Duality symmetries in string theory

Ashoke Sen

Mehta Research Institute of Mathematics and Mathematical Physics, Chhatnag Road, Jhusi, Allahabad 211 019, India

In this article I review duality symmetries in string theory and quantum field theories.

1. Introduction

One of the questions which has been around almost since the beginning of civilization is: what are we and everything around us made of? The conventional approach to our search for the answer to this question has been based on the so-called reductionist approach. According to this approach we try to describe matter by its elementary constituents, and develop a theory which describes various properties of these constituents. Given a complete understanding of the dynamics of these constituent particles, we should in principle be able to derive the various properties of all other matter. Thus for example, the property of a hydrogen molecule can be explained by knowing the properties of its constituents, the hydrogen atoms; the property of a hydrogen atom is explained from the property of its constituent proton and electron; and the property of a proton itself is explained in terms of its constituents, namely the quarks. Of course in practice it is almost never possible to carry out such analysis exactly, and one needs various approximation schemes. This problem is particularly significant while dealing with systems with large number of constituent particles, as in condensed matter physics; and there we need to develop new techniques to analyse such systems. But in principle (e.g. if we had a large enough computer) there is no obstruction to deriving the properties of matter from those of its constituents.

The above approach requires us to make a clear distinction between particles which are elementary and those which are composite: made of two or more elementary particles. After the advent of relativistic quantum mechanics, particularly quantum field theory we were forced to modify this viewpoint somewhat due to the possibility of particle production in collisions (note 1). Thus for example if we bring together an electron and a positron, they can annihilate and produce photons. This might lead us to believe that a photon should be considered as being made of electron and positron. Indeed, according to the principles of relativistic quantum mechanics, the photon can, for a short time, exist as an electron–positron pair. Similarly, the mediators of weak interaction, $W^{\pm}$ and Z, can decay into other elementary particles. As a concrete example we can take the Z boson, which can decay into an electron–positron pair. Due to such phenomena, strictly speaking there is no concrete experiment possible even in principle which can distinguish a composite particle from an elementary particle. Our present understanding of the physics of elementary particles, based on the standard model, treats electron, positron, various quarks and antiquarks (which are the building blocks of the proton, neutron and the π mesons) as well as $W^{\pm}$, Z, and the photon as elementary particles. This theory has been enormously successsful in explaining all known experimental results involving these particles to a very high degree of accuracy. Still, we should keep in mind that it is in principle possible to devise a theory in which some of these particles are bound states of the other particles. For example, one might have a theory in which the Z-boson appears as a bound state of the electron and the positron (note 2). The success of the standard model can be reproduced since it is in principle possible to ensure, by choosing suitable interaction between the electrons, positrons, quarks and other elementary particles in this new theory, that all the predictions of this new theory agree exactly with that of the standard model. Such a theory will involve a very complicated interaction between the elementary particles, including action at a distance also known as *non-local interactions* (note 3). On the other hand, the standard model is based on a very simple set of interactions between the particles which are considered as elementary in this model. Hence, even if we have such an alternative theory where the Z boson is regarded as a bound state of the electron and the positron, we would still regard the standard model, and not this (hypothetical) alternative theory, as fundamental. This in turn would require us to regard the Z boson (as well as the electron and the positron) as elementary particle, as this is how the standard model is formulated.

To summarize this discussion, in relativistic quantum mechanics we distinguish between elementary and composite particles by demanding that the theory formulated in terms of elementary particles has simple interactions. This seems to be a highly subjective notion, but there are definite criteria for what we call simple interactions: the interactions described by *renormalizable local quantum field theories*. For understanding the rest of

e-mail: sen@mri.ernet.in

this article it is not necessary to know the precise definition of a renormalizable local quantum field theory, but one should remember that these are special kind of theories, and not all quantum field theories fall into this category. The standard model is based on such a quantum field theory. In this theory the elementary particles include electrons, quarks, photon, $W^{\pm}$ and Z bosons etc. On the other hand the proton is a composite particle made of quarks interacting via the exchange of gluons.

String theory is based on the same reductionist approach, although the elementary constituents, instead of being particles, are strings (note 4). According to this viewpoint, the spectrum of particles in string theory can be divided into two classes: the elementary particles which are just different excitations modes of a single string, and the composite particles which are made of more than one (quite often infinite number of) strings. The theory is completely described by specifying the interaction between the elementary constituents, namely the strings. The interactions involving composite objects can be derived by knowing the interaction between the elementary strings, although the calculations involved may be quite difficult in some cases.

Before I conclude this section I would like to make a note of the terminology without which this article may be somewhat confusing. By elementary particles in a quantum field theory I shall refer to only those particles which are the elementary quanta associated with the fields of the quantum field theory. In a layperson's language this amounts to taking a wave associated with the field (e.g. the electromagnetic wave in case of electrodynamics) and associating a particle with it using the wave-particle equivalence of quantum mechanics. Any other kind of particle in the theory which is not associated with the fields in this fashion will be refered to as composite particle. Examples of such composite particles will involve bound states of two or more elementary particles, e.g. the hydrogen atom which is a bound state of electron and proton (which in turn is a bound state of quarks), and also *solitons*. Solitons are solutions of the classical field equations (analog of Maxwell's equation in electrodynamics) with energy localized around a given point in space. When we make the classical field theory into a quantum theory, these solitons behave like particles composed of infinite number of elementary particles. There is no known particle which can be identified as a soliton in the standard model, but many quantum field theories have such particles in their spectrum.

In string theory by elementary particles I shall refer to those 'particle'-like states which arise as quantum states of a single string. All other particles will be referred to as composite particles. These will include bound states and solitons, and also the *Dirichlet branes* described in the previous article.

2. Duality

In this section I shall describe the basic ideas of duality and then discuss some examples. Let us start with the definition.

