1
Introduction

There were two major breakthroughs that revolutionized theoretical physics in the twentieth century: general relativity and quantum mechanics. General relativity is central to our current understanding of the large-scale expansion of the Universe. It gives small corrections to the predictions of Newtonian gravity for the motion of planets and the deflection of light rays, and it predicts the existence of gravitational radiation and black holes. Its description of the gravitational force in terms of the curvature of space-time has fundamentally changed our view of space and time: they are now viewed as dynamical. Quantum mechanics, on the other hand, is the essential tool for understanding microscopic physics. The evidence continues to build that it is an exact property of Nature. Certainly, its exact validity is a basic assumption in all string theory research.

The understanding of the fundamental laws of Nature is surely incomplete until general relativity and quantum mechanics are successfully reconciled and unified. That this is very challenging can be seen from many different viewpoints. The concepts, observables and types of calculations that characterize the two subjects are strikingly different. Moreover, until about 1980 the two fields developed almost independently of one another. Very few physicists were experts in both. With the goal of unifying both subjects, string theory has dramatically altered the sociology as well as the science.

In relativistic quantum mechanics, called quantum field theory, one requires that two fields that are defined at space-time points with a space-like separation should commute (or anticommute if they are fermionic). In the gravitational context one doesn't know whether or not two space-time points have a space-like separation until the metric has been computed, which is part of the dynamical problem. Worse yet, the metric is subject to quantum fluctuations just like other quantum fields. Clearly, these are rather challenging issues. Another set of challenges is associated with the quantum description of black holes and the description of the Universe in the very early stages of its history.

The most straightforward attempts to combine quantum mechanics and general relativity, in the framework of perturbative quantum field theory, run into problems due to uncontrollable infinities. Ultraviolet divergences are a characteristic feature of radiative corrections to gravitational processes, and they become worse at each order in perturbation theory. Because Newton's constant is proportional to $(\text{length})^2$ in four dimensions, simple power-counting arguments show that it is not possible to remove these infinities by the conventional renormalization methods of quantum field theory. Detailed calculations demonstrate that there is no miracle that invalidates this simple dimensional analysis.[1]

String theory purports to overcome these difficulties and to provide a consistent quantum theory of gravity. How the theory does this is not yet understood in full detail. As we have learned time and time again, string theory contains many deep truths that are there to be discovered. Gradually a consistent picture is emerging of how this remarkable and fascinating theory deals with the many challenges that need to be addressed for a successful unification of quantum mechanics and general relativity.

1.1 Historical origins

String theory arose in the late 1960s in an attempt to understand the strong nuclear force. This is the force that is responsible for holding protons and neutrons together inside the nucleus of an atom as well as quarks together inside the protons and neutrons. A theory based on fundamental one-dimensional extended objects, called strings, rather than point-like particles, can account qualitatively for various features of the strong nuclear force and the strongly interacting particles (or hadrons).

The basic idea in the string description of the strong interactions is that specific particles correspond to specific oscillation modes (or quantum states) of the string. This proposal gives a very satisfying unified picture in that it postulates a single fundamental object (namely, the string) to explain the myriad of different observed hadrons, as indicated in Fig. 1.1.

In the early 1970s another theory of the strong nuclear force – called quantum chromodynamics (or QCD) – was developed. As a result of this, as well as various technical problems in the string theory approach, string

1 Some physicists believe that perturbative renormalizability is not a fundamental requirement and try to "quantize" pure general relativity despite its nonrenormalizability. Loop quantum gravity is an example of this approach. Whatever one thinks of the logic, it is fair to say that despite a considerable amount of effort such attempts have not yet been very fruitful.

theory fell out of favor. The current viewpoint is that this program made good sense, and so it has again become an active area of research. The concrete string theory that describes the strong interaction is still not known, though one now has a much better understanding of how to approach the problem.

String theory turned out to be well suited for an even more ambitious purpose: the construction of a quantum theory that unifies the description of gravity and the other fundamental forces of nature. In principle, it has the potential to provide a complete understanding of particle physics and of cosmology. Even though this is still a distant dream, it is clear that in this fascinating theory surprises arise over and over.

