond to anything new, but the other SL(2,Z) factor represents the S-duality transformation discussed in section 3.4. Since this limit gives us back the decompactified type IIB string theory, we conclude that type IIB string theory in ten dimensions has a self-duality group SL(2,Z).

Thus we see that all the dualities discussed so far can be 'derived' from a single duality conjecture, − the one between type I and SO(32) heterotic string theories in ten dimensions. In the next section we shall see more examples of dualities which can be derived from the ones that we have already discussed.

6 Duality in Theories with Less than Sixteen Supersymmetry Generators

So far our discussion has been focussed on theories with 16 or more supersymmetry charges. As was pointed out in section 3, for these theories the non-renormalization theorems for the low energy effective action and the spectrum of BPS states are particularly powerful. This makes it easy to test duality conjectures involving these theories. In this section we shall extend our discussion to theories with eight supercharges. Examples of such theories are provided by N=2 supersymmetric theories in four dimensions. We shall see that these theories have a very rich structure, and although the non-renormalization theorems are less powerful here, they are still powerful enough to provide us with some of the most striking tests of duality conjectures involving these theories. The material covered in this section is based mainly on refs.[86, 87, 84].

6.1 Construction of dual pair of theories with eight supercharges

For definiteness we shall focus our attention on N=2 supersymmetric theories in four dimensions. There are several ways to get theories with $N = 2$ supersymmetry in four dimensions. Two of them are:

1. Type IIA/IIB on Calabi-Yau 3-folds: In our convention an n-fold describes an n complex or $2n$ real dimensional manifold. In ten dimensions type II theories have 32 supersymmetry generators. Compactification on a Calabi-Yau 3-fold breaks 3/4 of the supersymmetry. Thus we are left with 8 supersymmetry generators in D=4, giving rise to N=2 supersymmetry.

2. Heterotic string theory on $K3 \times T^2$: In ten dimensions heterotic string theory has sixteen supersymmetry generators. Compactification on $K3 \times T^2$ breaks half of the supersymmetry. Thus we have a theory with eight supersymmetry generators, again giving N=2 supersymmetry in four dimensions. It is also possible to construct more general class of four dimensional heterotic string theories with the same number of

supersymmetries where the background does not have the product structure $K3 \times T^2$[122, 86].

The question we would like to ask is: is it possible to construct pairs of N=2 supersymmetric type II and heterotic string compactifications in four dimensions which will be non-perturbatively dual to each other? Historically such dual pairs were first constructed by trial and error[86] and then a more systematic approach was developed[87, 84, 88, 89, 90]. However we shall begin by describing the systematic approach, and then describe how one tests these dualities. The systematic construction of such dual pairs can be carried out by application of fiberwise duality transformation as described in the last section. The steps involved in this construction are as follows:

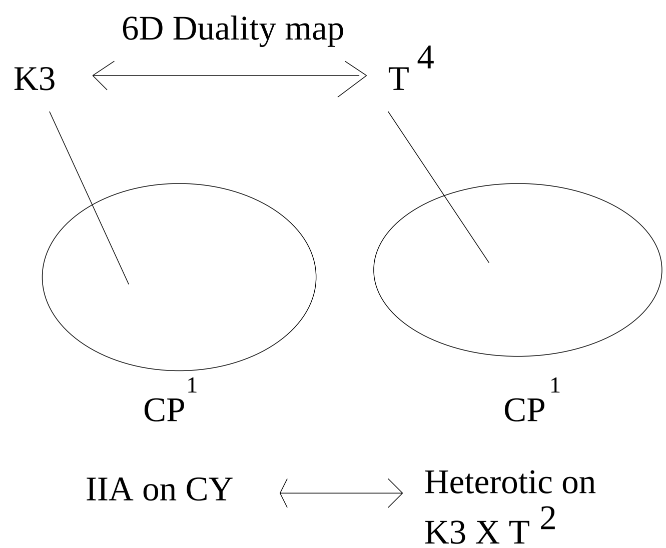


Figure 16: Construction of dual pair of N=2 supersymmetric string theories in four dimensions from the dual pair of theories in six dimensions.

- Start from the conjectured duality (Type IIA on K3) $\leftrightarrow$ (Heterotic on T^4).
- Choose a CP^1 base.
- Construct a Calabi-Yau 3-fold by fibering K3 over the base CP^1. One can construct a whole class of Calabi-Yau manifolds this way by choosing different ways of varying K3 over CP^1.

- For type IIA on each such Calabi-Yau 3-fold we can get a dual heterotic compactification by replacing the type IIA on $K3$ by heterotic on T^4 on each fiber according to the duality map. This gives heterotic string theory on a manifold obtained by varying T^4 on CP^1 according to the duality map. Typically this manifold turns out to be $K3 \times T^2$ or some variant of this. This model is expected to be dual to the type IIA string theory on the Calabi-Yau manifold that we started with. Thus we get a duality map

$$\text{(Type IIA on CY)} \leftrightarrow \text{(Heterotic on K3}\times T^2\text{)}$$

This construction has been illustrated in Fig.16. Note that the original duality map gives a precise relationship between the moduli of type IIA on K3 and heterotic on T^4. On the heterotic side the moduli involve background gauge fields on T^4 besides the shape and size of T^4. Thus for a specific Calabi-Yau, knowing how K3 varies over CP^1, we can find out how on the heterotic side the background gauge fields on T^4 vary as we move along CP^1. This gives the gauge field configuration on $K3 \times T^2$. Different Calabi-Yau manifolds will give rise to different gauge fields on K3$\times T^2$.

We shall illustrate this procedure with the example of a pair of Z_2 orbifolds of the form[87]:

$$\text{(IIA on } K3 \times T^2/h_A \cdot g) \leftrightarrow \text{(Heterotic on } T^4 \times T^2/h_B \cdot g)$$

where g acts on T^2 by changing the sign of both its coordinates, h_A is a specific involution of K3 known as the Enriques involution, and h_B is the image of this transformation on the heterotic side. By our previous argument relating orbifolds to fibered spaces, these two theories are expected to be dual to each other via fiberwise duality transformation. $K3 \times T^2/(h_A \cdot g)$ can be shown to describe a Calabi-Yau manifold. Thus the theory on the left-hand side corresponds to type IIA string theory compactified on this Calabi-Yau manifold. In order to determine the theory on the heterotic side, we need to determine h_B. We shall now describe this procedure in some detail.

In order to determine h_B, we need to study the relationship between the fields appearing in type IIA on K3 and heterotic on T^4. The low energy effective action of both the theories and the origin of the various massless fields in these theories were discussed in section 3.3. We shall focus our attention on the gauge fields. As discussed there, in the type IIA on K3, 22 of the gauge fields come from decomposing the three form field along the harmonic two forms on K3. Now, h_A, being a geometric transformation on K3, has known action on the harmonic forms $\omega^{(p)}$. For this particular example, h_A corresponds to

- exchanging ten of the $\omega^{(p)}$ with ten others and
- changing the sign of two more $\omega^{(p)}$.

This translates into a similar action on the fields $\mathcal{A}_\mu^{(p)}$ defined in eq.(3.36). Furthermore, h_A leaves the other two gauge fields, coming from the ten dimensional gauge field A_μ and the dual of $C_{\mu\nu\rho}$ invariant. We can now translate this into an action on the gauge fields in heterotic on T^4. It turns out that the action on the heterotic side is given by:

- exchanging the gauge fields in the two E_8 factors ,
- exchanging ($G_{9\mu}$, $B_{9\mu}$) with ($G_{8\mu}$, $B_{8\mu}$), and,
- changing the sign of ($G_{7\mu}$ and $B_{7\mu}$).

This translates into the following geometric action in heterotic string theory on T^4:[23]

- exchange of two E_8 factors in the gauge group,
- $x^8 \leftrightarrow x^9$,
- $x^7 \to -x^7$.

This is h_B.[24] It turns out that modding out heterotic string theory on T^6 by the transformation $h_B \cdot g$ produces an N=2 supersymmetric theory. Thus this construction gives a type II - heterotic dual pair with N=2 supersymmetry.

Using the idea of fiberwise duality transformation we can construct many more examples of heterotic - type IIA dual pairs in four dimensions with N=2 supersymmetry. Quite often using mirror symmetry[123] we can also relate this to IIB string theory on a mirror Calabi-Yau manifold.

[23]Here we are regarding this theory as the $E_8 \times E_8$ heterotic string theory compactified on T^4. By the duality between the two heterotic string theories upon compactification on a circle, this is equivalent to SO(32) heterotic string theory compactified on T^4.

[24] We need to add to this a non-geometrical shift involving half of a lattice vector in Λ_{24} in order to get a modular invariant theory on the heterotic side. This transformation is not visible in perturbative type IIA theory.

6.2 Test of duality conjectures involving theories with eight supercharges

Given such a dual pair of theories constructed by application of fiberwise duality transformation, the next question will be: how do we test if these theories are really dual to each other? After all, as we have seen, there is no rigorous proof that fiberwise duality transformation always produces a correct dual pair of theories, particularly when the fiber degenerates at some points / regions in the base. Unlike in the case of theories with sixteen supercharges, one cannot directly compare the tree level low energy effective action in the two theories, as they undergo quantum corrections in general. Furthermore, in this theory the spectrum of BPS saturated states can change discontinuously as we move in the moduli space[95]. Hence the spectrum computed at weak coupling cannot always be trusted at strong coupling. Nevertheless there are some non-renormalization theorems which allow us to test these proposed dualities, as we shall now describe.

Matter multiplets in N=2 supersymmetric theories in four dimensions are of two types. (For a review, see [95].) They are

- vector multiplet containing one vector, one complex scalar, and two Majorana fermions, and
- hypermultiplet containing two complex scalars and two Majorana fermions.

Let us consider a theory at a generic point in the moduli space where the massless matter fields include only abelian gauge fields and neutral hypermultiplets. Let $\vec{\phi}$ denote the complex scalars in the vector multiplet, and $\vec{\psi}$ denote the complex scalars in the hypermultiplet. The N=2 supersymmetry requires that there is no coupling between the vector and the hypermultiplets in those terms in the low energy effective action S_{eff} which contain at most two space-time derivatives[91]. Thus the scalar kinetic terms appearing in the lagrangian density associated with S_{eff} must be of the form:

$$G^V_{m\bar{n}}(\vec{\phi})\partial_\mu\phi^m\partial^\mu\bar{\phi}^n + G^H_{\alpha\bar{\beta}}(\vec{\psi})\partial_\mu\psi^\alpha\partial^\mu\bar{\psi}^\beta\,, \tag{6.1}$$

where G^V and G^H are appropriate metrics in the vector and the hypermultiplet moduli spaces. The kinetic terms of the vectors and the fermionic fields are related to these scalar kinetic terms by the requirement of N=2 supersymmetry.

This decoupling between the hyper- and the vector- multiplet moduli spaces by itself is not of much help, since each term may be independently modified by quantum corrections.[25] But in string theory we have some extra ingredient[86, 84]. Recall that the coupling constant in string theory involves the dilaton. Thus quantum corrections to a given term must involve a coupling to the dilaton. Now consider the following two special cases.

1. The dilaton belongs to a hypermultiplet. Then there can be no correction to the vector multiplet kinetic term since such corrections will give a coupling between the dilaton and the vector multiplet.

2. The dilaton belongs to a vector multiplet. In this case the same argument shows that there can be no correction to the hypermultiplet kinetic term.

In type IIA/IIB string theory on Calabi-Yau manifold the dilaton belongs to a hypermultiplet. Thus in these theories the vector multiplet kinetic term, calculated at the tree level, is exact. On the other hand in heterotic on $K3 \times T^2$, the dilaton is in the vector multiplet. Thus the hypermultiplet kinetic term, calculated at the tree level, is exact. Using this information we can adopt the following strategy for testing duality.[26]

1. Take a type II - heterotic dual pair and calculate the vector multiplet kinetic term exactly from the tree level analysis on the type II side.

2. Using the map between the fields in the type II and the heterotic theory, we can rewrite the exact vector multiplet kinetic term in terms of the heterotic variables.

3. In particular the heterotic variables include the heterotic dilaton Φ_H which is in the vector multiplet. So we can now expand the exact answer in powers of e^{Φ_H} and compare this answer with the explicit calculations in heterotic string perturbation theory. Typically the expansion involves tree, one loop, and non-perturbative terms. (There is no perturbative contribution in the heterotic theory beyond one loop due to some Adler-Bardeen type non-renormalization theorems.) Thus one can compare

[25] There are however strong restrictions on what kind of metric G_V and G_H should describe. In particular G_V must describe a special Kahler geometry[124, 91], whereas G_H must describe a quaternionic geometry[125]. However, these restrictions do not fix G_V and G_H completely.

[26] Here we describe the test using the vector multiplet kinetic term, but a similar analysis should be possible with the hypermultiplet kinetic term as well.

the expected tree and one loop terms, calculated explicitly in the heterotic string theory, with the expansion of the exact answer.

The results of the above calculation in heterotic and type II string theories agree in all the cases tested[86, 92, 93]! This agreement is quite remarkable, since the one loop calculation is highly non-trivial on the heterotic side, and involves integrals over the moduli space of the torus. Indeed, the agreement between the two answers is a consequence of highly non-trivial mathematical identities.

Given that the tree and one loop results in the heterotic string theory agree with the expansion of the exact result on the type II side, one might ask if a similar agreement can be found for the non-perturbative contribution from the heterotic string theory as well. From the exact answer calculated from the type II side we know what this contribution should be. But we cannot calculate it directly on the heterotic side, since there is no non-perturbative formulation of string theory. However, one can take an appropriate limit in which the stringy effects on the heterotic side disappear and the theory reduces to some appropriate $N = 2$ supersymmetric quantum field theory.[27] Thus now the calculation of these non-perturbative effects on the heterotic side reduces to a calculation in the N=2 supersymmetric field theory. This can be carried out using the method developed by Seiberg and Witten[95]. Again there is perfect agreement with the results from the type II side[94]. Besides providing a non-trivial test of string duality, this also shows that the complete Seiberg-Witten[95] results (and more) are contained in the classical geometry of Calabi-Yau spaces!