2.1 Definition

In order to understand the notion of duality in quantum field theory or string theory, we need to first recall the steps which are normally followed in defining a quantum theory. These are as follows:

- Begin with a classical system described by a certain set of dynamical degrees of freedom with a certain Hamiltonian (note 5).
- Quantise the system by replacing the Poisson brackets between the canonically conjugate variables by commutator brackets (note 6).

Duality in its most general form is a statement of equivalence between two or more 'apparently different' quantum theories. Here by apparently different theories we refer to theories whose corresponding classical theories are genuinely different, i.e. there is no change of variables which can relate these classical theories. Two such theories will be called dual to each other if they are identical as quantum theories, i.e. if there is a unitary transformation relating the Hilbert spaces of the two theories under which all correlation functions in one theory are mapped exactly to the corresponding correlation functions in the other theory. Thus a dual pair of theories represent two theories which are identical as quantum theories, but yet their classical limits are genuinely different.

One might wonder how this could be possible. The classical limit of the theory can be regarded as the $\hbar \to 0$ limit of the quantum theory, and so if two theories are the same as quantum theory, i.e. they are same for finite $\hbar$, how can they look different in the $\hbar \to 0$ limit? The key to understanding this phenomenon is to note that $\hbar \to 0$ does not define a unique limit. In order to define this limit uniquely, we must also specify which quantities are kept fixed in this limit. In general, we may be able to define different classical limits by holding fixed different sets of quantities as we take $\hbar \to 0$ limit. The resulting classical theories are very different, and yet they are the classical limits of the same underlying quantum theory (note 7).

I shall illustrate this through an example. There are quantum field theories (as well as string theories) which contain particles carrying electric charge as well as particles carying magnetic charge in their spectrum. If e denotes the quantum of electric charge, g denotes the quantum of magnetic charge, and c denotes the speed of light in the vacuum, then they satisfy a quantization rule:

$$eg = 2\pi\hbar c. \tag{2.1}$$

There are many examples of quantum field theories and string theories of this type which admit two possible classical limits. We can take $\hbar \to 0$ keeping g fixed, or $\hbar \to 0$ keeping e fixed. These classical theories are not equivalent, although they are different limits of the same underlying quantum theory. For example in the first limit, the magnetically charged particles arise as solitons, while electrically charged particles arise as elementary particles. On the other hand in the second limit, the electrically charged particles arise as solitons, while the magnetically charged particles arise as elementary particles. For future reference, I shall describe the $\hbar \to 0$ limit with g fixed as the first classical theory, while the $\hbar \to 0$ limit with e fixed as the second classical theory. Also I shall refer to the quantization of these theories as the first and the second quantum theories respectively, although they describe the same theory. In this notation, the first and the second theories are dual to each other. This particular kind of duality is known as electric-magnetic duality.

From the above discussion we see that under duality, the elementary particles of first theory gets mapped to the composite particles of the second theory and vice versa. In other words the same particle may be considered elementary in one description and composite in the other description. Thus in theories possessing dual descriptions; the question of whether a given particle is elementary or composite has different answers, depending on which description we use for the theory (note 8).

Typically in a quantum field theory, or a string theory we compute scattering amplitude/cross section for scattering involving various particles in the spectrum. Such an amplitude can be expressed as a power series in $\hbar$, with the leading contribution known as the classical contribution, and the other terms known as quantum corrections. It is clear that such a rearrangement of terms depends crucially on what quantities are considered to be $\hbar$ independent (e.g. kept fixed in the 'classical limit' $\hbar \to 0$). In particular in the example of the previous paragraph, if K denotes some physical quantity, then quantization of the first theory will give a power series expansion in $\hbar$ with coefficients regarded as functions of g:

$$K = \sum_{n=0}^{\infty} K_n'(g)\hbar^n. \tag{2.2}$$

On the other hand, quantization of the second theory will give a power series expansion of the same quantity with coefficients regarded as functions of e:

$$K = \sum_{n=0}^{\infty} K_n(e)\hbar^n. \tag{2.3}$$

Although these two expansions describe the same quantity K, they look very different. In particular the leading term K_0 in the expansion (2.3) will contain information about the non-leading terms in the expansion (2.2). Thus we see that the classical result in one theory may contain information about quantum effects in the dual theory, i.e. duality mixes up classical and quantum effects!

Since $\hbar$ is not a dimensionless parameter, it is useful to reanalyse the situation in terms of expansion in dimensionless parameters. For this, let us introduce dimensionless parameters:

$$\tilde{e} = e / \sqrt{2\pi\hbar c}, \quad \tilde{g} = g / \sqrt{2\pi\hbar c}. \tag{2.4}$$

For future reference we note that the parameter $\tilde{e}$ is known as the *coupling constant* of the first theory, since it measures the strength of the interaction or coupling between two elementary particles. (As stated earlier, in the first theory the electrically charged particles appear as elementary particles.) By the same token, $\tilde{g}$ is the coupling constant of the second theory. In terms of $\tilde{e}$ and $\tilde{g}$, eq. (2.1) takes the form

$$\tilde{e}\tilde{g} = 1. \tag{2.5}$$

Let us now consider the expansion given in (2.3). Since $\hbar$ is a dimensionful parameter, different terms in the expansion (2.2) or (2.3) have different dimensions. We can remedy this situation by defining new coefficients

$$L_n \equiv K_n(e)e^{2n} / (2\pi c)^n, \quad L_n' \equiv K_n'(g)g^{2n} / (2\pi c)^n, \tag{2.6}$$

so that (2.2) and (2.3) now take the form:

$$K = \sum_n L_n' \tilde{e}^{2n}, \tag{2.7}$$

and

$$K = \sum_n L_n \tilde{g}^{2n} \tag{2.8}$$

respectively. Since $\tilde{e}$ and $\tilde{g}$ are dimensionless parameters, each term in the expansion must have the same dimension. Furthermore since g remains fixed in the $\hbar \to 0$ limit of the first theory, all the coefficients L_n' are finite in this limit. Similarly all the coefficients K_n' are finite in the $\hbar \to 0$ limit of the second theory. This shows that the expansion (2.2) can be regarded as an expansion in the dimensionless parameter $\tilde{e}$, while expansion (2.3) can be regarded as an expansion in the dimensionless parameter $\tilde{g}$. Since $\tilde{e}$ and $\tilde{g}$ are related by eq. (2.5), it is clear that the individual coefficients of expansion in the two series are going to be very different, although both series represent the expansion of the same physical quantity. In future discussion, I shall always use these dimensionless coupling constants in the perturbation expansion, although one can always convert this to an expansion in $\hbar$ by reversing the procedure followed here.