1.2 General features

Even though string theory is not yet fully formulated, and we cannot yet give a detailed description of how the standard model of elementary particles should emerge at low energies, or how the Universe originated, there are some general features of the theory that have been well understood. These are features that seem to be quite generic irrespective of what the final formulation of string theory might be.

Gravity

The first general feature of string theory, and perhaps the most important, is that general relativity is naturally incorporated in the theory. The theory gets modified at very short distances/high energies but at ordinary distances and energies it is present in exactly the form as proposed by Einstein. This is significant, because general relativity is arising within the framework of a

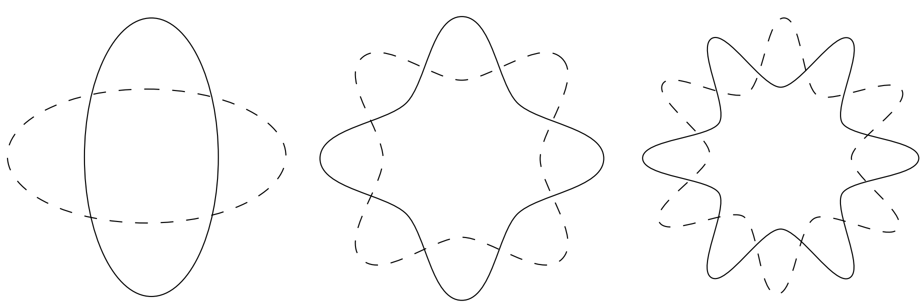

Fig. 1.1. Different particles are different vibrational modes of a string.

consistent quantum theory. Ordinary quantum field theory does not allow gravity to exist; string theory requires it.

Yang–Mills gauge theory

In order to fulfill the goal of describing all of elementary particle physics, the presence of a graviton in the string spectrum is not enough. One also needs to account for the standard model, which is a Yang–Mills theory based on the gauge group $SU(3)\times SU(2)\times U(1)$. The appearance of Yang–Mills gauge theories of the sort that comprise the standard model is a general feature of string theory. Moreover, matter can appear in complex chiral representations, which is an essential feature of the standard model. However, it is not yet understood why the specific $SU(3) \times SU(2) \times U(1)$ gauge theory with three generations of quarks and leptons is singled out in nature.

Supersymmetry

The third general feature of string theory is that its consistency requires supersymmetry, which is a symmetry that relates bosons to fermions is required. There exist nonsupersymmetric bosonic string theories (discussed in Chapters 2 and 3), but lacking fermions, they are completely unrealistic. The mathematical consistency of string theories with fermions depends crucially on local supersymmetry. Supersymmetry is a generic feature of all potentially realistic string theories. The fact that this symmetry has not yet been discovered is an indication that the characteristic energy scale of supersymmetry breaking and the masses of supersymmetry partners of known particles are above experimentally determined lower bounds.

Space-time supersymmetry is one of the major predictions of superstring theory that could be confirmed experimentally at accessible energies. A variety of arguments, not specific to string theory, suggest that the characteristic energy scale associated with supersymmetry breaking should be related to the electroweak scale, in other words in the range 100 GeV to a few TeV. If this is correct, superpartners should be observable at the CERN Large Hadron Collider (LHC), which is scheduled to begin operating in 2007.

Extra dimensions of space

In contrast to many theories in physics, superstring theories are able to predict the dimension of the space-time in which they live. The theory is only consistent in a ten-dimensional space-time and in some cases an eleventh dimension is also possible.

To make contact between string theory and the four-dimensional world of everyday experience, the most straightforward possibility is that six or seven of the dimensions are compactified on an internal manifold, whose size is sufficiently small to have escaped detection. For purposes of particle physics, the other four dimensions should give our four-dimensional space-time. Of course, for purposes of cosmology, other (time-dependent) geometries may also arise.

Fig. 1.2. From far away a two-dimensional cylinder looks one-dimensional.