7 M-theory

So far we have discussed dualities that relate known string theories. However, sometime analysis similar to those that lead to various duality conjectures can also lead to the discovery of new theories. One such theory is a conjectured theory living in eleven dimensions. This theory is now known as M-theory. In this section we shall give a brief description of this theory following refs.[26, 100, 101, 103, 104, 102, 105].

[27]This is in the same spirit as in the case of toroidal compactification of heterotic string theory, where, by going near a special point in the moduli space, we can effectively get an N=4 supersymmetric Yang-Mills theory.

7.1 M-theory in eleven dimensions

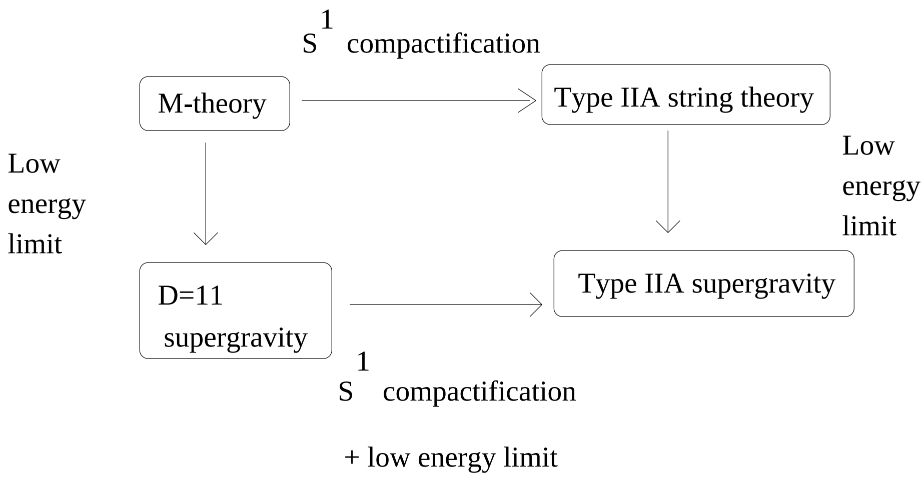


Figure 17: The relationship between M-theory and various other supergravity / string theories.

The arguments leading to the existence of M-theory goes as follows[96, 26]. Take type IIA string theory in ten dimensions. The low energy effective action of this theory is non-chiral N=2 supergravity in ten dimensions. It is well known that this can be obtained from the dimensional reduction of $N = 1$ supergravity in eleven dimensions[97]. More specifically, the relationship between the two theories is as follows. The bosonic fields in $N = 1$ supergravity theory in eleven dimensions consist of the metric $g^{(S)}_{MN}$ and a rank three anti-symmetric tensor field $C^{(S)}_{MNP}$ $(0 \le M, N \le 10)$. The bosonic part of the action of this theory is given by[126]

$$
\begin{aligned}
S_{SG} \quad = \quad & \frac{1}{(2\pi)^8} \int d^{11}x \Big[\sqrt{-g^{(S)}} \Big(R^{(S)} - \frac{1}{48} G^{(S)2} \Big) \\
& - \frac{1}{(12)^4} \varepsilon^{\mu_0 \cdots \mu_{10}} C^{(S)}_{\mu_0 \mu_1 \mu_2} G^{(S)}_{\mu_3 \cdots \mu_8} G^{(S)}_{\mu_7 \cdots \mu_{10}} \Big] , \qquad (7.1)
\end{aligned}
$$

where $G^{(S)} \sim dC^{(S)}$ is the four form field strength associated with the three form field $C^{(S)}$. In writing down the above equation we have set the eleven dimensional Planck mass to unity (or equivalently we can say that we have absorbed it into a redefinition of the metric.) Let us now compactify this supergravity theory on a circle of radius $R(\sim \sqrt{g^{(S)}_{10,10}})$ measured in the supergravity metric $g^{(S)}_{MN}$ and ignore (for the time being) the Kaluza-Klein modes carrying momentum in the internal direction. Then the effective action in the dimensionally reduced theory agrees with that of type IIA string theory given in (3.33) under the identification[97]:

$$\sqrt{g^{(S)}_{10,10}} = e^{\Phi/3}, \qquad g^{(S)}_{\mu\nu} \simeq e^{-\Phi/12} g_{\mu\nu} \qquad g^{(S)}_{10\mu} \simeq e^{2\Phi/3} A_\mu,$$
$$C^{(S)}_{\mu\nu\rho} \simeq C_{\mu\nu\rho}, \qquad C^{(S)}_{10\mu\nu} \simeq B_{\mu\nu}, \qquad (0 \le \mu, \nu \le 9)\,. \tag{7.2}$$

Here $\simeq$ denotes equality up to additive terms involving second and higher powers in fields. We are using the convention that $\Phi = 0$ corresponds to compactification on a circle of unit radius. Note that as the radius $R(\sim \sqrt{g^{(S)}_{10,10}})$ approaches ∞, $\Phi \to \infty$. This corresponds to strong coupling limit of the type IIA string theory. This leads one to the conjecture[96, 26] that *in the strong coupling limit type IIA string theory approaches an 11 dimensional Lorentz invariant theory, whose low energy limit is 11-dimensional N=1 supergravity.* This theory has been called M-theory. The situation is illustrated in Fig.17. Part of the conjecture is just the definition of M-theory as the strong coupling limit of type IIA string theory. The non-trivial part of the conjecture is that it describes a Lorentz invariant theory in eleven dimensions.

The evidence for the existence of an eleven dimensional theory, as discussed so far, has been analogous to the evidence for various duality conjectures based on the comparison of their low energy effective action. One might ask if there are more precise tests involving the spectrum of BPS states. There are indeed such tests. M theory on S^1 will have Kaluza-Klein modes representing states in the eleven dimensional N=1 supergravity multiplet carrying momentum along the compact x^{10} direction. These are BPS states, and can be shown to belong to the 256 dimensional ultra-short representation of the supersymmetry algebra. The charge quantum number characterizing such a state is the momentum (k/R) along S^1. Thus for every integer k we should find such BPS states in type IIA string theory in ten dimensions. In M-theory these states carry k units of $g^{(S)}_{10\mu}$ charge. Since $g^{(S)}_{10\mu}$ gets mapped to A_μ under the M-theory - IIA duality, these states must carry k units of A_μ charge in type IIA string theory. If we now recall that in type IIA string theory A_μ arises in the RR sector, we see that these states cannot come from elementary string states, as elementary string excitations are neutral under RR sector gauge fields. However Dirichlet 0-branes in this theory do carry A_μ charge. In particular the state with $k = 1$ corresponds to a single Dirichlet zero brane. As usual, the collective coordinate dynamics of the 0-branes is determined from the dynamics of massless open string states with ends lying on the D0-brane, and in this case is described by the dimensional reduction of N=1 super-

Maxwell theory from (9+1) to (0+1) dimensions. This theory has sixteen fermion zero modes whose quantization leads to a $2^8 = 256$ fold degenerate state. Thus we see that we indeed have an ultra-short multiplet with unit A_μ charge, as predicted by the M-theory - IIA duality conjecture.

What about states with $k > 1$? In type IIA string theory these must arise as bound states of k D0-branes. Dynamics of collective coordinates of k D0 branes is given by the dimensional reduction of N=1 supersymmetric U(k) gauge theory from (9+1) to (0+1) dimensions. Thus the number of ultra-short multiplets with k-units of A_μ charge is determined in terms of the number of normalizable supersymmetric ground states of this quantum mechanical system. Finding these bound states is much more difficult than the bound state problems discussed earlier. The main obstacle to this analysis is that a charge k state has the same energy as k charge 1 states at rest. Thus the bound states we are looking for sit at the bottom of a continuum. Such states are difficult to study. For $k = 2$ such a bound state with the correct degeneracy has been found[98]. The analysis for higher k still remains to be done.

The analysis can be simplified by compactifying M-theory on T^2 and considering the Kaluza-Klein modes carrying (k_1, k_2) units of momenta along the two S^1's. Assuming that the two S^1's are orthogonal, and have radii R_1 and R_2 respectively, the mass of such a state, up to a proportionality factor, is

$$\sqrt{\Big(\frac{k_1}{R_1}\Big)^2 + \Big(\frac{k_2}{R_2}\Big)^2}\,. \tag{7.3}$$

For (k_1, k_2) relatively prime, such a state has strictly less energy than the sum of the masses of any other set of states with the same total charge[99]. Thus one should be able to find these states in type IIA string theory on S^1 (which, according to the conjecture, is equivalent to M-theory on T^2) without encountering the difficulties mentioned earlier. By following the same kind of argument, these states can be shown to be in one to one correspondence to a class of supersymmetric vacua in a (1+1) dimensional supersymmetric gauge theory compactified on a circle.[28] All such states have been found with degeneracy as predicted by the M-theory - IIA duality.

There are also other consistency checks on the proposed M-theory - IIA duality. Consider M-theory on T^2. According to M-theory - type IIA duality, it is dual to IIA on

[28] In fact, these states are related via an $R \to (1/R)$ duality transformation to the ultra-short multiplets in type IIB on S^1 discussed in section 4.2.

S^1. But we know that IIA on S^1 is related by T-duality to IIB on S^1. Thus we have a duality between M-theory on T^2 and IIB on S^1. Now IIB on S^1 has an SL(2,Z) strong-weak coupling duality inherited from ten dimensional type IIB string theory. Thus one might ask, what does it correspond to in M-theory on T^2? One can find the answer by using the known map between the massless fields in the two theories, and the action of SL(2,Z) in type IIB string theory. It turns out that this SL(2,Z) symmetry in M-theory is simply the group of global diffeomorphisms of T^2[68, 101, 100]. Thus we again have an example of 'duality of dualities'. The SL(2,Z) of IIB is a non-perturbative symmetry. But in M-theory on T^2 it is simply a consequence of the diffeomorphism invariance of the 11-dimensional theory.

Turning this analysis around we see that this also supports the ansatz that M-theory, defined as the strong coupling limit of IIA, is a fully Lorentz invariant theory in eleven dimension. The argument goes as follows:

- First of all, from Lorentz invariance of type IIA string theory we know that we have Lorentz invariance in coordinates $x^0, \ldots x^9$ when all the coordinates $x^0, \ldots x^9$ are non-compact.

- Then from the conjectured SL(2,Z) duality symmetry of type IIB string theory we know that we have an exchange symmetry between the 9th and the 10th coordinate of M-theory when these coordinates are compact. In the limit when the radius of both the compact circles are taken to be large, this would mean that we should have Lorentz invariance in coordinates $x^0, \ldots x^{10}$.

7.2 Compactification of M-theory

Given the existence of M-theory, we can now construct new vacua of the theory by compactifying M-theory on various manifolds. (For a review of compactification of eleven dimensional supergravity, see [198]. For example, we can consider M-theory compactified on K3, Calabi-Yau, and various orbifolds. These can all be regarded as appropriate strong coupling limits of type IIA compactification on the same manifold. But in general these cannot be regarded as perturbative string vacua. The essential feature of this strong coupling limit is the emergence of Lorentz invariance in one higher dimension. For example, M-theory on a Calabi-Yau manifold gives a five dimensional theory with N=1 supersymmetry[197]. Such a theory cannot be constructed by conventional compactification of type IIA string theory at weak coupling.

Of course in many cases these non-perturbative vacua are related to perturbative string vacua by conjectured duality relations. These duality conjectures can be arrived at by using arguments very similar to those used in arriving at string duality conjectures. Some examples of such conjectured dualities are given below[102, 103, 104, 105]:

M-theory on		
S^1/Z_2	$\leftrightarrow$	$(E_8 \times E_8)$ heterotic in D=10
$K3$	$\leftrightarrow$	Heterotic/Type I on T^3
T^5/Z_2	$\leftrightarrow$	IIB on K3
T^8/Z_2	$\leftrightarrow$	Type I/Heterotic on T^7
T^9/Z_2	$\leftrightarrow$	Type IIB on T^8/Z_2

In each case Z_2 acts by reversing the sign of all the coordinates of T^n; for odd n this is also accompanied by a reversal of sign of $C^{(S)}_{MNP}$. Each of these duality conjectures satisfy the consistency condition that the theory on the right hand side, upon further compactification on a circle, is dual to type IIA string theory compactified on the manifold on the left hand side.

The duality between M-theory on S^1/Z_2 and the $E_8 \times E_8$ heterotic string theory is particularly amusing. Here the Z_2 transformation acts by reversing the orientation of S^1, together with a change of sign of the three form field $C^{(S)}_{MNP}$. S^1/Z_2 denotes a real line segment bounded by the two fixed points on S^1. It turns out that the two E_8 gauge multiplets arise from 'twisted sector' of the theory and sit at the two ends of this line segment. The supergravity sector, on the other hand, sits in the bulk. Now in the conventional heterotic string compactification on Calabi-Yau spaces, all the observed gauge bosons and charged particles come from one E_8 and are neutral under the second E_8[187]. The second E_8, known as the hidden sector or the shadow world, is expected to be responsible for supersymmetry breaking. In the M-theory picture these two sectors are physically separated in space. In other words, the real world and the shadow world live at two ends of the line and interact only via the exchange of supergravity multiplets propagating in the bulk[196]! It has been suggested that this physical separation could be as large as a millimeter[106]! This limit comes from the analysis of the fifth force experiment, since if this dimension is too large, we should have inverse cube law for the gravitational force instead of inverse square law. No such direct limit comes from the inverse square law of gauge interaction, since gauge fields live on the boundary of S^1/Z_2 and hence do not get affected by the existence of this extra dimension.