At this point the reader may recall our discussion in the last section. There I pointed out that in relativistic quantum mechanics, there is no strict distinction between elementary and composite particles; for example it may even be possible to regard the Z boson as a bound state of electron and positron by introducing suitably complicated interaction among the electron, positron and the quarks. In principle this can also be taken as an example of duality – duality between the standard model and the (hypothetical) new theory where the Z boson is a bound state of electron and the positron. However such alternative descriptions, although possible, are not particularly useful or illuminating, since this hypothetical dual theory will involve very complicated interaction between its elementary particles and will not be useful for anything. In this article we shall focus on only those kinds of dualities which relate two apparently different theories, *each with simple interaction rules.* In case of quantum field theories this would require that each of the two theories correspond to a renormalizable local quantum field theory (note 9), whereas in the case of string theory, this would require that each of the two theories is governed by simple interactions of the kind described in the previous article.

Finally, for readers familiar with the Ising model in statistical mechanics, one can draw an analogy with the duality in the Ising model. There the duality relates the high temperature phase of the theory to the low temperature phase, and under this duality the order and the disorder parameters get exchanged. The role of temperature in the Ising model is played here by $\hbar$, or, in terms of dimensionless numbers, the coupling constant $\tilde{e}$. The order parameter is the analog of the field associated with an elementary particle, whereas the disorder parameter is the analog of the field associated with a composite particle. Duality transformation exchanges them.

2.2 Examples

I shall now give some concrete examples of dualities. As I have already said, such examples exist both in quantum field theories and in string theories (and as you will learn from a subsequent article, also between quantum field theory and string theory). Since you have already learned about many string theories in the previous article, I shall begin with examples of dualities involving string theories. At the end I shall give an example of duality in a quantum field theory.

1. You have learned that in ten (9 space, 1 time) dimensions there are five different consistent string theories, two of which are Type-I and SO(32) heterotic string theories. In the classical limit these theories appear to be very different. Indeed, the Type-I theory contains both closed and open strings in its spectrum of elementary particles, whereas the heterotic string theory has only closed strings. However, it has been found that as quantum theories they are the same. This implies, among other things, that the complete spectrum of states (including elementary and composite particles) in the two theories are identical. The dimensionless coupling constants in the two theories are related by a formula similar to eq. (2.5) (note 10).
2. You have also learned that starting with a string theory in ten dimensions, we can get a lower-dimensional theory by compactification, in which some of the space directions are curled up into a small compact manifold. The example that I am going to discuss now involves string theories in which some of the dimensions have been compactified. The first of these is the SO(32) heterotic string theory, with four of the directions compactified on a four dimensional space known as T^4 – the four-dimensional torus. This space is not very difficult to describe; take a four dimensional Euclidean space, pick four mutually orthogonal directions, and make each of these four directions into circles instead of infinite line. The resulting theory has 5 infinite space-like dimension and one time dimension (which is always infinite). The second theory involves Type-IIA string theory, with four dimensions compactified on another four-dimensional space called $K3$. Unlike the space T^4 which is simple to describe, $K3$ is a very complicated space. Indeed, many of the geometric properties of this space (e.g. the distance between a pair of points in this space) are unknown to this date. The resulting theory again has five infinite space-like directions and one time direction.

   It turns out that these two theories are quantum mechanically equivalent, although classically they look very different. The dimensionless coupling constants of the two theories are again related by an equation analogous to (2.5).
3. This example will again involve compactified string theories, but with simpler compact spaces. Consider Type-IIA string theory, with one direction compactified on a circle. The resulting theory has 8 infinite space-like dimensions and 1 time dimension. Now consider Type-IIB string theory, and again compactify on a circle. The resulting theory again has 8 in-

finite space-like dimensions and 1 time dimension. It turns out that these two theories describe the same quantum theory. But in this case something special happens; they also describe the same classical theory! In fact, if we compare the $\hbar$ expansion of the two theories, they agree to each order in $\hbar$. Put another way, the dimensionless coupling constants of the two theories are the same!

From this discussion it would seem that this example does not quite fit in our scheme; if the theories are the same as classical theories then it seems almost inevitable that quantization will also give the same theory; so where is the surprise? The surprise here lies in the fact that they are the same as classical theories! Type-IIA and Type-IIB string theories in ten dimensions are very different theories. So it is certainly a nontrivial fact that upon compactifying two different theories on a circle we get the same theory. The surprise becomes deeper when one compares the radii of the circle of compactification of the two theories. Let R_A denote the radius of the compact circle in the first theory, and R_B denote the radius of the compact circle in the second theory. It turns out that the two theories describe the same classical (and quantum) theory only if these two radii satisfy the relation

$$R_A R_B = \alpha', \tag{2.9}$$

where α' is a constant of dimension length square, and is related to the mass per unit length T of the string via the relation:

$$\alpha' = \frac{\hbar}{2\pi c T}. \tag{2.10}$$

From this relation we see that smaller is the value of R_A, larger is the value of R_B (and vice versa). Thus for this theory there is no single answer to what is the radius of the compact direction. It depends on whether we use the description based on the Type-IIA string theory, or the description based on the Type-IIB string theory.