The idea of an extra compact dimension was first discussed by Kaluza and Klein in the 1920s. Their goal was to construct a unified description of electromagnetism and gravity in four dimensions by compactifying five-dimensional general relativity on a circle. Even though we now know that this is not how electromagnetism arises, the essence of this beautiful approach reappears in string theory. The Kaluza–Klein idea, nowadays referred to as *compactification*, can be illustrated in terms of the two cylinders of Fig. 1.2. The surface of the first cylinder is two-dimensional. However, if the radius of the circle becomes extremely small, or equivalently if the cylinder is viewed from a large distance, the cylinder looks effectively one-dimensional. One now imagines that the long dimension of the cylinder is replaced by our four-dimensional space-time and the short dimension by an appropriate six, or seven-dimensional compact manifold. At large distances or low energies the compact internal space cannot be seen and the world looks effectively four-dimensional. As discussed in Chapters 9 and 10, even if the internal manifolds are invisible, their topological properties determine the particle content and structure of the four-dimensional theory. In the mid-1980s Calabi–Yau manifolds were first considered for compactifying six extra dimensions, and they were shown to be phenomenologically rather promising, even though some serious drawbacks (such as the moduli space problem discussed in Chapter 10) posed a problem for the predictive power of string theory. In contrast to the circle, Calabi–Yau manifolds do not have isometries, and part of their role is to break symmetries rather than to make them.

The size of strings

In conventional quantum field theory the elementary particles are mathematical points, whereas in perturbative string theory the fundamental objects are one-dimensional loops (of zero thickness). Strings have a characteristic length scale, denoted l_s, which can be estimated by dimensional analysis. Since string theory is a relativistic quantum theory that includes gravity it must involve the fundamental constants c (the speed of light), $\hbar$ (Planck's constant divided by 2π), and G (Newton's gravitational constant). From these one can form a length, known as the Planck length

$$l_{\rm p} = \left(\frac{\hbar G}{c^3}\right)^{1/2} = 1.6 \times 10^{-33}\,\text{cm}.$$

Similarly, the Planck mass is

$$m_{\rm p} = \left(\frac{\hbar c}{G}\right)^{1/2} = 1.2 \times 10^{19}\,\text{GeV}/c^2.$$

The Planck scale is the natural first guess for a rough estimate of the fundamental string length scale as well as the characteristic size of compact extra dimensions. Experiments at energies far below the Planck energy cannot resolve distances as short as the Planck length. Thus, at such energies, strings can be accurately approximated by point particles. This explains why quantum field theory has been so successful in describing our world.

1.3 Basic string theory

As a string evolves in time it sweeps out a two-dimensional surface in spacetime, which is called the string *world sheet* of the string. This is the string counterpart of the world line for a point particle. In quantum field theory, analyzed in perturbation theory, contributions to amplitudes are associated with Feynman diagrams, which depict possible configurations of world lines. In particular, interactions correspond to junctions of world lines. Similarly, perturbation expansions in string theory involve string world sheets of various topologies.

The existence of interactions in string theory can be understood as a consequence of world-sheet topology rather than of a local singularity on the world sheet. This difference from point-particle theories has two important implications. First, in string theory the structure of interactions is uniquely determined by the free theory. There are no arbitrary interactions to be chosen. Second, since string interactions are not associated with short-distance singularities, string theory amplitudes have no ultraviolet divergences. The string scale $1/l_s$ acts as a UV cutoff.

World-volume actions and the critical dimension

A string can be regarded as a special case of a p-brane, which is an object with p spatial dimensions and tension (or energy density) T_p. In fact, various p-branes do appear in superstring theory as nonperturbative excitations. The classical motion of a p-brane extremizes the $(p+1)$-dimensional volume V that it sweeps out in space-time. Thus there is a p-brane action that is given by $S_p = -T_p V$. In the case of the fundamental string, which has $p = 1$, V is the area of the string world sheet and the action is called the Nambu–Goto action.