Many of the listed duality conjectures involving M-theory (in fact all except the first one) can be derived by fiberwise duality transformation[105].[29] Let us for example consider the duality

$$(\text{M theory on } T^5/Z_2) \leftrightarrow (\text{type IIB on K3})$$

The Z_2 generator is $\mathcal{I}_5 \cdot \sigma$ where $\mathcal{I}_5$ changes the sign of all five coordinates $(x^6, \ldots x^{10})$ on T^5, and σ denotes the transformation $C^{(S)}_{MNP} \to -C^{(S)}_{MNP}$. Let us express this as $(\mathcal{I}_1 \cdot \sigma) \cdot \mathcal{I}_4$ where $\mathcal{I}_1$ changes the sign of x^{10}, and $\mathcal{I}_4$ changes the sign of $(x^6, \ldots x^9)$. We now use the result of fiberwise duality transformation:

$$(\text{A on } K_A \times \mathcal{D}/(h_A \cdot g)) \equiv (\text{B on } K_B \times \mathcal{D}/(h_B \cdot g))$$

by choosing A on K_A to be M-theory on S^1, B on K_B to be type IIA string theory, h_A to be $\mathcal{I}_1 \cdot \sigma$, h_B to be $(-1)^{F_L}$ (this can be shown to be the image of h_A in the type IIA string theory), $\mathcal{D}$ to be T^4 spanned by $x^6, \ldots x^9$, and g to be $\mathcal{I}_4$. Thus we get the duality

$$\text{M-theory on } (S^1 \times T^4/\mathcal{I}_1 \cdot \sigma \cdot \mathcal{I}_4) \leftrightarrow \text{IIA on } T^4/(-1)^{F_L} \cdot \mathcal{I}_4$$

The theory on the left hand side is M-theory on T^5/Z_2. On the other hand, if we take the theory on the right hand side and make an $R \to (1/R)$ duality transformation on one of the circles, it converts

- type IIA theory to type IIB theory, and
- $(-1)^{F_L} \cdot \mathcal{I}_4$ into $\mathcal{I}_4$.

Thus the theory on the right is dual to type IIB on $T^4/\mathcal{I}_4$, which is a special case of type IIB on $K3$. Thus we get the duality:

$$(\text{M-theory on } T^5/Z_2) \leftrightarrow (\text{IIB on K3})$$

[29]The duality between $E_8 \times E_8$ heterotic string theory and M-theory on S^1 can be 'derived' from other known duality conjectures by taking the infinite radius limit of a lower dimensional duality relation[82].

This duality was first conjectured in [103, 104].

As in the case of type IIA string theory, we can get non-perturbative enhancement of gauge symmetries in M-theory when the compact manifold develops singularities[26, 193, 194]. M-theory contains classical membrane and five-brane soliton solutions carrying electric and magnetic charges of $C^{(S)}_{MNP}$ respectively[8]. The extra massless states required for this symmetry enhancement come from membranes wrapped around the collapsed two cycles of the singular manifold.

8 F-Theory

Just as M-theory can be used to describe non-perturbative compactification of type IIA string theory, F-theory describes non-perturbative compactification of type IIB string theory[107, 108]. However, unlike M-theory, it does not correspond to a Lorentz invariant higher dimensional theory, although, as we shall see, an auxiliary manifold with two extra dimensions plays a crucial role in the construction of F-theory compactification. This section will be based mainly on refs.[107, 108, 112].

8.1 Definition of F-theory

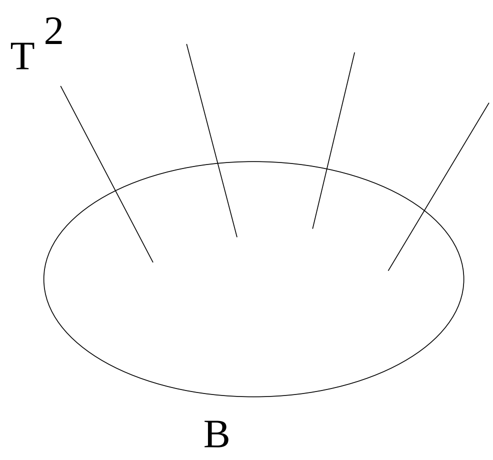


Figure 18: An elliptically fibered manifold $\mathcal{M}$ with base $\mathcal{B}$.

In conventional perturbative type IIB compactification one takes the dilaton-axion field λ (defined in section 3.4) to be constant. F-theory is a novel way of compactifying type IIB theory that avoids this restriction and allows the string coupling to vary over the compact manifold. The starting point in this construction is an elliptically fibered Calabi-Yau manifold defined as follows. Let $\mathcal{B}$ be a manifold (which we shall call base manifold)of real dimension d and $\mathcal{M}$ be another manifold of real dimension $d+2$, obtained by erecting at every point of $\mathcal{B}$ a copy of a torus T^2, with the moduli of T^2 varying over $\mathcal{B}$. This situation is illustrated in Fig.18. $\mathcal{M}$ is called an elliptically fibered manifold. We shall choose the base $\mathcal{B}$ and the fibration in such a way that $\mathcal{M}$ describes a Calabi-Yau manifold. Let $\vec{z}$ denote the complex coordinate on $\mathcal{B}$, and $\tau(\vec{z})$ denote the complex structure of T^2 as a function of $\vec{z}$. Then F-theory on $\mathcal{M}$ is defined to be type IIB string theory compactified on $\mathcal{B}$ with

$$\lambda(\vec{z}) = \tau(\vec{z})\,. \tag{8.1}$$

Note that the size of the fiber torus does not appear in eq.(8.1). Thus F-theory on a manifold $\mathcal{M}$ is insensitive to a subset of the moduli of $\mathcal{M}$ which describe how the size of the fiber torus varies over the base.

In order that $\mathcal{M}$ is well defined, $\lambda = \tau(\vec{z})$ must come back to its original value only up to an SL(2,Z) transformation as we move along a closed cycle on $\mathcal{B}$. Due to the presence of this non-perturbative duality transformation in the monodromy group, conventional type IIB perturbation theory cannot be used to describe this system. In particular, there are points in $\mathcal{B}$ where $Im(\lambda)$ is of order unity, and hence type IIB theory is strongly coupled. For example, suppose that $Im(\lambda)$ is large in one region of the manifold, and also that as we go around a closed curve starting from this region, there is an SL(2,Z) transformation by the matrix $\begin{pmatrix} p & q \\ r & s \end{pmatrix}$ so that λ comes back to a value near p/r. Then at some point on the curve $Im(\lambda)$ must be finite, and hence the string theory is strongly coupled.

From this note it would seem that although F-theory describes a novel way of compactifying type IIB string theory, we cannot extract any information about such a theory, since string perturbation theory cannot be used to analyse this system. However, it turns out that we can learn quite a lot about these theories by using various known duality relations. For this consider F-theory on $\mathcal{M} \times S^1$, *i.e.* type IIB theory on $\mathcal{B} \times S^1$ with $\lambda(\vec{z}) = \tau(\vec{z})$. Now, as we have discussed in the last section, type IIB theory on S^1 is dual to M-theory on T^2, with λ being the modular parameter of T^2. Thus we can now apply fiberwise duality transformation illustrated in Fig.19 to relate the F-theory compactification on $\mathcal{M} \times S^1$ to an M-theory compactification. Under this duality the modulus of T^2 on the right hand side must be set equal to $\lambda(\vec{z})$ of the theory on the left, which,

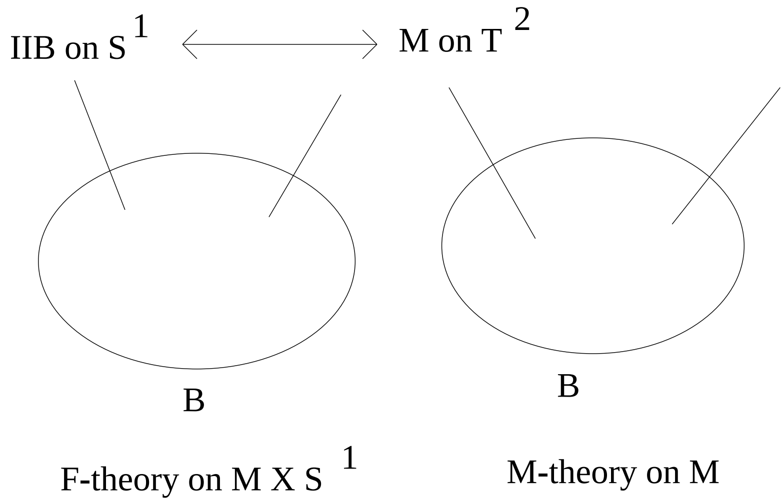


Figure 19: Fiberwise application of duality to relate F-theory on $\mathcal{M} \times S^1$ to M-theory on $\mathcal{M}$.

according to eq.(8.1), is just $\tau(\vec{z})$. This means that the manifold on the right hand side is the original manifold $\mathcal{M}$ that we started with. This gives the duality

$$\text{F-theory on } \mathcal{M} \times S^1 \leftrightarrow \text{M-theory on } \mathcal{M}$$

Thus many of the properties of F-theory on $\mathcal{M}$ (*e.g.* the number of supersymmetries, spectrum of massless states, etc.) can be studied from that of M-theory on $\mathcal{M}$ and then taking appropriate limit in which the size of S^1 on the F-theory side goes to ∞. Using this one can show for example that if $\mathcal{M}$ is K3, then we preserve half of the space-time supersymmetries of the type IIB theory, whereas if $\mathcal{M}$ is a Calabi-Yau manifold then we preserve $1/4$ of the supersymmetry.

In order to gain more insight into various F-theory compactifications, we need to develop a convenient formalism for describing elliptically fibered manifolds. Our starting point will be the equation describing a torus:

$$y^2 = x^3 + fx + g\,, \tag{8.2}$$

where x and y are complex variables, and f and g are complex parameters. For every pair of constants f and g the above equation describes a one complex dimensional surface, which can be shown to be a torus. The modular parameter τ of the torus is given by

$$j(\tau) = \frac{4.(24f)^3}{27g^2 + 4f^3}\,, \tag{8.3}$$

where $j(\tau)$ is a known function of τ. In fact it is the unique modular invariant function with a single pole at $\tau = i\infty$ and zeroes at $\tau = e^{i\pi/3}$. The overall normalization of j is chosen in such a way that $j(i) = (24)^3$.

In order to describe elliptically fibered manifold on some base $\mathcal{B}$ we simply need to make f and g depend on the coordinates of $\mathcal{B}$. We shall illustrate this with the help of an elliptically fibered K3. Here we choose the base $\mathcal{B}$ to be CP^1. Thus the elliptically fibered manifold $\mathcal{M}$ is described by the equation

$$y^2 = x^3 + f(z)x + g(z)\,, \tag{8.4}$$

where z is the coordinate on CP^1. It can be shown that $\mathcal{M}$ describes a K3 manifold provided $f(z)$ is a polynomial of degree 8, and $g(z)$ is a polynomial of degree 12 in z. The modular parameter of the torus varies over the base CP^1 according to the relation:

$$j(\tau(z)) = \frac{4.(24f(z))^3}{27g(z)^2 + 4f(z)^3}\,. \tag{8.5}$$

By definition F-theory on this elliptically fibered K3 is type IIB string theory compactified on CP^1 with:

$$\lambda = \tau(z)\,. \tag{8.6}$$

In order to specify the background completely we also need to specify the metric on the base CP^1. This can be calculated from the low energy effective field theory when the size of $\mathcal{B}$ is sufficiently large, and the answer is[109]

$$ds^2 = F(\tau(z), \bar\tau(\bar z))\, dz d\bar z \Big(\prod_{i=1}^{24} (z - z_i)^{-1/12} (\bar z - \bar z_i)^{-1/12} \Big)\,, \tag{8.7}$$

where z_i are the zeroes of $\Delta \equiv (4f^3 + 27g^2)$, and

$$F(\tau, \bar\tau) = (\tau_2)\eta(\tau)^2 \bar\eta(\bar\tau)^2\,. \tag{8.8}$$

$\eta(\tau)$ denotes the Dedekind function and τ_2 is the imaginary part of τ.

Similarly we can describe more complicated F-theory compactifications by choosing more complicated base $\mathcal{B}$. For example, consider the base $CP^1 \times CP^1$ labelled by a pair of complex coordinates (z, w)[108]. We can get an elliptically fibered manifold on this base by the equation

$$y^2 = x^3 + f(z, w)x + g(z, w)\,. \tag{8.9}$$

In order that this manifold is Calabi-Yau, we need $f(z, w)$ to be a polynomial of degree (8,8) in (z, w), and $g(z, w)$ to be a polynomial of degree (12,12) in (z, w). F-theory on such a manifold is by definition a configuration of $\lambda(z, w)$ described by the equation

$$j(\lambda(z, w)) = \frac{4.(24f)^3}{27g^2 + 4f^3} . \tag{8.10}$$

8.2 Dualities involving F-theory

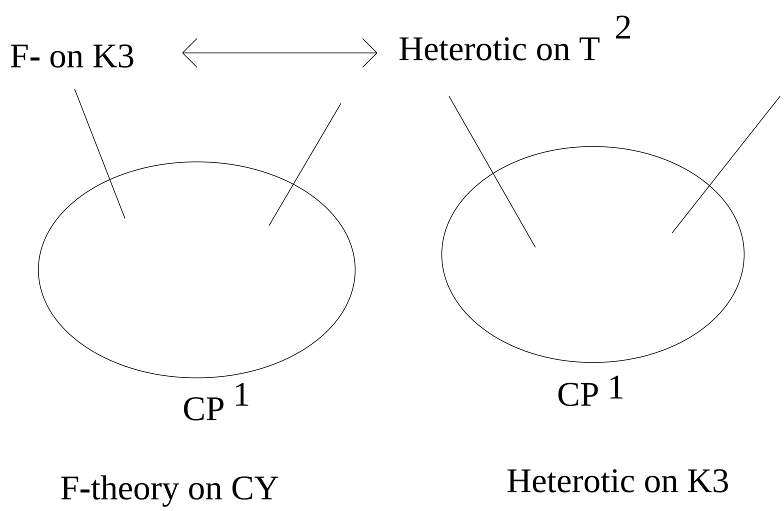


Figure 20: Fiberwise application of duality to relate F-theory on Calabi-Yau to heterotic string theory on K3.