What this example teaches us is that even in classical string theory, there is no invariant notion of geometry of a compact space-time. The same classical string theory may have two different descriptions with very different geometries. Thus in string theory not only the notion of elementary particles and classical limit is picture dependent, but the notion of geometry is also picture dependent!

Dualities of the kind discussed here, which do not mix up classical and quantum effects or elementary and composite particles, but involves map between different geometries, have a special name. These are known as T-dualities.

4. Another example of T-duality is the duality between SO(32) heterotic string theory with one direction compactified on a circle of radius R, and $E_8 \times E_8$ heterotic string theory with one direction compactified on a circle of radius (α'/R). Due to this duality, the duality between the SO(32)-heterotic string theory compactified on T^4 and Type-IIA string theory compactified on $K3$ also holds if we replace the SO(32)-heterotic string theory by the $E_8 \times E_8$ heterotic string theory.
5. So far I have discussed examples of dualities which map one string theory to another string theory. But there are also examples of duality transformations which take a particular string theory to the same string theory, but different value of the coupling constant labelling the theory. These theories are known as self-dual theories. An example of this is the Type-IIB string theory in 9 + 1 dimension. Besides the fundamental constants $\hbar$, c and α', Type-IIB string theory is parametrized by a dimensionless coupling constant (note 11). It turns out that for any two values of this coupling constant, say $\tilde{e}$ and $\tilde{e}'$, related by the relation

$$\tilde{e}' = \frac{1}{\tilde{e}}, \tag{2.11}$$

the corresponding Type-IIB string theories are quantum mechanically equivalent. However, if one considers perturbation expansion of any physical quantity in the coupling constant, clearly the coefficients of expansion in powers of $\tilde{e}$ will be very different from the coefficients of expansion in powers of $\tilde{e}' = \tilde{e}^{-1}$. By following the example of electric-magnetic duality discussed earlier, the expansion in $\tilde{e}$ can be reagrded as an expansion in $\hbar$ in the first theory, and the expansion in $\tilde{e}'$ can be regarded as an expansion in $\hbar$ in the second theory. Thus this transformation mixes up the classical and quantum effects in the two theories, and is an example of a duality transformation in the same sense as defined in the last subsection.
6. Finally let me give an example of duality in a quantum field theory. There is a special supersymmetric gauge theory, known as N = 4 supersymmetric SU(2)-gauge theory, in four-dimensional space-time (3 space, 1 time). This theory has electrically charged particles as elementary particles and magnetically charged particles as solitons. The strength of the interaction between the elementary electrically charged particles is controlled by a dimensionless coupling constant $\tilde{e} = e/\sqrt{2\pi\hbar c}$ where e is the quantum of electric charge. This theory turns out to be self-dual under an electric-magnetic duality transformation of the kind described earlier. In other

words, the theory at coupling constant $\tilde{e}$ is equivalent to the same theory at coupling constant $\tilde{e}' = (1/\tilde{e})$. We shall refer to the theory with coupling constant $\tilde{e}$ as the first theory, and the theory with coupling constant $\tilde{e}'$ as the second theory. The duality transformation relating the two theories maps an electrically charged elementary particle of the first theory to a magnetically charged soliton of the second theory. Note that the value of one quantum of magnetic charge (after rescaling according to eq. (2.4)) in this second theory is $(1/\tilde{e}') = \tilde{e}$ – the same as the value of the electric charge quantum in first theory. Similarly the duality transformation maps a magnetically charged soliton of the first theory to the electrically charged elementary particle of the second theory.

3. Testing duality conjectures

So far I have discussed the basic notion of duality symmetries, and have given some examples in the context of string theory. However I have not addressed one basic question: how do we guess, test or prove the existence of duality between two apparently different theories? At a conceptual level the answer is simple; since duality is a statement of quantum equivalence between two apparently different classical theories, one should compute various physical quantitities in both quantum theories and compare answers. If the two theories are dual to each other then the answers should agree. The problem however is that typically in a quantum field theory or a string theory one can never perform an exact calculation. Instead what one has is a perturbation expansion in $\hbar$, which, by following the procedure outlined in the last section, can be converted into a perturbation expansion in some dimensionless coupling constant. But there is no reason for the individual terms in the perturbation expansion in a pair of dual theories to agree. Indeed, in the electric-magnetic duality example discussed in the last section, one description of the theory gives an expansion in the dimensionless coupling constant $\tilde{e}$ defined in eq. (2.4), while the other description gives a perturbation expansion in $(1/\tilde{e})$. By knowing the first few terms in both the perturbation series we cannot determine if the two answers are the same. In fact, not only can we not prove duality this way, we cannot even test duality, as duality does not predict any simple relationship between the two sets of expansion coefficients (note 12).

This is where supersymmetry comes to our rescue. As you have learned from the previous article, supersymmetry is not a symmetry that is observed in nature and hence must be broken below some energy scale, but it is a property of a wide class of (compactified) string theories. In theories with supersymmetry, there are often restrictions on the kind of quantum corrections which can modify a classical answer. In particular in some supersymmetric string theories there are special physical quantities for which there are no quantum corrections; thus the classical answer (the leading term in $\hbar$ expansion) is the complete quantum answer. The theorems which guarantee absence of quantum corrections are usually referred to as nonrenormalization theorems, and such physical quantities are known as nonrenormalized quantities. For such theories we are in a better position to test duality, since we can now compute some physical quantities exactly and hence compare their values in the two different descriptions to see if they agree. In some cases a given physical quantity may be exactly computable in one description due to absence of quantum corrections *in that description*, but can only be computed as a perturbation expansion in the dual description. In this case one can take the exact answer from the first description, expand it in Taylor series in the coupling constant of the theory in the dual description, and check if each term in the Taylor series expansion agrees with the explicit computation in the second description using perturbation theory.

