Chapter 1

A Brief History of String Theory

To help introduce the topics which follow and their significance in high energy physics, in this chapter we will briefly give a non-technical historical account of the development of string theory to date, focusing on its achievements, failures and prospects. This will also help to motivate the vast interest in string theory within the particle theory community. It will further give an overview of the material which will follow.

In conventional quantum field theory, the fundamental objects are mathematical points in spacetime, modeling the elementary point particles of nature. String theory is a rather radical generalization of quantum field theory whereby the fundamental objects are extended one-dimensional lines or loops (Fig. 1.1). The various elementary particles observed in nature correspond to different vibrational modes of the string. While we cannot see a string (yet) in nature, if we are very far away from it we will be able to see its point-like oscillations, and hence measure the elementary particles that it produces. The main advantage of this description is that while there are many particles, there is only one string. This indicates that strings could serve as a good starting point for a unified field theory of the fundamental interactions.

This is the idea that emerged by the end of the 1960s from several years of intensive studies of dual models of hadron resonances [Veneziano (1968)]. In this setting, string theory attempts to describe the strong nuclear force. The excitement over this formalism arose from the fact that string S-matrix scattering amplitudes agreed with those found in meson scattering experiments at the time. The inclusion of fermions into the model led to the notion of a supersymmetric string, or "superstring" for short [Neveu and Schwarz (1971); Ramond (1971)]. The massive particles sit on "Regge trajectories" in this setting.

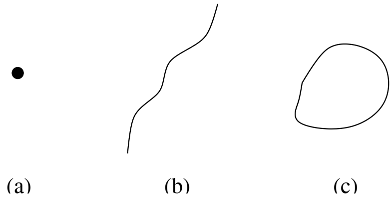

Fig. 1.1 (a) A point particle. (b) An open string. (c) A closed string.

However, around 1973 the interest in string theory quickly began to fade, mainly because quantum chromodynamics became recognized as the correct quantum field theory of the strong interactions. In addition, string theories possessed various undesirable features which made them inappropriate for a theory of hadrons. Among these were the large number of extra spacetime dimensions demanded by string theory, and the existence of massless particles other than the spin 1 gluon in the spectrum of string states.

In 1974 the interest in string theory was revived for another reason [Scherk and Schwarz (1974); Yoneya (1974)]. It was found that, among the massless string states, there is a spin 2 particle that interacts like a graviton. In fact, the only consistent interactions of massless spin 2 particles are gravitational interactions. Thus string theory naturally includes general relativity, and it was thereby proposed as a unified theory of the fundamental forces of nature, including gravity, rather than a theory of hadrons. This situation is in marked contrast to that in ordinary quantum field theory, which does not allow gravity to exist because its scattering amplitudes that involve graviton exchanges are severely plagued by non-renormalizable ultraviolet divergences (Fig. 1.2). On the other hand, string theory is a consistent quantum theory, free from ultraviolet divergences, which necessarily *requires* gravitation for its overall consistency.

With these facts it is possible to estimate the energy or length scale at which strings should be observed in nature. Since string theory is a relativistic quantum theory that includes gravity, it must involve the corresponding three fundamental constants, namely the speed of light c, the reduced Planck constant $\hbar$, and the Newtonian gravitational constant G. These three constants may combined into a constant with dimensions of length. The characteristic length scale of strings may thereby be estimated

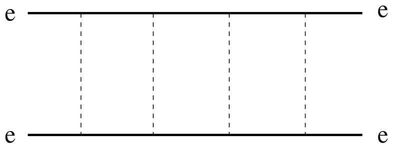

Fig. 1.2 A non-renormalizable ultraviolet divergent Feynman diagram in quantum gravity. The dashed lines depict graviton exchanges.

by the *Planck length* of quantum gravity:

$$\boxed{\ell_{\mathrm{P}} = \left(\frac{\hbar G}{c^3}\right)^{3/2} = 1.6 \times 10^{-33}\ \mathrm{cm}\ .} \tag{1.1}$$

This is to be compared with the typical size of hadrons, which is of the order of 10^{-13} cm. The corresponding energy scale is known as the *Planck mass*:

$$\boxed{m_{\mathrm{P}} = \left(\frac{\hbar c}{G}\right)^{1/2} = 1.2 \times 10^{19}\ \mathrm{GeV}/c^2\ .} \tag{1.2}$$

These scales indicate the reasons why strings have not been observed in nature thus far. Present day particle accelerators run at energies $\ll m_{\mathrm{P}}c^2$ and thus cannot resolve distances as short as the Planck length. At such energies, strings look like point particles, because at very large distance scales compared to the Planck length all one can observe is the string's center of mass motion, which is point-like. Thus at these present day scales, strings are accurately described by quantum field theory.

For many of the subsequent years superstring theory began showing great promise as a unified quantum theory of all the fundamental forces including gravity. Some of the general features which were discovered are:

- General relativity gets modified at very short distances/high energies (below the Planck scale), but at ordinary distances and energies it is present in string theory in exactly the same form as Einstein's theory.