There are many conjectured dualities involving F-theory compactifications. Some examples are given below[108]:

$$\begin{array}{ccc} \text{F-theory on K3} & \leftrightarrow & \text{Heterotic on } T^2 \\ \text{F-theory on CY 3-fold} & \leftrightarrow & \text{Heterotic on K3} \end{array}$$

All the duality conjectures involving F-theory have the following property:
If F-theory on $\mathcal{M}$ is dual to some string theory S compactified on a manifold $\mathcal{K}$, then M-theory on $\mathcal{M}$ must be dual to the same string theory compactified on $\mathcal{K} \times S^1$, and type IIA on $\mathcal{M}$ must be dual to the same string theory compactified on $\mathcal{K} \times T^2$. These results follow from the duality between type IIB on S^1 and M-theory on T^2, and that between M-theory on S^1 and type IIA string theory.

All conjectured dualities involving F-theory on Calabi-Yau 3-folds and more complicated manifolds can be derived from the fiberwise duality transformation[108, 110, 111]. For example, for Calabi-Yau 3-fold, this is done by representing the Calabi-Yau 3-fold as K3 fibered over CP^1, and replacing F-theory on $K3$ by heterotic on T^2 fiberwise. This has been illustrated in Fig.20. The theory on the right hand side of this figure represents heterotic string theory on K3 with apropriate gauge field background.

We shall now show how to derive the parent duality

$$(\text{F-theory on K3}) \leftrightarrow (\text{Heterotic on } T^2)$$

from other known duality conjectures[112]. Recall that for this background:

$$\begin{aligned} j(\lambda(z)) &= \frac{4.(24f(z))^3}{27g(z)^2 + 4f(z)^3} \\ ds^2 &= F(\lambda, \bar{\lambda})\, dz d\bar{z} \Big(\prod_{i=1}^{24} (z - z_i)^{-1/12} (\bar{z} - \bar{z}_i)^{-1/12} \Big) , \end{aligned} \tag{8.11}$$

where $f(z)$ and $g(z)$ are polynomials of degree 8 and 12 respectively, and z_i are the zeroes of $\Delta \equiv (4f^3 + 27g^2)$. The strategy is to try to go to a special point in the moduli space where λ, instead of varying over CP^1, becomes a constant. At this special point the theory reduces to a conventional compactification of type IIB string theory. Examining eq.(8.11) we see that this requires f^3/g^2 to be a constant. If we now recall that f is a polynomial of degree 8 in z and g is a polynomial of degree 12 in z, we see that for f^3/g^2 to be constant, we need

$$f = \alpha \phi^2, \qquad g = \phi^3 \, , \tag{8.12}$$

where ϕ is a polynomial of degree 4 in z, and α is a constant. Using the freedom of an overall rescaling of ϕ which does not change the value of λ, we can take

$$\phi = \prod_{m=1}^{4} (z - z_m) \, . \tag{8.13}$$

This gives

$$\Delta \equiv 4f^3 + 27g^2 = (4\alpha^3 + 27) \prod_{m=1}^{4} (z - z_m)^6 \, , \tag{8.14}$$

$$j(\lambda) = 4.(24\alpha)^3/(4\alpha^3 + 27) \, , \tag{8.15}$$

and

$$ds^2 = F(\lambda, \bar{\lambda})\, dz d\bar{z} \prod_{m=1}^{4} (z - z_m)^{-1/2} (\bar{z} - \bar{z}_m)^{-1/2} \, . \tag{8.16}$$

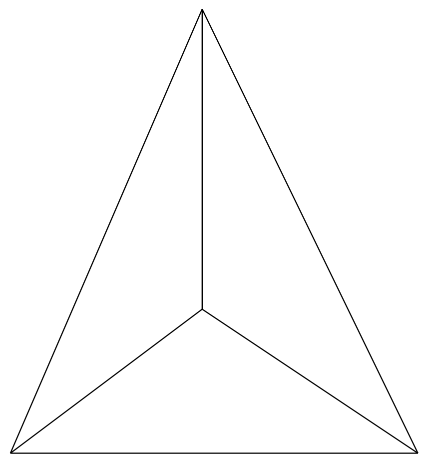


Figure 21: CP^1 with a flat metric except at four points.

Since λ is now a constant, the metric can be simplified by going to a new coordinate system w defined by:

$$dw = \prod_{m=1}^{4} (z - z_m)^{-1/2} dz\,. \tag{8.17}$$

Then

$$ds^2 = C\, dw\, d\bar{w}\,, \tag{8.18}$$

where $C = F(\lambda, \bar{\lambda})$ is a constant. Thus the metric is flat! But this poses a puzzle, since we know that the base is CP^1, and that we cannot put a flat metric on CP^1 since it has non-zero Euler number. The resolution to this puzzle comes from noting that if w_m are the images of $z = z_m$ in the w plane, then, near $z = z_m$,

$$(w - w_m) \sim (z - z_m)^{1/2}\,. \tag{8.19}$$

This gives rise to a deficit angle of π at each of these four points. Thus the base has flat metric everywhere except for conical singularities at these four points. This represents a regular tetrahedron as shown in Fig.21. This can be also be identified as the orbifold $T^2/\mathcal{I}_2$. Here w is the complex coordinate on T^2, and $\mathcal{I}_2$ denotes the transformation $w \to -w$. z on the other hand, is the coordinate on $T^2/\mathcal{I}_2$. The points z_m are the fixed points of $\mathcal{I}_2$.

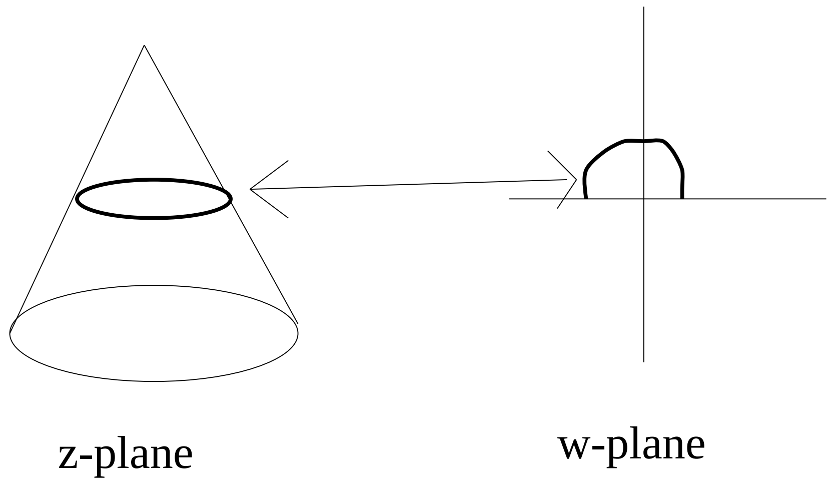


Figure 22: A closed curve around $z = z_m$ in the z-plane and its image in the w-plane.

This analysis would suggest that at this special point in the moduli space F-theory on K3 has reduced to type IIB on $T^2/\mathcal{I}_2$ with constant λ given in eq.(8.15). However, there is a further subtlety. Recall that going once around a fixed point in the z plane corresponds to going from w to $-w$ as illustrated in Fig.22. The relevant question to ask would be: is there any twist by some internal symmetry transformation g of type IIB theory as we go around a fixed point of z? If there is such a twist, then the Z_2 orbifold group will be generated by $w \to -w$, accompanied by the transformation g. In order to find g we need to study the effect of going around the point $z = z_m$ once. To do this, recall the equation describing this particular K3:

$$y^2 = x^3 + \alpha x \prod_{m=1}^{4} (z - z_m)^2 + \prod_{m=1}^{4} (z - z_m)^3 \,. \tag{8.20}$$

Let us take z around z_1 once through the parametrization

$$\prod_{m=1}^{4} (z - z_m) = e^{2\pi i t} \prod_{m=1}^{4} (z_{initial} - z_m) \,, \tag{8.21}$$

and continuously changing t from 0 to 1. Also during this change, focus on a point on the fiber torus and follow its trajectory. This can be achieved by choosing:

$$x = x_{initial} e^{2\pi i t}, \qquad y = y_{initial} e^{3\pi i t} \,. \tag{8.22}$$

This point lies on the surface (8.20) for all values of t if the initial point $(x_{initial}, y_{initial}, z_{initial})$ lies on this surface. At the end of this process when $t = 1$, we do not return to the original point but to $(x_{initial}, -y_{initial}, z_{initial})$. To see what this transformation correponds to if we choose a more conventional coordinate system to the fiber torus, let us denote by u is the conventional flat coordinate on the fiber torus, so that the torus is described through the identification $u \equiv u + 1 \equiv u + \tau$. It can be shown that u is related to (x, y) through the relation $du = Kdx/y$ where K is a constant. Thus $(x, y) \to (x, -y)$ corresponds to $u \to -u$, *i.e.* reversing the orientation of both the circles on the fiber torus. This is nothing but an SL(2,Z) transformation with matrix $\begin{pmatrix} -1 & \\ & -1 \end{pmatrix}$. Thus we see that as we move around any of the points $z = z_m$ on a the base once, we make an SL(2,Z) transformation with this matrix. But now recall that the SL(2,Z) transformation with the matrix $\begin{pmatrix} -1 & \\ & -1 \end{pmatrix}$ can be identified to the transformation $(-1)^{F_L} \cdot \Omega$ as discussed in section 3.4. Thus as we go around the point $z = z_m$ we make a global symmetry transformation by $(-1)^{F_L} \cdot \Omega$. This shows that the F-theory on K3 at this particular point in the moduli space can be identified to

$$\text{Type IIB on } T^2/\mathcal{I}_2 \cdot (-1)^{F_L} \cdot \Omega$$

We have encountered this theory earlier in section 5.4. As discussed there, by making $R \to (1/R)$ duality transformation on both coordinates of T^2, we can relate this theory type IIB on T^2/Ω. But since type IIB modded out by Ω is type I theory, we see that this model can be identified to type I on T^2. Finally, using type I - SO(32) heterotic duality in ten dimension, we can relate this model to heterotic string theory on T^2. This finally establishes the equivalence between

$$\text{(F-theory on K3) and (Heterotic on } T^2)$$

Similar strategy has been used to establish many other dualities involving F-theory compactification[188].

As in the case of type IIA string theory and M-theory, F-theory on a singular manifold can also develop enhanced non-abelian gauge symmetry[108, 189, 190]. In this case the extra massless states required for the symmetry enhancement come from open string states lying on the base[191, 192].

The F-theory type compactification can be generalized in the following way. One starts with type II string theory compactified on T^n with a duality group G, and compactifies it further on a base $\mathcal{B}$ with the monodromy on the base being a subgroup of G. Such compactifications have been discussed in [195].

9 Microscopic Derivation of Black Hole Entropy

One of the major stumbling blocks to our understanding of nature has been the apparent incompatibility between quantum mechanics and general relativity. Three of the four known forces of nature – strong, weak and electromagnetic interactions – are very well explained by quantum field fheory. The current model of elementary particle physics – the standard model – has explained most of the observed phenomena involving these three interactions. However, this is not so for gravity.

There are many problems of quantizing gravity using quantum field theory. First of all, gravity is not perturbatively renormalizable *i.e.* the usual rules in quantum field theory for extracting finite answers for all physical quantities from infinite answers at the intermediate stages of calculation, are not valid for gravity. As mentioned in section 1, this problem is automatically solved in string theory. However, during the last twenty five years a more serious objection to the compatibility of general relativity and quantum mechanics has been raised[127, 128]. This is the problem to which we now turn. The discussion in this section will follow closely refs.[133, 136, 137, 138, 121, 142].

9.1 Problem with black holes in quantum mechanics

The starting point is the existence of a class of classical solutions in general relativity (possibly coupled to other fields) known as black holes. Classically black holes are completely black. In other words, objects can fall into the black hole, but nothing can ever come out of a black hole. (In more technical terms, one says that black holes have event horizons.) Black holes in general relativity also satisfy a classical no hair theorem which states that a black hole solution is completely characterized by its mass, angular momentum, and gauge charges. Thus all other information (quantum numbers) carried by an object falling into the black hole is lost for ever.

However due to the work of Bekenstein, Hawking and others during the last twenty five years it has become clear that once quantum effects are taken into account, this picture of black hole undergoes dramatic modification. In particular two things happen.

- Black holes emit thermal radiation at temperature[127]:

$$T = \frac{\kappa}{2\pi}\,, \tag{9.1}$$

where κ is the surface gravity of the black hole (the acceleration due to gravity felt by a static observer at the event horizon).

- Black holes carry entropy[128]:

$$S = \frac{1}{4G_N}A\,. \tag{9.2}$$

where A is the area of the event horizon and G_N is the Newton's constant.