It should be clear from this discussion that such analysis can never provide a proof of duality, since in these theories the nonrenormalization theorems hold only for a small subset of physical quantities. In order to prove duality between two theories we need to show that *all* physical quantities in the two theories agree. Nevertheless, many nontrivial tests of duality have been provided by these nonrenormalization theorems. Also we should emphasize that although supersymmetry is necessary for testing duality, in general there is no reason why duality should only be a property of supersymmetric theories. In fact there are several conjectured dualities between nonsupersymmetric theories which have been derived by starting from a dual pair of supersymmetric theories, and then breaking supersymmetry in both theories by following a specific set of rules. Thus the conjectured dualities involving these nonsupersymmetric theories are on a reasonably solid footing, although we cannot directly carry out a test of duality involving these theories.

We shall now discuss two examples of such nonrenormalized quantities.

3.1 Spectrum of BPS states

One of the intrinsic properties of a quantum theory is the spectrum of states in the theory. Thus if there are two different descriptions of the same theory, the spectrum of states must be identical in the two descriptions. But since in general the spectrum depends on the coupling constant labelling the theory (e.g. the parameter $\tilde{e}$ appearing in the electric-magnetic duality example) we need to ensure that when we compare the spectrum of the two theories, we choose the corresponding coupling constants in a way that they are related to each other by the duality relation analogous to eq. (2.5). This, however, is not possible in general, since typically we know the spectrum only as a perturbation expansion in the coupling constant. Thus for example in the example involving electric-magnetic duality, in one description the spectrum is known as a series expansion in $\tilde{e}$, and in the other description it is known as a series expansion in $(1/\tilde{e})$. Since we only know the first few coefficients in each series expansion, it is in general impossible to compare the spectrum in the two theories, and determine if it is the same.

But in special supersymmetric (field or string) theories there is a special class of single particle states with the property that the dependence of the mass of the particle on the coupling constant of the theory is known exactly. These special states are known as BPS states, named after Bogomol'nyi, Prasad and Sommerfeld. Furthermore, if a BPS state carrying a given set of charge quantum numbers is part of the spectrum for one value of the coupling constant, then it remains part of the spectrum for any other value of the coupling constant (note 13). Thus the spectrum of BPS states in a given theory can be computed by first computing the spectrum of BPS states for a small value of the coupling constant (where perturbation theory is valid), and then using the known BPS mass formula to compute the mass of each of these BPS states at any arbitrary value of the coupling constant. This gives us a method for computing the spectrum of BPS states in a theory for all values of the coupling constant. Thus given a pair of theories, we can explicitly compute the spectrum of BPS states in each theory, and compare the answers. If they do not agree then the two theories cannot be dual to each other. On the other hand if they agree then there is strong reason to believe that they are indeed dual to each other.

Let us now look at an example. As stated earlier, the N = 4 supersymmetric SU(2) gauge theory has been conjectured to be dual to itself under an electric-magnetic duality transformation. Under this duality the electrically charged states of the first theory get mapped to the magnetically charged states in the second theory and vice versa. Now it can be checked explicitly that at small $\tilde{e}$ the theory contains 16 BPS states carrying one quantum $(e = \tilde{e}\sqrt{2\pi\hbar c})$ of electric charge (note 14). Due to the BPS nature of the state, we can conclude that there are 16 BPS states with one quantum of electric charge for all values of the coupling constant $\tilde{e}$, and furthermore, their mass can be determined using the BPS mass formula. Under duality map this theory goes to the same theory, but with coupling constant $\tilde{e}' = (1/\tilde{e}')$, and the 16 BPS states carrying one quantum of electric charge becomes 16 BPS states carrying one quantum of magnetic charge. Thus the self-duality of the N = 4 supersymmetric field theory predicts that the theory must have 16 magnetically charged BPS states carrying one quantum of magnetic charge for all values of $\tilde{e}'$. The mass of these states is determined by the BPS mass formula.

So the question is: are these states present in the theory? If they are present, then it will lend support to the conjectured self-duality of the theory. On the other hand if these states are not present then this will tell us that this conjecture is false. Since the spectrum of BPS states can be computed for any value of the coupling constant $\tilde{e}$, we can do the analysis for small $\tilde{e}'$ where we can use perturbation theory. One finds that these states are indeed present in this theory, lending support to the conjectured duality.

Incidentally, it turns out that this N = 4 supersymmetric SU(2)-gauge theory has a much larger (in fact infinite set) of duality transformations. A typical transformation is characterized by four integers p, q, r and s satisfying the relation

$$ps - qr = 1, \tag{3.1}$$

and transforms a BPS state carrying one quantum of electric charge to a BPS state carrying p quanta of electric charge and r quanta of magnetic charge. This also transforms the coupling constant in a complicated way, but we do not need to know it for this discussion. Now it is a simple exercise to show that given four integers p, q, r and s satifying eq. (3.1), p and r cannot have a common factor. Such pair of integers are known as *relatively prime integers*. Furthermore, it can also be shown that given a pair of relatively prime integers p and r, we can always find integers s and q satisfying eq. (3.1). Thus duality predicts that for every pair of relatively prime integers (p, r) the N = 4 supersymmetric gauge theory must have 16 BPS states carrying r quanta of magnetic charge and p quanta of electric charge.

This prediction has been explicitly verified for all states carrying one or two quanta of magnetic charge, i.e. for $r = 1$ and $r = 2$. For $r = 1$ this requires showing the existence of BPS states with one quantum of magnetic charge and arbitrary integer quanta of electric charge, whereas for $r = 2$ it involves showing the existence of BPS states carrying two quanta of magnetic charge and arbitrary *odd* integer quanta of electric charge. Extending these results to higher values of r requires highly sophisticated mathematical analysis. Although there is no conclusive proof of the existence of these states yet, the progress has been quite encouraging. But already the existence of appropriate BPS states with one and two quanta of magnetic charge and appropriate quanta of electric charge gives us strong evidence that the self-duality conjecture of this theory is indeed correct. The BPS states with two quanta of magnetic charge appear as quantum mechanical bound states of two BPS states, each carrying one quantum of magnetic charge. Thus here we see an explicit example where a duality transformation takes an elementary particle to a bound state of two particles and not just a single soliton.