Classically, the Nambu–Goto action is equivalent to the string sigma-model action

$$S_\sigma = -\frac{T}{2}\int \sqrt{-h}h^{\alpha\beta}\eta_{\mu\nu}\partial_\alpha X^\mu \partial_\beta X^\nu d\sigma d\tau,$$

where $h_{\alpha\beta}(\sigma,\tau)$ is an auxiliary world-sheet metric, $h = \det h_{\alpha\beta}$, and $h^{\alpha\beta}$ is the inverse of $h_{\alpha\beta}$. The functions $X^\mu(\sigma,\tau)$ describe the space-time embedding of the string world sheet. The Euler–Lagrange equation for $h^{\alpha\beta}$ can be used to eliminate it from the action and recover the Nambu–Goto action.

Quantum mechanically, the story is more subtle. Instead of eliminating h via its classical field equations, one should perform a Feynman path integral, using standard machinery to deal with the local symmetries and gauge fixing. When this is done correctly, one finds that there is a conformal anomaly unless the space-time dimension is $D = 26$. These matters are explored in Chapters 2 and 3. An analogous analysis for superstrings gives the critical dimension $D = 10$.

Closed strings and open strings

The parameter τ in the embedding functions $X^\mu(\sigma,\tau)$ is the world-sheet time coordinate and σ parametrizes the string at a given world-sheet time. For a closed string, which is topologically a circle, one should impose periodicity in the spatial parameter σ. Choosing its range to be π one identifies both ends of the string $X^\mu(\sigma, \tau) = X^\mu(\sigma + \pi, \tau)$. All string theories contain closed strings, and the graviton always appears as a massless mode in the closed-string spectrum of critical string theories.

For an open string, which is topologically a line interval, each end can be required to satisfy either Neumann or Dirichlet boundary conditions (for each value of μ). The Dirichlet condition specifies a space-time hypersurface on which the string ends. The only way this makes sense is if the open string ends on a physical object, which is called a D-brane. (D stands for Dirichlet.) If all the open-string boundary conditions are Neumann, then the ends of the string can be anywhere in the space-time. The modern interpretation is that this means that space-time-filling D-branes are present.

Perturbation theory

Perturbation theory is useful in a quantum theory that has a small dimensionless coupling constant, such as quantum electrodynamics (QED), since it allows one to compute physical quantities as expansions in the small parameter. In QED the small parameter is the fine-structure constant $\alpha \sim 1/137$. For a physical quantity $T(\alpha)$, one computes (using Feynman diagrams)

$$T(\alpha) = T_0 + \alpha T_1 + \alpha^2 T_2 + \ldots$$

Perturbation series are usually asymptotic expansions with zero radius of convergence. Still, they can be useful, if the expansion parameter is small, because the first terms in the expansion provide an accurate approximation.

The heterotic and type II superstring theories contain oriented closed strings only. As a result, the only world sheets in their perturbation expansions are closed oriented Riemann surfaces. There is a unique world-sheet topology at each order of the perturbation expansion, and its contribution is UV finite. The fact that there is just one string theory Feynman diagram at each order in the perturbation expansion is in striking contrast to the large number of Feynman diagrams that appear in quantum field theory. In the case of string theory there is no particular reason to expect the coupling constant $g_{\rm s}$ to be small. So it is unlikely that a realistic vacuum could be analyzed accurately using only perturbation theory. For this reason, it is important to understand nonperturbative effects in string theory.

Superstrings

The *first superstring revolution* began in 1984 with the discovery that quantum mechanical consistency of a ten-dimensional theory with $\mathcal{N} = 1$ supersymmetry requires a local Yang–Mills gauge symmetry based on one of two possible Lie algebras: $SO(32)$ or $E_8 \times E_8$. As is explained in Chapter 5, only for these two choices do certain quantum mechanical anomalies cancel. The fact that only these two groups are possible suggested that string theory has a very constrained structure, and therefore it might be very predictive. [2]

When one uses the superstring formalism for both left-moving modes and right-moving modes, the supersymmetries associated with the left-movers and the right-movers can have either opposite handedness or the same handedness. These two possibilities give different theories called the type IIA and type IIB superstring theories, respectively. A third possibility, called type I superstring theory, can be derived from the type IIB theory by modding out by its left–right symmetry, a procedure called orientifold projection. The strings that survive this projection are unoriented. The type I and type II superstring theories are described in Chapters 4 and 5 using formalisms with world-sheet and space-time supersymmetry, respectively.