Black holes satisfy the usual laws of thermodynamics in terms of these variables. In particular the first law of black hole thermodynamics states that

$$dM = TdS\,. \tag{9.3}$$

This relates the change in the mass M of a black hole to its change in entropy S and the Hawking temperature T. The second law of black hole thermodynamics states that

$$dS \geq 0\,, \tag{9.4}$$

i.e. the sum of the entropy of the black hole and the usual thermodynamic entropy of its surroundings increases with time. In the presence of U(1) gauge charges we can also define chemical potential associated with each gauge charge. In the presence of these charges the laws of thermodynamics are modified in the usual manner.

It is this thermodynamic description of the black hole that causes an apparent conflict with quantum mechanics. This is best illustrated by considering the following thought experiment.

- Consider a black hole formed out of the collapse of a pure state. We can imagine a spherical shell of matter described by an s-wave state collapsing to form a black hole.

- It will then emit thermal radiation and at the end evaporate completely. If the outgoing radiation is really thermal, then the final state is a mixed state.

Thus the net result of this two step process is the evolution of a pure state to a mixed state, in conflict with the rules of quantum mechanics.

At this stage, it is useful to compare this with the phenomenon of thermal radiation from a star (or any other hot object). If an object in a pure quantum state is thrown into a star, it comes out as thermal radiation, so why doesn't this contradict quantum mechanics? The answer to this is that the thermal description of the radiation from a star is a result of averaging over the microstates of the star. We could, in principle, start from a pure quantum state of the star, and give a microscopic description of the radiation coming out of the star. In this description a pure state evolves to a pure state. In other words, although on average the star emits thermal radiation, the radiation coming out of the star has subtle dependence on what goes into the star, and a detailed analysis of this radiation can be used to completely reconstruct the initial state.

Let us now come back to black holes. Why can't the same reasoning be used for black holes to resolve the apparent conflict with quantum mechanics? The reason is that there is no similar microscopic description of the radiation from a black hole in conventional semiclassical gravity − the approximation that is used in demonstrating that black holes emit thermal radiation. This is related to the problem that there is no understanding of the black hole entropy in terms of counting of microstates. In other words, the entropy formula (9.2) is derived purely in analogy with thermodynamics, but not as the logarithm of density of states as in statistical mechanics. Thus there is no possibility of giving a quantum mechanical description of Hawking radiation by studying the evolution of individual microstates of the black hole.

Since string theory claims to be a consistent quantum theory of gravity, it should be able to explain black hole entropy and Hawking radiation in terms of conventional quantum mechanics. We shall now discuss some of the attempts to explain black hole thermodynamics in string theory.

9.2 Black holes as elementary string states

As we have discussed earlier (see, for example, eq.(4.9)) the spectrum of an elementary string contains an infinite tower of states. Since the Schwarzschild radius (the radius of the event horizon) of a black hole is proportional to its mass, for sufficiently large mass the Schwarzschild radius of a black hole will be larger than the string scale. In that case an elementary string state of the same mass will lie inside its Schwarzschild radius, and become a black hole[129]. This opens up the possibility of a statistical description of black hole entropy as follows[130]. For a given mass M, the microscopic entropy S_{micro} can be defined to be the logarithm of the number of elementary string states at that mass level. We can then compare this with the Bekenstein-Hawking entropy S_{BH} of the black hole with the same mass. If the two expressions agree, then we can give a statistical interpretation of black hole entropy by attributing it to the degeneracy of string states. Unfortunately this attempt fails, since one finds that

$$S_{micro} \propto M, \qquad S_{BH} \propto M^2 . \tag{9.5}$$

The above formulae hold for chargeless black holes, but a similar discrepancy is present even for black holes carrying electric charge. This seems to be a severe blow to the attempt at giving a microscopic description of Bekenstein-Hawking entropy in string theory. But one should keep in mind that in the region of parameter space where the elementary string state becomes a black hole, there are strong coupling effects, and hence the mass of an elementary string state can get renormalized[130]. Thus the parameter M that appears in the computation of S_{BH} and the M that appears in the computation of S_{micro} may not be the same, but may be related by a renormalization factor.

Various attempts have been made to get out of this difficulty for Schwarschild black holes[130, 131, 132, 186], but we shall not discuss them here. Instead, we shall try to get out of this impasse by working with states for which there is no mass renormalization, namely the BPS states[133]. Since the degeneracy of a BPS state does not change as we change the coupling (at least for theories with ≥ 16 supersymmetry charges) we can proceed as follows. First we compute the degeneracy of BPS states at weak coupling where the microscopic description of the state is reliable. Then we increase the string coupling constant to a sufficiently large value where the state becomes a black hole. For this black hole, S_{micro} should be given by the logarithm of the degeneracy computed at weak coupling. This leads to the following strategy for comparing S_{micro} and S_{BH}:

1. Identify the BPS states among elementary string states and calculate their degeneracy. This gives S_{micro}.

2. Identify BPS black holes (also known as extremal black holes) with the same quantum numbers and find S_{BH} by computing the area of the event horizon.

3. Compare the two expressions.

We shall illustrate this with the help of a specific model − heterotic string theory compactified on T^6[133]. As discussed above eq.(4.9), this theory has two classes of U(1) charges, $\vec{k}_L$ and $\vec{k}_R$. From eqs.(4.9), (4.10) we see that among the elementary string states, BPS states are those satisfying[30]

$$m^2 = (2\vec{k}_R^2/\langle\lambda_2\rangle)\,, \qquad N_L = \frac{1}{4}\langle\lambda_2\rangle\Big(m^2 - \frac{2\vec{k}_L^2}{\langle\lambda_2\rangle}\Big) + 1\,. \tag{9.6}$$

The degeneracy of these states can be calculated from eq.(4.12). For large N_L, this gives

$$d(m, k_L, \langle\lambda_2\rangle) \sim \exp(4\pi\sqrt{N_L}) \sim \exp\left(2\pi\sqrt{\langle\lambda_2\rangle}\sqrt{m^2 - \frac{2\vec{k}_L^2}{\langle\lambda_2\rangle}}\right). \tag{9.7}$$

This gives

$$S_{micro} \simeq 2\pi\sqrt{\langle\lambda_2\rangle}\sqrt{m^2 - \frac{2\vec{k}_L^2}{\langle\lambda_2\rangle}}\,. \tag{9.8}$$

Our next task is to calculate the Bekenstein-Hawking entropy S_{BH} for black holes carrying the same quantum numbers. It turns out that the black holes carrying these quantum numbers have vanishing area of the event horizon, and hence $S_{BH} = 0$. Thus again we seem to have run into a contradiction!

However, again there is a subtlety. In constructing the black hole solution, one uses the low energy effective field theory which is valid only when the curvature is much smaller than the string scale. But since this black hole has vanishing area of the event horizon, it follows that the solution actually has a curvature singularity at the horizon.[31] Thus we would expect that the solution near the horizon will be modified by the higher derivative terms in the string effective action, and hence might give a different answer for the area of the event horizon.

In order to estimate what the modified area will be, one needs to understand what kind of corrections we must include in the effective action. Typically in string theory there are two types of correction − the ones due to string world-sheet effects (the higher

[30] When we regard elementary strings as classical black hole solutions, the field λ_2 varies as a function of the radial distance from the origin. Due to this fact we are specifically using the notation $\langle\lambda_2\rangle$ to denote the asymptotic value of λ_2, *i.e.* the expectation value of λ_2 in the vacuum.

[31] One might wonder whether such a solution can be called a black hole at all, but the reason that they are called black holes is that they can be obtained as a limit of black hole solutions with non-singular horizons with finite area. As the mass of the black hole approaches the Bogomol'nyi bound (9.6), we get the singular black hole.

derivative terms), and the ones due to the string loop effects. In the present case when one examines the classical black hole solution one finds that the field λ_2 approaches ∞ as we approach the event horizon. Thus the string coupling $\lambda_2^{-1/2}$ vanishes in this region and we do not expect the string loop corrections to be significant. This leaves us with the string world-sheet corrections. Although we cannot explicitly calculate the effect of these corrections, we can use a scaling argument to determine the form of these corrections up to an overall numerical factor. The argument goes as follows. If we use the string metric (as opposed to the canonical metric) to describe the black hole solution, then one finds that with a suitable choice of coordinate system the solution near the origin becomes completely independent of all parameters m, $\vec{k}_L$ and $\langle\lambda_2\rangle$, except for an additive factor of

$$-\ln\left(\sqrt{\langle\lambda_2\rangle}\sqrt{m^2-\frac{2\vec{k}_L^2}{\langle\lambda_2\rangle}}\right), \tag{9.9}$$

in the expression for the dilaton. This additive constant does not affect the string world-sheet lagrangian. Thus whatever be the effect of the corrections due to the string world-sheet effects, these corrections are universal, and do not depend on any parameters. Thus after taking into account the string world-sheet corrections, the area of the event horizon, as measured in the string metric, will be a universal numerical constant, independent of all external parameters. However since the Bekenstein-Hawking entropy is to be identified with the area of the event horizon measured in the *canonical metric*, which differs from the string metric by a factor of $e^{-\Phi}$, and since Φ has an additive factor (9.9) that depends on the parameters, the modified Bekenstein-Hawking entropy will be given by,

$$S_{BH}=C\sqrt{\langle\lambda_2\rangle}\sqrt{m^2-\frac{2\vec{k}_L^2}{\langle\lambda_2\rangle}}, \tag{9.10}$$

where C is an unknown numerical constant. This is in complete agreement with the answer for S_{micro} given in (9.8) if we choose

$$C=2\pi\,. \tag{9.11}$$

Note, in particular, that S_{micro} and S_{BH} have the same functional dependence on m, $\vec{k}_L$ and $\langle\lambda_2\rangle$. Considering the fact that these are all dimensionless parameters (we are working in units $\hbar=1, c=1, \alpha'=1$) this agreement is impressive. This calculation can also be extended to other toroidal compactification of heterotic string theory[134], and to non-toroidal compactification of heterotic and type II string theories[135].

9.3 Black holes and D-branes

Although the result described in the previous subsection is encouraging, we would like to do better and compare S_{BH} and S_{micro} without encountering any undetermined numerical factor. The strategy is to try to identify black hole solutions which are

- BPS states, and
- have non-vanishing area of the event horizon even without stringy corrections.

It turns out that there are indeed such black holes present in the theory, but they do not carry the same quantum numbers as elementary string states. Instead they carry the same quantum numbers as a configuration of D-branes. Thus in order to calculate the microscopic entropy we need to calculate the degeneracy of this D-brane configuration. As has already been discussed earlier, the dynamics of collective coordinates of D-branes is given by the massless open string states propagating on the D-branes. Thus we can explicitly determine the Hamiltonian describing this dynamics and calculate the degeneracy of states to calculate S_{micro}. This is precisely what is done.

Thus our strategy is as follows:

1. Identify a BPS black hole with non-vanishing area of the event horizon and calculate S_{BH} from this area.
2. Identify the D-brane configuration carrying the same quantum numbers as this black hole and calculate S_{micro} by computing the degeneracy of these states.
3. Compare the two answers.

The analysis is simplest in five dimensions, so we shall concentrate on this case[136, 18]. We focus on type IIB string theory on T^5. The D-brane configuration that we consider has

- Q_5 D-5-branes wrapped on T^5,
- Q_1 D-1-branes wrapped on one of the circles S^1 of T^5, and
- $-n$ units of momentum along S^1. If R denotes the radius of S^1, this corresponds to a momentum of $-n/R$.

The counting of states for this D-brane system can be done as follows. The world-volume theory of a system of Q_5 parallel D-5 branes is a supersymmeric $U(Q_5)$ gauge theory in $(5+1)$ dimensions, obtained by the dimensional reduction of N=1 supersymmetric $U(Q_5)$ gauge theory in (9+1) dimensions. It can be shown that a single D-1 brane inside Q_5 coincident D-5 branes can be identified to a single instanton in this $U(Q_5)$ gauge theory[154].[32] Thus a system of Q_1 parallel D-1 branes inside Q_5 coincident D-5 branes can be described as a system of Q_1 instantons in the $U(Q_5)$ gauge theory. The moduli of this Q_1 instanton solution act as collective coordinates of this system. These moduli span a $4Q_1Q_5$ dimensional hyper-kahler manifold. As a result the low energy dynamics of a system of Q_1 D-1 branes inside Q_5 D-5 branes is described by a (1+1) dimensional supersymmetric σ-model with this instanton moduli space as the target space. Since this space is hyper-kahler, the corresponding supersymmetric σ-model is conformally invariant[155]. Furthermore, since the moduli space has dimension $4Q_1Q_5$, the central charge is given by:

$$c = \frac{3}{2} \cdot 4Q_1Q_5 = 6Q_1Q_5 \, , \tag{9.12}$$

taking into account the contribution of 1 from each scalar and (1/2) from each fermion. The BPS states in this theory correspond to states with $L_0 = 0$. $L_0 - \bar{L}_0$ represents the total number of momentum units carried by the system along S^1; for this system this is equal to $-n$. For large n the degeneracy of such states can be computed[199]. The answer is

$$d(Q_1, Q_5, n) \sim \exp\Big(2\pi\sqrt{\frac{cn}{6}}\Big) = \exp(2\pi\sqrt{Q_1Q_5n}) \, . \tag{9.13}$$

Since this argument is somewhat abstract, we shall now give a simplified description of the counting of states of this system[137, 70, 138, 156]. First consider the case when there is one D-5 brane and Q_1 D-1 branes inside the D-5 brane (so that the D-1 branes are free to move inside the D5-brane but not free to leave the D-5 brane). Now, since Q_1 D-1 branes, each wrapped once around S^1, has the same charge as a single D-1 brane wrapped Q_1 times, we must also include this configuration in our counting of states. It turns out that the contribution to the total degeneracy is dominated by the later configuration, so we can restrict our attention to this configuration. Now consider the case when there are Q_5 D-5 branes instead of just one. Again the dominant contribution comes from the

[32] An instanton is a classical solution in Yang-Mills theory in four euclidean dimensions. Thus in the (5+1) dimensional gauge theory it represents a solution that is independent of time and one spatial directions, *i.e.* a static string.

configuration where instead of Q_5 D-5 branes each wrapped once around $S^1 \times T^4$, we have a single D-5 brane wrapped Q_5 times. Thus we have a configuration of a single D1 brane of length $2\pi R Q_1$ and a single D-5 brane of length $2\pi R Q_5$ along the direction of S^1. We shall call these branes long D-1 and long D-5 branes respectively. Now take Q_1 and Q_5 to be relatively prime. In that case, when we go around the long D-1 brane once by travelling Q_1 times around the circle S^1, we do not come back to the same point on the long D-5 brane. On the other hand, if we go around the long D-5 brane once by travelling Q_5 times around S^1, we do not come back to the same point on the long D-1 brane. In fact, we need to go around the long D-1 brane Q_5 times in order to come back to the same point on the long D-5 brane and the long D-1 brane. This amounts to going around S^1 $Q_1 Q_5$ times. Thus to an open string stretched between the D-1 brane and the D-5 brane the configuration will appear to be that of a single D-string of length $2\pi R Q_1 Q_5$, which is free to move in the four transverse directions inside a D-5 brane[138]. This gives four bosonic collective coordinates X^i. Due to supersymmetry, this system will also have four Majorana fermions λ^i moving on the D-string world-sheet.