The procedure of using BPS states to test duality has been used extensively in many examples, involving quantum field theories as well as string theories. In string theory duality transformations map an elementary particle not only to solitons and their bound states, but also to D-branes, and various bound states of D-branes and solitons. This is a fascinating subject, but the general principle remains the same, and so I shall not discuss them here. Instead I shall turn to another kind of test of duality – based on the study of interactions rather than the spectrum.

3.2 Effective lagrangian density

In the previous subsection I discussed a method of testing duality by comparing the spectrum of particles in the two theories. But a quantum field theory or a string theory is characterized not only by the spectrum of particles that it contains, but also by how these particles interact. These interactions control, for example, how two particles scatter or how a particle decays into other particles. If two theories are dual to each other, then they must have identical interactions. The difficulty in checking explicitly if this is so lies again in the fact that in a quantum theory the interaction between particles can only be computed as a perturbation expansion in the coupling constant of the theory; and the individual terms in the perturbation expansion in a dual pair of theories need not agree.

Information about interactions involving massless particles with small external momenta can be encoded in a function of various fields and their derivatives known as the *effective lagrangian density* (henceforth denoted by $\mathcal{L}_{\rm eff}$). If two theories are dual, then they must be described by the same $\mathcal{L}_{\rm eff}$ (possibly after suitable change of variables). In general $\mathcal{L}_{\rm eff}$ cannot be computed exactly, but can only be computed using perturbation theory. As discussed earlier, by knowing the first few terms in the perturbation expansion of $\mathcal{L}_{\rm eff}$ of two theories, we cannot compare them to see if they are dual to each other. However in certain supersymmetric theories, certain terms in their $\mathcal{L}_{\rm eff}$ have the property that they do not get corrected by quantum effects, and hence the answer to the leading order in the perturbation theory is the exact answer. Thus in these cases we can test a duality conjecture by comparing these particular terms in $\mathcal{L}_{\rm eff}$ in the two theories. If they are the same then there is a good chance that they might be dual to each other. On the other hand if they are not the same, then the two theories are definitely not dual to each other.

There are also cases where supersymmetry prevents quantum correction to a set of terms in $\mathcal{L}_{\rm eff}$ in only one of the two theories which we are comparing. Let us call this theory the first theory, and the other theory the second theory. In this case we can compute these terms in $\mathcal{L}_{\rm eff}$ exactly in the first theory, and then expand this in Taylor series in the coupling constant of the second theory. The individual terms in this series can then be compared with the perturbation expansion of $\mathcal{L}_{\rm eff}$ of the second theory. If they agree, the theories are likely to be dual to each other. If they do not agree then the two theories cannot be dual to each other. Turning this procedure around, we see that if we find such a dual pair of theories, then information about the quantum effects in the second theory (higher order terms in the perturbation series in $\mathcal{L}_{\rm eff}$) is encoded in the purely classical contribution to $\mathcal{L}_{\rm eff}$ in the first theory.

There are many example involving supersymmetric string and field theories where comparison of $\mathcal{L}_{\rm eff}$ has led to tests of duality. In fact many of the duality conjectures were arrived at by comparing the $\mathcal{L}_{\rm eff}$ of the two theories.

4. Application of duality

As already discussed earlier, discovery of duality symmetries have radically changed our understanding of the constituents of matter by bringing in a sort of democracy between all particles – elementary and composite. But besides this it has improved our understanding of string theory in several other ways. I shall mention a few of them here.

4.1 Computational application

Duality implies the existence of two or more descriptions of the same theory. This allows us to get more information about a theory than is possible by using a single description. As discussed earlier, in any given string theory we can calculate a physical quantity only as a perturbation expansion in the coupling constant. This gives results for small values of the coupling constant, but does not tell us anything about what happens at large or finite values of the coupling constant. The only exceptions are quantities satisfying non-renormalization theorems. But now, for theories which admit a dual description, we can compute the same physical quantities as a perturbation expansion in the coupling constant of the dual theory. These results are valid when the coupling constant of the dual theory is small. but this typically correspond to large or finite value of the coupling constant of the original theory. Thus by exploiting duality symmetries we can get information about a given theory for large or finite values of the coupling constant – a task which was thought to be almost impossible before the advent of duality.

Quite often for supersymmetric theories we can recover remarkable amount of information by combining the results from duality with various nonrenormalization theorems. As an example we can mention certain class of supersymmetric string theories – known as N = 2 supersymmetric string theories in four dimensions – which have two different descriptions; as a compactification of one of the heterotic string theories, and also as a compactification of one of the Type-II string theories. Certain terms in $\mathcal{L}_{\text{eff}}$ are not modified by quantum effects in the Type-II description but are modified in the heterotic description. Certain other kind of terms in $\mathcal{L}_{\text{eff}}$ are not modified by quantum effects in the heterotic description, but are modified in the Type-II description. Thus combining the results from the two descriptions, we can find an exact answer for both kinds of terms in $\mathcal{L}_{\text{eff}}$. This certainly would not have been possible in absence of duality.