- "Standard model type" Yang–Mills gauge theories arise very naturally in string theory. However, the reasons why the gauge group $SU(3) \times SU(2) \times U(1)$ of the standard model should be singled out is not yet fully understood.
- String theory predicts supersymmetry, because its mathematical consistency depends crucially on it. This is a generic feature of string theory that has not yet been discovered experimentally.

This was the situation for some years, and again the interest in string theory within the high energy physics community began to fade. Different versions of superstring theory existed, but none of them resembled very closely the structure of the standard model.

Things took a sharp turn in 1985 with the birth of what is known as the "first superstring revolution". The dramatic achievement at this time was the realization of how to cancel certain mathematical inconsistencies in quantum string theory. This is known as Green–Schwarz anomaly cancellation [Green and Schwarz (1984)] and its main consequence is that it leaves us with only five consistent superstring theories, each living in ten spacetime dimensions. These five theories are called Type I, Type IIA, Type IIB, $SO(32)$ heterotic, and $E_8 \times E_8$ heterotic. The terminology will be explained later on. For now, we simply note the supersymmetric Yang–Mills gauge groups that arise in these theories. The Type I theories have gauge group $SO(32)$, both Type II theories have $U(1)$, and the heterotic theories have gauge groups as in their names. Of particular phenomenological interest was the $E_8 \times E_8$ heterotic string, because from it one could construct grand unified field theories starting from the exceptional gauge group E_6.

The spacetime dimensionality problem is reconciled through the notion of "compactification". Putting six of the spatial directions on a "small" six-dimensional compact space, smaller than the resolution of the most powerful microscope, makes the 9+1 dimensional spacetime look 3+1 dimensional, as in our observable world. The six dimensional manifolds are restricted by string dynamics to be "Calabi–Yau spaces" [Candelas *et al* (1985)]. These compactifications have tantalizingly similar features to the standard model. However, no complete quantitative agreement has been found yet between the two theories, such as the masses of the various elementary particles. This reason, and others, once again led to the demise of string theory towards the end of the 1980s. Furthermore, at that stage one only understood how to formulate superstring theories in terms of divergent perturbation series analogous to quantum field theory. Like in quantum chromodynamics, it is unlikely that a realistic vacuum can be accurately analysed within perturbation theory. Without a good understanding of nonperturbative effects (such as the analogs of QCD instantons), superstring theory cannot give explicit, quantitative predictions for a grand unified model.

This was the state of affairs until around 1995 when the "second superstring revolution" set in. For the first time, it was understood how to go beyond the perturbation expansion of string theory via "dualities" which probe nonperturbative features of string theory [Font *et al* (1990); Hull and Townsend (1995); Kachru and Vafa (1995); Schwarz (1995); Sen (1994)]. The three major implications of these discoveries were:

- *Dualities relate all five superstring theories in ten dimensions to one another.*

The different theories are just perturbative expansions of a unique underlying theory $\mathcal{U}$ about five different, consistent quantum vacua [Schwarz (1996); Schwarz (1997)]. Thus there is a completely unique theory of nature, whose equation of motion admits many vacua. This is of course a most desirable property of a unified theory.

- *The theory $\mathcal{U}$ also has a solution called "M-Theory" which lives in 11 spacetime dimensions [Duff (1996); Townsend (1995); Witten (1995)].*

The low-energy limit of M-Theory is 11-dimensional supergravity [Cremmer, Julia and Scheck (1978)]. All five superstring theories can be thought of as originating from M-Theory [Duff (1996); Schwarz (1997)] (see Fig. 1.3). The underlying theory $\mathcal{U}$ is depicted in Fig. 1.4.

- *In addition to the fundamental strings, the theory $\mathcal{U}$ admits a variety of extended nonperturbative excitations called "p-branes", where p is the number of spatial extensions of the objects [Horowitz and Strominger (1991)].*

Especially important in this regard are the "Dirichlet p-branes" [Dai, Leigh and Polchinski (1989); Hořava (1989); Polchinski (1995)], or "D-branes" for short, which are p-dimensional soliton-like hyperplanes in spacetime whose quantum dynamics are governed by the theory of *open strings* whose ends are constrained to move on them (Fig. 1.5).