For each of these four bosonic and fermionic coordinates there are left moving modes as well as right moving modes on the D-1 brane. A quantum of the mth left (right) moving mode carries $-m$ (m) units of momentum along S^1, with each unit of momentum now being equal to $1/(RQ_1Q_5)$. We need a state with a total of $-nQ_1Q_5$ units of momentum. It turns out that in order to saturate the BPS mass formula, we need to concentrate on states containing only quanta of left moving modes and no quanta of right moving modes. If there are N^i_m quanta of the mth left moving mode of X^i, and n^i_m quanta of the mth left moving mode of λ^i, then in order to get $-Q_1Q_5n$ units of momentum along S^1 we need:

$$Q_1 Q_5 n = \sum_{i=1}^{4} \sum_{m=1}^{\infty} m(N^i_m + n^i_m)\,. \tag{9.14}$$

The degeneracy $d(Q_1, Q_5, n)$ is the number of ways we can choose integers N^i_m and n^i_m satisfying the above relation. This can be computed using standard procedure, and the answer is

$$d(Q_1, Q_5, n) \sim \exp(2\pi\sqrt{Q_1 Q_5 n})\,, \tag{9.15}$$

which is the same as (9.13). This gives,

$$S_{micro} = \ln d \simeq 2\pi\sqrt{Q_1 Q_5 n}\,. \tag{9.16}$$

We now need to compare this result with the Bekenstein-Hawking entropy of the black hole solution of the low energy effective field theory carrying the same set of charges and the same mass. In the normalization convention of eqs.(3.1), (3.2) and (3.33) the five dimensional Newton's constant is given by,

$$G_N^5 = \frac{\pi}{4}\,. \tag{9.17}$$

Instead of writing down the full black hole solution, we shall only write down the canonical metric for this black hole in the five non-compact directions, since this is what is required to compute the area of the event horizon, and hence S_{BH}. The metric is[18, 19][33]

$$ds^2 = -\lambda^{-2/3}dt^2 + \lambda^{1/3}[dr^2 + r^2 d\Omega_3^2]\,, \tag{9.18}$$

where,

$$\lambda = (1 + r_1^2/r^2)(1 + r_5^2/r^2)(1 + r_n^2/r^2)\,, \tag{9.19}$$

$$r_1^2 = (RV)^{2/3} g^{-1/2} Q_1/V, \qquad r_5^2 = (RV)^{2/3} g^{1/2} Q_5, \quad r_n^2 = (RV)^{2/3} n/R^2 V\,. \tag{9.20}$$

$d\Omega_3$ denotes line element on a unit three sphere, $(2\pi)^4 V$ is the volume of T^4 and R is the radius of S^1, both measured in the ten dimensional canonical metric, and $g(\equiv e^{\langle\Phi^{(10)}\rangle/2})$ is the string coupling constant *in ten dimension*. Here T^4 denotes the subspace of the full T^5 that does not include the special S^1 on which the D-1 branes are wrapped. The event horizon is located at $r = 0$, and the area A of the even horizon can be easily computed from eqs.(9.18), (9.19) to be $2\pi^2 r_1 r_5 r_n$. This gives,

$$S_{BH} = \frac{A}{4G_N^5} = (2\pi)\sqrt{Q_1 Q_5 n}\,. \tag{9.21}$$

This is in exact agreement with S_{micro} computed in eq.(9.16).

Similar agreement between the Bekenstein-Hawking entropy and the microscopic entropy has been demonstrated for black holes carrying angular momentum[144], and also for black holes in four dimensions[145]. The analysis can also be easily extended to compactification of type II string theory on $K3 \times S^1$ with sixteen supersymmetry charges. More non-trivial case is the extension to black holes in theories with eight supercharges. This has been done in many cases, and in every case that has been studied, the Bekenstein-Hawking entropy of the BPS saturated black hole agrees with the microscopic entropy

[33] Our ten dimensional metric, in which R and V are measured, differs from that of [18, 19] by a factor of $g^{-1/2}$, and our five dimensional metric differs from that of [18, 19] by a factor of $(RV)^{2/3} g^{-1/2}$.

computed from analyzing the dynamics of the brane configuration[143]. An alternate approach to calculating S_{micro} has been advocated in [149] that also gives answer in agreement with S_{BH}.

Given the success of the D-brane dynamics in giving a microscopic description of the entropy of an extremal black hole, one might wonder if a similar analysis can be carried out for black holes which are not extremal, but are nearly extremal[137, 139]. For the D-brane configuration such states are obtained by relaxing the requirement that there is no quanta of right-moving modes. However, we restrict the number of such excitations so that the interaction between the left- and the right-moving modes can be neglected; this is known as the dilute gas approximation. Let the left moving modes carry a total momentum $-N_L/(Q_1 Q_5 R)$ along S^1 and the right moving modes carry a total momentum $N_R/(Q_1 Q_5 R)$ along S^1. Then we require

$$N_L - N_R = Q_1 Q_5 n \, . \tag{9.22}$$

The degeneracy of states is obtained by computing the number of ways these momenta can be distributed between different left- and the right-moving modes. The microscopic entropy, computed this way, is given by

$$S_{micro} = 2\pi(\sqrt{N_L} + \sqrt{N_R}) \, . \tag{9.23}$$

Since for this configuration the mass is no longer given by the BPS formula, the black hole solution also gets modified and gives a new S_{BH}. The answer turns out to be

$$S_{BH} = 2\pi(\sqrt{N_L} + \sqrt{N_R}) \, , \tag{9.24}$$

again in perfect agreement with (9.23). In the expression for S_{BH} the combination $N_L - N_R$ enters through the dependence of the solution on the various charges, whereas the combination $N_L + N_R$ enters through the dependence of the solution on the mass of the black hole which is now an independent parameter.

A priori we should not have expected such an agreement between S_{BH} and S_{micro} for these black holes, since for non-BPS states one did not expect any non-renormalization theorem to hold and there was no reason why the two answers computed in different domains of validity should agree. An explanation for this agreement was provided later with the help of a new non-renormalization theorem[140].

Having found agreement between S_{BH} and S_{micro}, one can now ask if we can also reproduce Hawking radiation from these black holes from the dynamics of D-branes[141, 121, 142]. It turns out that extremal black holes have zero temperature and hence do not Hawking radiate. This is consistent with the fact that in the microscopic description, BPS states are stable and hence cannot decay into other states. But non-extremal black holes do Hawking radiate and we can try to compare this radiation with the radiation due to the decay of a non-BPS D-brane. Computation of the Hawking radiation rate from the near extremal black hole can be done by standard technique. One subtlety comes from the fact that although the black hole horizon gives out thermal radiation, there is a frequency dependent filtering of this radiation as it passes through the black hole background and reaches the asymptotic observer. This effect is known as the grey body factor, and can be computed by knowing the background fields associated with the specific black hole solution under consideration. For the specific non-extremal black holes that we are considering, the net Hawking radiation of a specific class of scalar particles, as seen by an asymptotic observer is given by[142],[34]

$$\Gamma_H \;=\; 2\pi^2 (RV)^{4/3} \frac{Q_1 Q_5}{V} \frac{\pi k_0}{2} \frac{d^4 k}{(2\pi)^4} \frac{1}{e^{\frac{k_0}{2T_L}} - 1} \frac{1}{e^{\frac{k_0}{2T_R}} - 1} \tag{9.25}$$

where

$$\begin{aligned}(T_R) &= \frac{(RV)^{-1/3}}{\pi R Q_1 Q_5}\sqrt{N_R} \\ (T_L) &= \frac{(RV)^{-1/3}}{\pi R Q_1 Q_5}\sqrt{N_L}\,,\end{aligned} \tag{9.26}$$

and d^4k denotes the four dimensional phase-space in this $(4+1)$ dimensional theory.

In the D-brane description, the radiation of these specific scalar particles is due to the annihilation of the left and right moving modes on the D-1 brane. This decay rate can be calculated by using standard string theoretic technique, and the answer is[121],

$$\Gamma_{micro} \;=\; 2\pi^2 (RV)^{4/3} \frac{Q_1 Q_5}{V} \frac{\pi k_0}{2} \frac{d^4 k}{(2\pi)^4} \frac{1}{e^{\frac{k_0}{2T_L}} - 1} \frac{1}{e^{\frac{k_0}{2T_R}} - 1}\,. \tag{9.27}$$

Again we see that this is in exact agreement with Γ_H!

[34] Again in comparing this expression to those in refs.[18, 19] we need to take into account the rescaling of the metric.

Thus we see that at least for a class of black holes we now have a concrete microscopic derivation of the Bekenstein-Hawking entropy and the phenomenon of Hawking radiation. The description is completely quantum mechanical. The only caveat is that these microscopic calculations are done in regions of parameter space where the system is not a black hole, and then, with the help of non-renormalization theorems, the answer is continued to the region where the system is a black hole. If we could understand in more detail how this continuation works, then perhaps we shall be able to show that there is really no conflict between Hawking radiation and quantum mechanics, in the same sense that thermal radiation from a star is not in conflict with quantum mechanics.

10 Matrix Theory

In section 7 we postulated the existence of an eleven dimensional theory, known as M-theory, with the following two properties:

- The low energy limit of M-theory is eleven dimensional N=1 supergravity.
- M-theory compactified on a circle is dual to type IIA string theory. As the radius of this circle goes to infinity, the type IIA coupling constant also goes to infinity.

From this we can define M-theory as the strong coupling limit of type IIA string theory. However this definition does not give us any clue as to how to systematically compute S-matrix elements in M-theory, since the type IIA string theory is defined only by the rules for its perturbation expansion in the coupling constant. Thus finding a non-perturbative definition of M-theory is of importance. One might try various approaches:

1. M-theory is the N=1 supergravity theory in D=11: This proposal by itself does not make sense since this supergravity theory is not renormalizable and hence is plagued by the usual ultra-violet divergences. One might argue that there is some intrinsic regularization that accompanies this supergravity theory, but till we find such a regularisation, describing M-theory as the supergravity theory remains an empty statement.
2. M-theory is a string theory: One might imagine that M-theory can be made finite by regarding it as the low energy limit of some string theory in the same way that the various supergravity theories in ten dimensions are made finite. However

nobody has been able to find any such string theory so far. There is also another compelling reason to believe that M-theory is not a string theory. With the sole exception of type I string theory, every other string theory has the property that the corresponding low energy supergravity theory contains the fundamental string as a supersymmetric soliton. N=1 supergravity in eleven dimensions does not contain any such soliton solution.

3. M-theory is a theory of membranes or five-branes: The eleven dimensional supergravity theory does contain membrane and five-brane like soliton solutions[8]. Thus one might argue that M-theory should be formulated as a theory of membranes or five-branes. The difficulty with this proposal is that unlike string theory, which is tractable due to the infinite dimensional conformal symmetry on its world-sheet, the world-volume theory of membranes or five-branes have no such infinite dimensional symmetry and hence are extremely difficult to handle. However, the Matrix theory that we are about to discuss does in some sense regard M-theory as a theory of membranes[157].

These difficulties in formulating M-theory has led to a radically new way of thinking about M-theory, and in fact string theories in general. This proposal, known as Matrix theory[158, 159], is based on describing the theory in terms of its Hamiltonian in the discrete light cone quantization (DLCQ). In this section we shall give a brief overview of this formulation. The contents of this section will form a small fraction of the material covered in [21, 22].