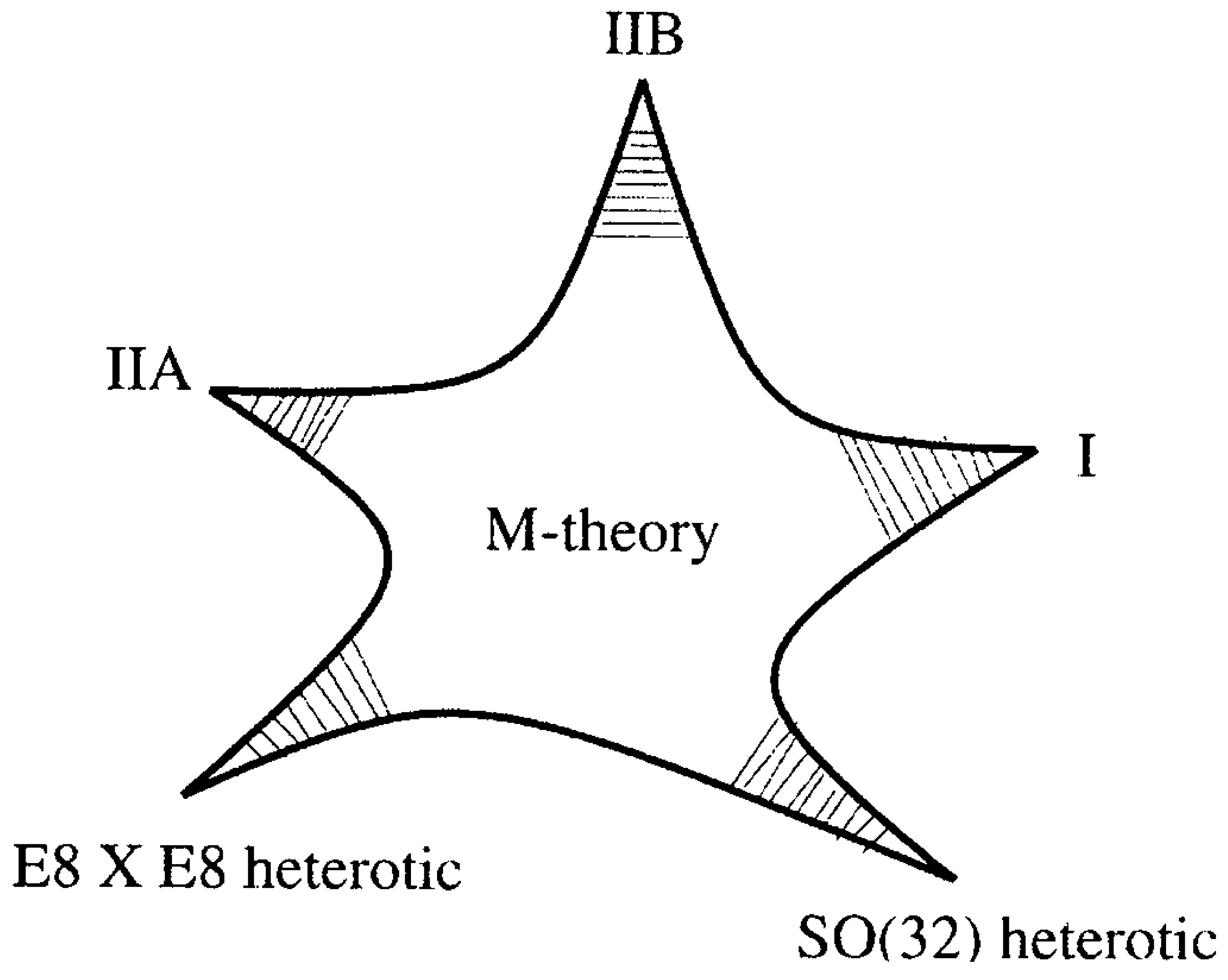

Figure 1. Unified picture of all string theories.

4.2 *Emergence of M-theory*

As you have seen in the previous article, there are five consistent string theories in ten dimensions. This is not a satisfactory situation; if there are five consistent theories, then how does nature choose between these theories? After the advent of duality we have seen that these five theories are not distinct theories, but they (and their various compactifications) often describe equivalent theories. Thus all these five string theories can be regarded as different limits of a single unified theory. This theory has been given the name M-theory.

The situation has been schematically illustrated in Figure 1. This diagram shows the parameter space of M-theory (note 15). We can identify the five corners as the classical limit of the five different string theories and their various compactifications. (Thus a given corner, instead of representing just one theory, represents a whole host of theories obtained by compactifying the parent theory.) The shaded region near the corners can be regarded as the weakly coupled version of the corresponding string theory where perturbation theory in the coupling constant can be trusted. The white region in the middle is the domain where the coupling constants of all the descriptions are large (or of order 1) so that the perturbation expansion of none of the string theories is a good description of the theory in this region. Understanding the theory in this region remains an open problem. If string theory describes nature, then presumably our universe corresponds to some point in the parameter space of M-theory. Finding this point, as well as understanding why we live at this point in the parameter space and not at any other point, also remains a challenging problem for the future.

Upon examining the parameter space of M-theory one discovers that there is one particular limit in which the theory behaves like an eleven (10 space, 1 time)- dimensional theory. This particular limit can be understood by starting with Type-IIA string theory in ten dimensions, and taking the limit in which its coupling constant approaches infinity (note 16). In this limit the spectrum and various scattering amplitudes involving particles carrying *small energy and momentum* agree with those computed from a well-known field theory, known as the N = 1 supergravity theory in eleven dimensions (note 17). Unfortunately the latter is a classical field theory, and one only knows how to compute the leading order terms in the $\hbar$ expansion of various quantities in this theory. Unlike a renormalizable quantum field theory or a string theory, one does not have a well-defined set of rules for computing this amplitude as a power series expansion in $\hbar$. Fortunately the same kind of dimensional analysis which we carried out in the example of electric-magnetic duality, when applied here, shows that the $\hbar$ expansion coincides with an expansion in powers of the energy and momenta of various external particles. Thus in the limit when the energy and momentum of each of the external particles are very small, the leading order terms in $\hbar$ expansion give the dominant contribution. The agreement between the leading order terms in the $\hbar$ expansion of the eleven dimensional supergravity theory and M-theory in this limit shows that in this particular 'classical limit', M-theory reduces to the eleven-dimensional N = 1 supergravity theory.