A more surprising possibility is to use the formalism of the 26-dimensional bosonic string for the left-movers and the formalism of the 10-dimensional superstring for the right-movers. The string theories constructed in this way are called "heterotic." Heterotic string theory is discussed in Chapter 7. The mismatch in space-time dimensions may sound strange, but it is actually exactly what is needed. The extra 16 left-moving dimensions must describe a torus with very special properties to give a consistent theory. There are precisely two distinct tori that have the required properties, and they correspond to the Lie algebras $SO(32)$ and $E_8 \times E_8$.

Altogether, there are five distinct superstring theories, each in ten dimensions. Three of them, the type I theory and the two heterotic theories, have $\mathcal{N} = 1$ supersymmetry in the ten-dimensional sense. The minimal spinor in ten dimensions has 16 real components, so these theories have 16 conserved supercharges. The type I superstring theory has the gauge group $SO(32)$, whereas the heterotic theories realize both $SO(32)$ and $E_8 \times E_8$. The other two theories, type IIA and type IIB, have $\mathcal{N} = 2$ supersymmetry or equivalently 32 supercharges.

1.4 Modern developments in superstring theory

The realization that there are five different superstring theories was somewhat puzzling. Certainly, there is only one Universe, so it would be most satisfying if there were only one possible theory. In the late 1980s it was

2 Anomaly analysis alone also allows $U(1)^{496}$ and $E_8 \times U(1)^{248}$. However, there are no string theories with these gauge groups.

realized that there is a property known as T-duality that relates the two type II theories and the two heterotic theories, so that they shouldn't really be regarded as distinct theories.

Progress in understanding nonperturbative phenomena was achieved in the 1990s. Nonperturbative S-dualities and the opening up of an eleventh dimension at strong coupling in certain cases led to new identifications. Once all of these correspondences are taken into account, one ends up with the best possible conclusion: there is a unique underlying theory. Some of these developments are summarized below and are discussed in detail in the later chapters.

T-duality

String theory exhibits many surprising properties. One of them, called T-duality, is discussed in Chapter 6. T-duality implies that in many cases two different geometries for the extra dimensions are physically equivalent! In the simplest example, a circle of radius R is equivalent to a circle of radius ℓ_s^2/R, where (as before) ℓ_s is the fundamental string length scale.

T-duality typically relates two different theories. For example, it relates the two type II and the two heterotic theories. Therefore, the type IIA and type IIB theories (also the two heterotic theories) should be regarded as a single theory. More precisely, they represent opposite ends of a continuum of geometries as one varies the radius of a circular dimension. This radius is not a parameter of the underlying theory. Rather, it arises as the vacuum expectation value of a scalar field, and it is determined dynamically.

There are also fancier examples of duality equivalences. For example, there is an equivalence of type IIA superstring theory compactified on a Calabi–Yau manifold and type IIB compactified on the "mirror" Calabi–Yau manifold. This mirror pairing of topologically distinct Calabi–Yau manifolds is discussed in Chapter 9. A surprising connection to T-duality will emerge.

S-duality

Another kind of duality – called S-duality – was discovered as part of the *second superstring revolution* in the mid-1990s. It is discussed in Chapter 8. S-duality relates the string coupling constant g_s to $1/g_s$ in the same way that T-duality relates R to ℓ_s^2/R. The two basic examples relate the type I superstring theory to the $SO(32)$ heterotic string theory and the type IIB superstring theory to itself. Thus, given our knowledge of the small g_s behavior of these theories, given by perturbation theory, we learn how