10.1 Discrete light-cone quantization (DLCQ)

We shall illustrate the procedure of DLCQ in the context of a scalar field theory. Let us begin with a free scalar field theory in $d+1$ dimension. The action of the system can be expressed as

$$S = \int dx^+ dx^- d^{d-1}\vec{x}_\perp (\partial_+\phi\partial_-\phi - \frac{1}{2}\sum_{i=1}^{d-1}\partial_i\phi\partial_i\phi)\,, \tag{10.1}$$

where

$$x^\pm = \frac{1}{\sqrt{2}}(x^0 \pm x^d)\,, \qquad \partial_\pm = \frac{\partial}{\partial x^\pm}\,, \tag{10.2}$$

and $\vec{x}_\perp \equiv (x^1, \ldots x^{d-1})$ denotes the transverse coordinates. We shall now canonically quantize this theory by regarding x^+ as time. Thus the momentum conjugate to ϕ is given by,

$$\pi = \frac{\delta S}{\delta(\partial_+\phi(x))} = \partial_-\phi\,. \tag{10.3}$$

Note that the relationship between ϕ and π given in eq.(10.3) does not involve a 'time' derivative, reflecting the fact that the original lagrangian is linear in $\partial_+\phi$. As a result eq.(10.3) represents a constraint. The canonical equal time commutation relations can be found by using the standard formalism for quantization of constrained system. However, we can simplify this problem by going to the Fourier transformed variables defined through the decomposition:

$$\phi(x^+, x^-, \vec{x}_\perp) \;=\; \int_0^\infty \frac{dk_-}{\sqrt{4\pi k_-}}[a(x^+, k_-, \vec{x}_\perp)e^{-ik_-x^-} + a^*(x^+, k_-, \vec{x}_\perp)e^{ik_-x^-}] \tag{10.4}$$

The action (10.1) now takes the form:

$$S = i\int dx^+ \int_0^\infty dk^- \int d^{d-1}\vec{x}_\perp [a^*(x^+, k_-, \vec{x}_\perp)\partial_+ a(x^+, k_-, \vec{x}_\perp) + \ldots]\,, \tag{10.5}$$

where $\ldots$ denote terms without any x^+ derivatives. Thus if we now regard a as the coordinate, its canonically conjugate momentum is given by,

$$\frac{\delta S}{\delta(\partial_+ a)} = ia^* \tag{10.6}$$

Upon going to the quantum theory, a^* should be regarded as the hermitian conjugate of the operator a. This gives the following equal time commutation rules:

$$\begin{aligned}
&[a(x^+, k_-, \vec{x}_\perp), a(x^+, l_-, \vec{y}_\perp)] = 0 = [a^\dagger(x^+, k_-, \vec{x}_\perp), a^\dagger(x^+, l_-, \vec{y}_\perp)] \\
&[a(x^+, k_-, \vec{x}_\perp), a^\dagger(x^+, l_-, \vec{y}_\perp)] = \delta(k_- - l_-)\delta(\vec{x}_\perp - \vec{y}_\perp)\,.
\end{aligned} \tag{10.7}$$

From this we see that $a^\dagger$ and a behave respectively as creation and annihilation operators for particles carrying momentum k_- located at the transverse location $\vec{x}_\perp$. Note that the argument k_- of a and $a^\dagger$ extends over positive values only. This can be traced to the fact that (10.3) is a constraint equation, hence only half of the degrees of freedom of ϕ are true coordinates, the other half being momenta.

We shall now compactify the light-like direction x^- on a circle of radius L. This of course is not a physically meaningful system since it has closed light-like curves and hence violates causality, but we simply use it as an infrared regulator and at the end take the

$L \to \infty$ limit to recover physically meaningful answers for various processes. In this case the momentum k_- is quantized as:

$$k_- = \frac{n}{L}\,. \tag{10.8}$$

Eqs.(10.4) and(10.5) are now modified as:

$$\begin{aligned}\phi(x^+, x^-, \vec{x}_\perp) &= a_0(x^+, \vec{x}_\perp) + \frac{1}{\sqrt{4\pi}} \sum_{n=1}^{\infty} \frac{1}{\sqrt{n}} \Big[a_n(x^+, \vec{x}_\perp) e^{-inx^-/L} \\ &\quad + a_n^*(x^+, \vec{x}_\perp) e^{inx^-/L} \Big]\end{aligned} \tag{10.9}$$

$$S = i \int dx^+ \int d^{d-1}\vec{x}_\perp \sum_{n=1}^{\infty} (a_n^* \partial_+ a_n + \ldots)\,, \tag{10.10}$$

This shows that ia_n^* is the momentum conjugate to a_n. In quantum theory a_n^* will be represented by the hermitian conjugate $a_n^\dagger$ of the operator a_n. This gives the following commutation relations:

$$\begin{aligned}&[a_n(x^+, \vec{x}_\perp), a_m(x^+, \vec{y}_\perp)] = 0 = [a_n^\dagger(x^+, \vec{x}_\perp), a_m^\dagger(x^+, \vec{y}_\perp)]\,, \\ &[a_n(x^+, \vec{x}_\perp), a_m^\dagger(x^+, \vec{y}_\perp)] = \delta_{mn} \delta(\vec{x}_\perp - \vec{y}_\perp)\,.\end{aligned} \tag{10.11}$$

Note that the action does not contain any term containing x^+ derivative of a_0. Thus we can interprete it as the Lagrange multiplier field and integrate it out. The hamiltonian computed from the action (10.1) takes the form:

$$H = \sum_{n=1}^{\infty} \frac{L}{2n} \int d^{d-1}\vec{x}_\perp \partial_i a_n^\dagger(x^+, k_-, \vec{x}_\perp) \partial_i a_n(x^+, k_-, \vec{x}_\perp)\,. \tag{10.12}$$

We can define the ground state of this system as the state annihilated by all the a_n's. The p-particle state is created by acting with p of the $a^\dagger$'s on this ground state.

Let us now consider the effect of introducing interactions in the original action (10.1). This will add new terms to the hamiltonian. In particular, since the interaction terms in the lagrangian will involve a_0, integrating out a_0 will in general produce infinite series of additional terms in H even if the lagrangian itself contains only a finite number of terms.[35] But as long as the interactions are invariant under translation, all the terms in

[35] If the theory contains moduli fields which have flat potential and hence can acquire arbitrary vacuum expectation value, then the process of integtrating out the zero modes of these fields requires choosing a definite set of vacuum expectation values of these fields. This corresponds to choosing a specific point in the moduli space. At different points of the moduli space we shall get different DLCQ theory. Thus the DLCQ formalism is not independent of the choice of background.

the Hamiltonian will conserve k_-. Let us now focus on a sector with $k_- = N/L$. This sector contains N-particle states with each particle carrying momentum $1/L$, but it may also contain states with less than N particles, with some particles carrying momentum larger than $1/L$. However, this sector does not cantain states with more than N particles, since all particles in this description carry strictly positive k_- in units of $1/L$. This suggests that the dynamics in this sector should be describable by an N-body *quantum mechanical hamiltonian* $\mathcal{H}_N^{DLCQ}$. (One should distinguish this from the hamiltonian H of the second quantized theory.) $\mathcal{H}_N^{DLCQ}$, by definition, describes the complete dynamics in the $k_- = N/L$ sector. Thus for example it should correctly reproduce the spectrum, as well as all scattering amplitudes in this sector. In particular, in this description a single particle state carrying N units of momentum should appear as a bound state of N particles in the spectrum of $\mathcal{H}_N^{DLCQ}$. Thus $\mathcal{H}_N^{DLCQ}$ must possess an appropriate bound state of this form. Similarly it must contain in its spectrum appropriate bound states representing p particle state carrying total momentum N/L for all $p < N$, and for all possible distribution of the minus component of the total momentum N/L among these p particles.

Given the action of an interacting quantum field theory, it should in principle be possible to construct $\mathcal{H}_N^{DLCQ}$. Conversely, if we are given $\mathcal{H}_N^{DLCQ}$ for all N, we can, in principle, completely reconstruct the spectrum and the S-matrix elements of the theory. For example, calculation of a specific n particle $\to$ m particle process will involve the following steps:

1. Calculate the scattering amplitude for a set of n states carrying momenta $k_-^i = n_i/L$ going to a set of m states carrying momenta $l_-^i = m_i/L$ with
$$\sum_{i=1}^{n} n_i = \sum_{i=1}^{m} m_i = N\,. \qquad (10.13)$$
2. Now take the limit $N \to \infty$, $L \to \infty$, $n_i \to \infty$ and $m_i \to \infty$ keeping fixed
$$N/L, \qquad k_-^i = n_i/L, \qquad l_-^i = m_i/L\,. \qquad (10.14)$$

In this limit we shall reproduce the physical scattering amplitude involving $(m + n)$ external legs. We are of course implicitly assuming that there is an unambiguous procedure for taking the large N limit.

Given that $\mathcal{H}_N^{DLCQ}$ defines a theory, we can now try to define M-theory by specifying $\mathcal{H}_N^{DLCQ}$ for M-theory. This is what is done in the Matrix-theory approach to M-theory.

10.2 DLCQ of M-theory

Although DLCQ is used for giving a non-perturbative definition of M-theory, it can in principle be used to describe any theory. We shall first give a general recipe for constructing $\mathcal{H}_N^{DLCQ}$ for any (compactified) string theory or M-theory, and then specialize to the case of M-theory[160, 161]. Let $\mathcal{T}$ be such a (compactified) string or M-theory. Typically $\mathcal{T}$ has one mass parameter M, which can be taken to be the Planck mass for M-theory, and $(\sqrt{\alpha'})^{-1}$ for string theory, and a set of dimensionless parameters, *e.g.* string coupling constant, the dimensions of the compact manifold measured in units of M^{-1}, etc. We shall label all these dimensionless parameters by $\{\vec{y}\}$. (From now on we shall display all factors of the mass parameter explicitly, and not work in the $\alpha' = 1$ unit, as we have been doing till now.) Let us denote by $\mathcal{H}_N^{DLCQ}(M, L, \{\vec{y}\})$ the DLCQ hamiltonian for this system. L as usual denotes the radius of the compact light-cone direction. We now propose the following recipe for constructing $\mathcal{H}_N^{DLCQ}$[160, 161].

1. Consider the same theory $\mathcal{T}$ with the same values of the dimensionless parameters $\{\vec{y}\}$, but a different value m of the mass parameter. Let us compactify this theory on a space-like circle S^1 of radius R. We shall call this theory the auxiliary theory.

2. When R is small, the Kaluza-Klein modes carrying momentum along S^1 are heavy, and one would expect that the dynamics of this system will be described by a non-relativistic quantum mechanical hamiltonian. Let us focus on the sector with N-particles each carrying momentum $1/R$ along S^1, and let us denote the N-body hamiltonian describing the dynamics of this system by $\mathcal{H}_N^{KK}(m, R, \{\vec{y}\})$. (We subtract off the rest mass energy N/R of these particles in defining $\mathcal{H}_N^{KK}$.)

3. $\mathcal{H}_N^{DLCQ}$ is constructed from $\mathcal{H}_N^{KK}$ by taking the limit

$$\mathcal{H}_N^{DLCQ}(M, L, \{\vec{y}\}) = \lim_{R\to 0} \mathcal{H}_N^{KK}(m = M\sqrt{L/R}, R, \{\vec{y}\}) \,. \tag{10.15}$$

We shall first explore some of the consequences of this recipe. Later we shall discuss how this recipe might be 'derived'.

First of all, note that since the duality transformations in $\mathcal{T}$ (before compactification on S^1) leave the momenta along the non-compact directions unchanged, it leaves the sector with N units of momentum along S^1 invariant after we compactify $\mathcal{T}$ on S^1. As a result $\mathcal{H}_N^{KK}$, and hence $\mathcal{H}_N^{DLCQ}$ defined in (10.15) is expected to possess the full set of duality symmetries of $\mathcal{T}$[162]-[171].

Second, note that although this recipe gives $\mathcal{H}_N^{DLCQ}$ for any theory $\mathcal{T}$, it is particularly useful for (compactified) M-theory, since in the $R \to 0$ limit M-theory on a circle of radius R gets mapped to *weakly coupled* type IIA string theory as discussed in section 7. As discussed there, states carrying momenta along S^1 correspond to D0 branes in the type IIA theory. Thus the recipe given above relates $\mathcal{H}_N^{DLCQ}$ to a specific weak coupling limit of the hamiltonian describing N D0 branes in type IIA string theory. We shall later construct this hamiltonian explicitly in some cases.

One might feel that the assertion made above is a bit too simplified, since D0-branes represent only a subset of particles in M-theory carrying momenta along S^1 – the eleven dimensional graviton and its supersymmetric partners. For example, if $\mathcal{T}$ corresponds to M-theory compactified on T^2, then $\mathcal{T}$ has solitonic states corresponding to the membrane of M-theory wrapped around T^2. These states are distinct from the supergravitons. Thus one would expect that in constructing $\mathcal{H}_N^{KK}$ (and hence $\mathcal{H}_N^{DLCQ}$) for this theory one needs to add to the D0-brane hamiltonian new degrees of freedom which are capable of describing these wrapped membrane states carrying momentum along S^1. However, due to a truly marvellous property of the D0-brane system, this is not necessary. It turns out that the D0-brane hamiltonian automatically contains the required degrees of freedom that gives rise to these new states. We shall explicitly see an example of this later.

Let us now apply this recipe to construct $\mathcal{H}_N^{DLCQ}$ for M-theory on T^n. Let M_p be the Planck mass of the M-theory, L_i be the radii of the n circles[36] which make up T^n and L be the radius of the light-like circle. Let m_p be the Planck mass of the auxiliary M-theory, R_i be the radii of the circles that make up the n-dimensional torus in this auxiliary M-theory, and R be the radius of the extra S^1 on which this auxiliary M-theory on T^n is further compactified. Then from (10.15) we get the following relation between the parameters of the two theories:

$$m_p = M_p\sqrt{L/R}, \qquad m_p R_i = M_p L_i, \tag{10.16}$$

where the second equation reflects that the dimensionless parameters obtained by taking the product of the Planck mass and the radii of the compact directions must be the same in the two theories. We now map this to the hamiltonian of a set of D0 branes in type

[36] For simplicity we are assuming that the torus is made of product of n circles, without any background field.