Although there is no well-defined quantum theory for the eleven-dimensional supergravity theory, M-theory is presumably a well-defined quantum theory. In other words in M-theory one should be able to compute corrections of order $\hbar$ and higher to the various scattering amplitudes. These corrections are important when the energy and momenta of the external particles are not small. Thus one can use M-theory as a definition of the quantum theory whose classical limit is the eleven-dimensional N = 1 supergravity theory. Thus not only does M-theory encompass all the known string theories, but it also defines novel quantum theories whose classical limits are well known classical field theories.

5. Conclusions

We conclude by summarizing the main points once more.

- Duality is an equivalence relation between two string theories or two quantum field theories whose classical limits are genuinely different. The classical limits of the two theories correspond to the $\hbar \to 0$ limit of the same underlying quantum theory, but keeping fixed different combinations of parameters.
- Under this equivalence, elementary particles in one theory may appear as composite particles in the dual theory, and vice versa. Thus the distinction between elementary and composite particles can no longer be regarded as a fundamental distinction.
- Typically in any string theory or quantum field theory, a physical quantity is computed using perturbation theory, as a series expansion in the coupling constant. Since duality only requires that the complete answer in the two theories should agree, but individual terms in the perturbation series need not agree, it is in general very difficult to test if a pair of theories are dual to each other. This is overcome by working with supersymmetric string theories or quantum field theories. In some of these theories there are special theorems which allow us to compute some of the physical quantities exactly by knowing the first term in the series expansion. Since these special physical quantities can be computed exactly in both theories, we can compare them in the two theories to check if they agree.
- Once we are convinced that a pair of theories are dual to each other, duality can be used to extract new information about any physical quantity in the theory. Since a given physical quantity can be computed in both theories as a series expansion in the respective coupling constants, and since the coupling constants in the two theories are not the same, these two series expansions contain complimentary information. Thus by combining these two series expansions we can learn more about the physical quantity than is possible by using any single description. In some special cases, this allows us to determine the quantity completely.
- In string theory duality also serves the purpose of unifying all five apparently different string theories. According to our present understanding, all of these string theories are simply different limits in the parameter space of a single underlying theory. This theory has been given the name M-theory, although at present our knowledge of M-theory has mostly been limited to those corners of the parameter space where it corresponds to a weakly coupled string theory.

Notes

1. I wish to thank R. Rajaraman for discussion on this point.
2. One could also try this exercise for the photon, but there are special problems in treating a massless particle as a bound state.
3. We can also try to do the reverse. According to the standard model, the proton, neutron and the π-mesons are considered as bound states of quarks and antiquarks. But we can in principle describe their interaction by an alternative theory in which each of them is considered as elementary particle. In order that the predictions of this new theory agree with those of the standard model, we need to introduce extremely complicated interaction between the elementary particles of the new theory.
4. Historically string theory started as an alternative to this reductionist viewpoint in which all particles were considered to be on equal footing, and the interaction between these particles were supposed to be governed by the requirement of consistency rather than as a result of interaction among some elementary constituents. It was realised only later that these interactions follow from the simple hypothesis that these various particles are different excitation modes of a string.
5. The number of degrees of freedom could be finite or infinite. It does not matter for our discussion.
6. There are other methods of quantising the system, e.g. using Feynman path integrals. For our discussion it does not matter which definition we choose.
7. I wish to thank B. Julia for a discussion on this point.
8. Although in this particular class of examples duality maps an elementary particle to a soliton, there are other examples where duality maps an elementary particle to a bound state of two or more particles. I shall discuss such an example later.
9. However, in many cases such effective dualities, relating a renormalizable quantum field theory to a nonrenormalizable quantum field theory have also played an important role in understanding behaviour of quantum field theories at low energies.
10. As was discussed in the previous article, in string theory the coupling constant is given by the value of a scalar field known as the dilaton. Whenever I refer to the coupling constant of a string theory, I shall mean the value of the dilaton field.
11. Actually, besides the coupling constant, Type-IIB string theory has another dimensionless parameter, but I shall consider the case where this parameter is set to zero. Like the coupling constant, this parameter is also related to the value of a scalar field of the theory.
12. Note that this kind of difficulty does not arise for testing T-duality. Since T-duality does not mix classical and quantum effects, the individual terms in the perturbation expansion should agree in the two theories. For this reason T-dualities are much easier to test, and were historically the first to be discovered.
13. Actually this property does not hold for all BPS states, but holds for a special class of BPS states. The analysis described in the text can be applied only to this special class of BPS states.
14. It is a consequence of supersymmetry that the number of BPS states in this theory always comes as a multiple of 16.
15. These parameters include the coupling constant, as well as the shape and size of the compact manifold if we are considering the case where some of the space-like dimensions have been compactified. Like the coupling constant in string theory, each

of these parameters in M-theory can be identified as the value of a scalar field of the theory.

. At a finite but large value of the coupling constant the theory behaves like the eleven dimensional supergravity theory with one of its dimensions compactified to a circle of large radius. As the coupling constant of the Type-IIA string theory decreases, the radius of the corresponding circle also decreases.

. Since this limit corresponds to strong coupling limit of Type-IIA string theory, no direct calculation is possible. But supersymmetry non-renormalization theorems guarantee that certain quantities computed at small value of the coupling can be trusted even for large value of the coupling. The comparison is done only for these quantities.

1. For a review and other references see A. Sen, An Introduction to Non-Perturbative String Theory, hep-th/9802051 (can be downloaded from http://xxx.lanl.gov/abs/hep-th/9802051).

ACKNOWLEDGEMENT. I wish to thank P. Durganandini, S. Mukhi and S. Rao for their comments on the manuscript.