We will not attempt any description of the theory $\mathcal{U}$, which at present is not very well understood. Rather, we wish to focus on the remarkable

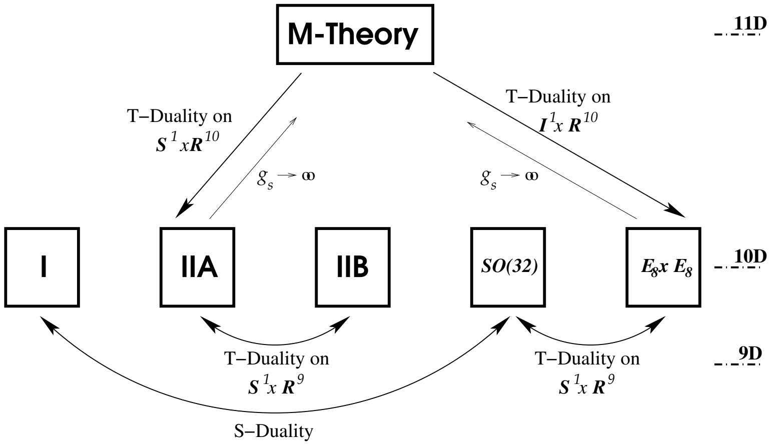

Fig. 1.3 The various duality transformations that relate the superstring theories in nine and ten dimensions. T-Duality inverts the radius R of the circle $\mathbf{S}^1$, or the length of the finite interval $\mathbf{I}^1$, along which a single direction of the spacetime is compactified, i.e. $R \mapsto \ell_{\rm P}^2/R$. S-duality inverts the (dimensionless) string coupling constant g_s, $g_s \mapsto 1/g_s$, and is the analog of electric-magnetic duality (or strong-weak coupling duality) in four-dimensional gauge theories. M-Theory originates as the strong coupling limit of either the Type IIA or $E_8 \times E_8$ heterotic string theories.

impact in high-energy physics that the discovery of D-branes has provided. Amongst other things, they have led to:

- Explicit realizations of nonperturbative string dualities [Polchinski (1995)]. For example, an elementary closed string state in Theory A (which is perturbative because its amplitudes are functions of the string coupling g_s) gets mapped under an S-duality transformation to a D-brane state in the dual Theory B (which depends on $1/g_s$ and is therefore nonperturbative).
- A microscopic explanation of black hole entropy and the rate of emission of thermal (Hawking) radiation for black holes in string theory [Callan and Maldacena (1996); Strominger and Vafa (1996)].
- The gauge theory/gravity (or AdS/CFT) correspondence [Aharony *et al* (2000); Maldacena (1998)]. D-branes carry gauge fields, while on the other hand they admit a dual description as solutions of the classical equations of motion of string theory and supergravity. Demanding that these two descriptions be equivalent implies, for

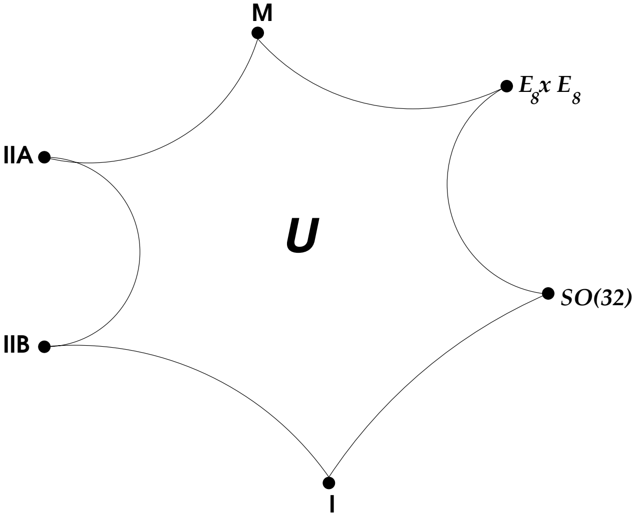

Fig. 1.4 The space $\mathcal{U}$ of quantum string vacua. At each node a weakly-coupled string description is possible.

some special cases, that string theory is equivalent to a gauge field theory. This is an explicit realization of the old ideas that Yang–Mills theory may be represented as some sort of string theory.

- Probes of short-distances in spacetime [Douglas *et al* (1997)], where quantum gravitational fluctuations become important and classical general relativity breaks down.
- Large radius compactifications, whereby extra compact dimensions of size $\gg (\text{TeV})^{-1}$ occur [Antoniadis *et al* (1998); Arkani-Hamed, Dimopoulos and Dvali (1998)]. This is the distance scale probed in present-day accelerator experiments, which has led to the hope that the extra dimensions required by string theory may actually be observable.
- Brane world scenarios, in which we model our world as a D-brane [Randall and Sundrum (1999a); Randall and Sundrum (1999b)]. This may be used to explain why gravity couples so weakly to matter, i.e. why the effective Planck mass in our 3+1-dimensional world is so large, and hence gives a potential explanation of the hierarchy problem $m_{\rm P} \gg m_{\rm weak}$.

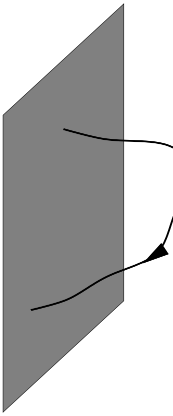

Fig. 1.5 A fundamental open string (wavy line) starting and ending (with Dirichlet boundary conditions) on a Dp-brane (shaded region) which is a $p+1$-dimensional hyperplane in spacetime.

In what follows we will give the necessary background into the description of D-brane dynamics which leads to these exciting developments in theoretical high-energy physics.