IIA string theory by identifying M-theory on S^1 of radius R with type IIA string theory according to the rules given in section 7. Let $g_S(\equiv e^{\Phi/2})$ and $m_S(\equiv (\alpha')^{-1/2})$ denote the coupling constant and the string mass in this type IIA theory. Then the first of eq.(7.2) gives

$$m_p R = g_S^{2/3}\,, \tag{10.17}$$

where we have explicitly put in the eleven dimensional Planck mass that was set to unity in (7.2). Another relation between m_p, m_S and g_S comes from restoring the appropriate factors of m_p, m_S and g_S in (7.1) and (3.33). In particular (7.1) contains a multiplicative factor of $(m_p)^9$, whereas (3.33) contains a multiplicative factor of $(\alpha')^{-4}g_S^{-2} = m_S^8 g_S^{-2}$ (analog of eq.(3.3) for the heterotic theory). Thus the equality between the two actions upon compactification of the eleven dimensional theory on S^1 requires that:

$$m_p^9 R = m_S^8 g_S^{-2}\,. \tag{10.18}$$

Inverting the relations (10.17) and (10.18) we get,

$$m_S = m_p^{3/2} R^{1/2}, \qquad g_S = (m_p R)^{3/2}\,. \tag{10.19}$$

If $\mathcal{H}_N^{D0}$ denotes the hamiltonian for N D0 branes (with the rest mass subtracted) in this theory, then, using eqs.(10.15), (10.16) and (10.19) we get,

$$\begin{aligned} & \mathcal{H}_N^{DLCQ}(M_p, L, \{R_i\}) \\ = \ & \lim_{R\to 0} \mathcal{H}_N^{D0}(m_S = M_p^{3/2} L^{3/4} R^{-1/4}, g_S = M_p^{3/2}(LR)^{3/4}, R_i = R^{1/2} L^{-1/2} L_i)\,. \end{aligned} \tag{10.20}$$

This gives sensible answer for all n up to 5 and does not give sensible answer for $n \geq 6$[160, 161]. Here we shall only discuss two special cases, $n = 0$ and $n = 2$.

First we consider the case $n = 0$. This corresponds to eleven dimensional M-theory. As seen from (10.20), as $R \to 0$, $g_S \to 0$ and $m_S \to \infty$. Thus we can use low energy and weak coupling approximation of type IIA string theory. This limit has been studied in detail in [172]. The action governing the dynamics of the D0 brane system in this limit is given by the dimensional reduction of $N = 1$ supersymmetric U(N) gauge theory in ten dimensions:

$$S \;\sim\; m_S^{-3} g_S^{-1} \int dt \Big[\sum_{m=1}^{9} Tr(\partial_t \Phi^m \partial_t \Phi^m) - \sum_{m<n=1}^{9} Tr([\Phi^m, \Phi^n]^2) \Big]$$

$$
\begin{aligned}
& +\text{fermionic terms} \\
= \; & M_p^{-6} L^{-3} \int dt \Big[\sum_{m=1}^{9} Tr(\partial_t \Phi^m \partial_t \Phi^m) - \sum_{m<n=1}^{9} Tr([\Phi^m, \Phi^n]^2) \Big] \\
& +\text{fermionic terms}
\end{aligned}
\tag{10.21}
$$

where Φ^m are $(N \times N)$ hermitian matrices. In going from the first to the second line we have used eq.(10.20). This gives a well-defined hamiltonian for DLCQ M-theory. The flat direction in the potential corresponds to a configuration where all the Φ^m's are simultaneously diagonalized. The N eigenvalues of Φ^m represent the m'th coordinate of the N different D0 branes.

Let us now consider M-theory on T^2. From eq.(10.20) we see that in the $R \to 0$ limit, the radii R_i vanish. We can remedy this problem by giving a different description of the same system by making an $R_i \to (1/m_S^2 R_i)$ duality transformation on both circles. This converts the original type IIA theory to type IIA theory on a dual torus $\widetilde{T}^2$ and the system of N D0 branes to a system of N D-2 branes wrapped on $\widetilde{T}^2$.[37] The new theory has parameters:

$$
\begin{aligned}
& \widetilde{g}_S = g_S/(m_S^2 R_1 R_2) = M_p^{-3/2} L^{1/4} R^{1/4} L_1^{-1} L_2^{-1}, \\
& \widetilde{m}_S = m_S = M_p^{3/2} L^{3/4} R^{-1/4}, \qquad \widetilde{R}_i = R_i^{-1} m_S^{-2} = M_p^{-3} L^{-1} L_i^{-1} \,.
\end{aligned}
\tag{10.22}
$$

In the $R \to 0$ limit this again gives a theory at weak coupling and large string mass. Furthermore the new radii $\widetilde{R}_i$ are finite. The dynamics of wrapped D-2 branes in this limit is described by a (2+1) dimensional $N = 8$ supersymmetric U(N) Yang-Mills theory compactified on $\widetilde{T}^2$. This theory has seven scalars Φ^m $(1 \le m \le 7)$ in the adjoint representation of the gauge group. The bosonic part of the lagrangian is given by:

$$
\begin{aligned}
S \;\sim\; & \widetilde{m}_S^{-1} \widetilde{g}_S^{-1} \int dt \int_0^{\widetilde{R}_1} dx^8 \int_0^{\widetilde{R}_2} dx^9 \Big[Tr(F_{\mu\nu} F^{\mu\nu}) \\
& + Tr(D_\mu \Phi^m D^\mu \Phi^m) - \sum_{m<n=1}^{7} Tr([\Phi^m, \Phi^n]^2) \Big]
\end{aligned}
$$

[37] Quite generally one can show that an $R_i \to 1/(m_S^2 R_i)$ duality transformation converts a Dirichlet boundary condition to Neumann boundary condition and vice versa[5]. Thus this duality transformation converts a D0 brane with Dirichlet boundary conditions along T^2 into a D2 brane wrapped on $\widetilde{T}^2$ which has Neumann boundary condition along $\widetilde{T}^2$.

$$= \quad L^{-1}L_1L_2\int dt\int_0^{\widetilde{R}_1}dx^8\int_0^{\widetilde{R}_2}dx^9\Big[Tr(F_{\mu\nu}F^{\mu\nu})$$
$$+Tr(D_\mu\Phi^m D^\mu\Phi^m)-\sum_{m<n=1}^{7}Tr([\Phi^m,\Phi^n]^2)\Big]\,, \tag{10.23}$$

where x^μ $(\mu = 0, 8, 9)$ denote coordinates along the D-2 brane world-volume, and the N eigenvalues of Φ_m represent the mth transverse coordinate of the N different D-2 branes.

Let us now address the problem alluded to earlier, namely that M-theory on T^2 contains solitonic states in the form of wrapped membranes. Thus the complete $\mathcal{H}_N^{DLCQ}$ must contain these states as well. In the auxiliary type IIA theory, these wrapped membranes correspond to D2-branes wrapped on T^2. Upon T-duality in both circles, these become D0 branes moving on the dual torus $\widetilde{T}^2$. Thus the relevant question is, does the system described in (10.23) automatically contain these states, or do we need to add new degrees of freedom in this system so as to be able to describe these states? It turns out that the D0-brane charge in this dual theory simply corresponds to the flux of the U(1) component of the magnetic field through $\widetilde{T}^2$. Thus a state with k D0-branes (which correspond to a membrane wrapped k times on the original torus) can be described by a specific excitation of the system (10.23) carrying k units of magnetic flux through $\widetilde{T}^2$. There is no need to add new degrees of freedom.

Finally let us give a 'derivation' of the recipe described at the beginning of this section following [161]. First of all, we note that in the auxiliary theory $\mathcal{T}$ on S^1, if we multiply all the masses by some constant λ, and simultaneously multiply all the lengths by λ^{-1}, then the hamiltonian gets multiplied by λ due to purely dimensional reasons. This gives the following identity:

$$\mathcal{H}_N^{KK}(m,R,\{\vec{y}\}) = \lambda^{-1}\mathcal{H}_N^{KK}(\lambda m,\lambda^{-1}R,\{\vec{y}\})\,. \tag{10.24}$$

where the first and the second arguments denote respectively the overall mass scale and the radius of S^1 as usual. Let us now choose:

$$r=\sqrt{RL}, \qquad m = M\sqrt{L/R}, \qquad \lambda=\sqrt{R/L}=(r/L). \tag{10.25}$$

Substituting this in (10.24) we get,

$$\mathcal{H}_N^{KK}(M\sqrt{L/R},R,\{\vec{y}\}) = \frac{L}{r}\mathcal{H}_N^{KK}(M,r,\{\vec{y}\})\,. \tag{10.26}$$

Thus the recipe (10.15) can now be rewritten as

$$\mathcal{H}_N^{DLCQ}(M, L, \{\vec{y}\}) = \lim_{r\to 0} \frac{L}{r} \mathcal{H}_N^{KK}(M, r, \{\vec{y}\})\,. \tag{10.27}$$

It is this form of the identity that we shall attempt to prove.

The basic idea behind this proof is to regard the light-like circle as an infinitely boosted space-like circle of zero radius[173, 174]. Let us start with theory $\mathcal{T}$ compactified on a space-like circle S^1 of radius r. If x and t denote the coordinate along S^1 and the time coordinate respectively, then we have an identification:

$$\begin{pmatrix} x \\ t \end{pmatrix} \equiv \begin{pmatrix} x \\ t \end{pmatrix} + 2\pi \begin{pmatrix} r \\ 0 \end{pmatrix} . \tag{10.28}$$

Let us now define new coordinates (x', t') and $x'^{\pm}$ as follows:

$$\begin{pmatrix} x' \\ t' \end{pmatrix} = \begin{pmatrix} x\cosh\alpha - t\sinh\alpha \\ t\cosh\alpha - x\sinh\alpha \end{pmatrix} , \tag{10.29}$$

$$x'^{\pm} = \frac{1}{\sqrt{2}}(t' \pm x')\,. \tag{10.30}$$

In this coordinate system eq.(10.28) takes the form:

$$\begin{pmatrix} x'^{+} \\ x'^{-} \end{pmatrix} \equiv \begin{pmatrix} x'^{+} \\ x'^{-} \end{pmatrix} - \sqrt{2}\pi r \begin{pmatrix} -e^{-\alpha} \\ e^{\alpha} \end{pmatrix} . \tag{10.31}$$

Now consider the limit $r \to 0$, $\alpha \to \infty$ keeping fixed

$$L \equiv \frac{r}{\sqrt{2}} e^{\alpha}. \tag{10.32}$$

In this limit eq.(10.31) reduces to

$$\begin{pmatrix} x'^{+} \\ x'^{-} \end{pmatrix} \equiv \begin{pmatrix} x'^{+} \\ x'^{-} \end{pmatrix} - 2\pi \begin{pmatrix} 0 \\ L \end{pmatrix} . \tag{10.33}$$

This is equivalent to compactifying x'^{-} on a circle of radius L.

Under this map, a system carrying momentum N/R along S^1 gets mapped to a system carrying total momentum $k'^{-} = N/L$ along the x'^{-} direction. Thus it is not surprising that there is a relation between the Hamiltonian describing the two systems. To find the precise relation between these two hamiltonians, we need to study the relation between the usual time coordinate t of the original theory and the light-cone time x'^{+} of the boosted theory. From eqs.(10.29), (10.30) it follows that:

$$\frac{\partial}{\partial x'^{+}} = \frac{1}{\sqrt{2}} e^{\alpha} \Big(\frac{\partial}{\partial t} + \frac{\partial}{\partial x} \Big)\,. \tag{10.34}$$

Since the quantum operators which generate $i(\partial/\partial x'^+)$, $i(\partial/\partial t)$ and $i(\partial/\partial x)$ are $\mathcal{H}_N^{DLCQ}$, $\mathcal{H}_N^{KK} + M_N$ and $-N/r$ respectively, with M_N being the rest mass of the N Kaluza-Klein modes in $\mathcal{T}$ on S^1, we see from eq.(10.32), (10.34) that in the $r \to 0$ limit with L fixed,

$$\mathcal{H}_N^{DLCQ}(M, L, \{\vec{y}\}) = \frac{L}{r}(\mathcal{H}_N^{KK}(M, r, \{\vec{y}\}) + M_N - \frac{N}{r}) = \frac{L}{r}\mathcal{H}_N^{KK}(M, r, \{\vec{y}\})\,, \qquad (10.35)$$

since $M_N = N/r$. This reproduces (10.27).

If we recall that $\mathcal{H}_N^{DLCQ}$ is supposed to describe the theory $\mathcal{T}$, whereas $\mathcal{H}_N^{KK}$ describes theory $\mathcal{T}$ compactified on a small circle, then by the above argument, quite generally we can reconstruct a theory by knowing its behaviour when compactified on a small circle. This seems counterintuitive, so let us examine the steps leading to this conclusion. They may be summarized as follows:

1. We start with a small circle.

2. We convert this to an almost light-like circle of finite radius via a large boost.

3. We then take the limit where the radius of this light-like circle goes to infinity.

As we can see, the key point in this proof is the assumption that a light-like circle can be considered as a space-like circle of zero radius in the limit of infinite boost. Of course, this may be taken as a definition of the light-like circle. However, we are interested in a definition in which the radius of the light-like circle acts as an infra-red regulator in the uncompactified theory, so that in the end by taking $L \to \infty$ limit we recover the amplitudes in the uncompactified theory. Clearly there is a possibility that these two definitions do not match[174]. Indeed there are explicit computations which show that these two definitions do not always match for finite N[175, 176, 177, 178, 179], although they do match for some specific terms in the supergraviton scattering amplitudes[158, 180, 173]. (Note that the 'proof' given above did not involve taking the $N \to \infty$ limit.) It has been suggested[21, 22] that this problem might go away in the $N \to \infty$ limit, but there is no compelling argument as of now in favour of this. This of course does not mean that Matrix theory is wrong, it is just that we do not know for sure if it is right, and even if it is right, we do not quite know why it is right. Perhaps the arguments of ref.[161] together with supersymmetry non-renormalization theorems and properties of the large N limit can be combined to constitute such a proof.