1 Introduction

1.1 Prehistory

The quest in physics has been historically dominated by unraveling the simplicity of the physical laws, moving more and more toward the elementary. Although this is not guaranteed to succeed indefinitely, it has been vindicated so far. The other organizing tendency of the human mind is toward "unification": finding a unique framework for describing seemingly disparate phenomena.

The physics of the late nineteenth and twentieth centuries is a series of discoveries and unifications. Maxwell unified electricity and magnetism. Einstein developed the general theory of relativity that unified the principle of relativity and gravity. In the late 1940s, there was a culmination of two decades' efforts in the unification of electromagnetism and quantum mechanics. In the 1960s and 1970s, the theory of weak and electromagnetic interactions was also unified. Moreover, around the same period there was also a wider conceptual unification. Three of the four fundamental forces known were described by gauge theories. The fourth, gravity, is also based on a local invariance, albeit of a different type, and so far stands apart.[1] The combined theory, containing the quantum field theories of the electroweak and strong interactions together with the classical theory of gravity, formed the Standard Model of fundamental interactions. It is based on the gauge group $SU(3)\times SU(2)\times U(1)$. Its spin-1 gauge bosons mediate the strong and electroweak interactions. The matter particles are quarks and leptons of spin $\frac{1}{2}$ in three copies (known as generations and differing widely in mass), and a spin-0 particle, the Higgs boson, still experimentally elusive, that is responsible for the spontaneous breaking of the electroweak gauge symmetry.

[1] Today, we have some intriguing evidence that even gravity may be a strong-coupling facet of an extra underlying four-dimensional gauge theory.

The Standard Model has been experimentally tested and has survived thirty years of accelerator experiments.[2] This highly successful theory, however, is not satisfactory:

- A classical theory, namely, gravity, described by general relativity, must be added to the Standard Model in order to agree with experimental data. This theory is not renormalizable at the quantum level. In other words, new input is needed in order to understand its high-energy behavior. This has been a challenge to the physics community since the 1930s and (apart from string theory) very little has been learned on this subject since then.

- The three SM interactions are not completely unified. The gauge group is semisimple. Gravity seems even further from unification with the gauge theories. A related problem is that the Standard Model contains many parameters that look *a priori* arbitrary.

- The model is unstable as we increase the energy (hierarchy problem of mass scales) and the theory loses predictivity as one starts moving far from current accelerator energies and closer to the Planck scale. Gauge bosons are protected from destabilizing corrections because of gauge invariance. The fermions are equally protected due to chiral symmetries. The real culprit is the Higgs boson.

Several attempts have been made to improve on the problems above.

The first attempts focused on improving on unification. They gave rise to the grand unified theories (GUTs). All interactions were collected in a simple group SU(5) in the beginning, but also SO(10), E_6, and others. The fermions of a given generation were organized in the (larger) representations of the GUT group. There were successes in this endeavor, including the prediction of $\sin^2\theta_W$ and the prediction of light right-handed neutrinos in some GUTs. However, there was a need for Higgs bosons to break the GUT symmetry to the SM group and the hierarchy problem took its toll by making it technically impossible to engineer a light electroweak Higgs.

The physics community realized that the focus must be on bypassing the hierarchy problem. A first idea attacked the problem at its root: it attempted to banish the Higgs boson as an elementary state and to replace it with extra fermionic degrees of freedom. It introduced a new gauge interaction (termed technicolor) which bounds these fermions strongly; one of the techni-hadrons should have the right properties to replace the elementary Higgs boson as responsible for the electroweak symmetry breaking. The negative side of this line of thought is that it relied on the nonperturbative physics of the technicolor interaction. Realistic model building turned out to be difficult and eventually this line of thought was mostly abandoned.

A competing idea relied on a new type of symmetry, supersymmetry, that connects bosons to fermions. This property turned out to be essential since it could force the bad-mannered spin-0 bosons to behave as well as their spin-$\frac{1}{2}$ partners. This works well, but supersymmetry stipulated that each SM fermion must have a spin-0 superpartner with equal mass. This being obviously false, supersymmetry must be spontaneously broken at

[2] With the exception of the neutrino sector that was suspected to be incomplete and is currently the source of interesting discoveries.

an energy scale not far away from today's accelerator energies. Further analysis indicated that the breaking of global supersymmetry produced superpartners whose masses were correlated with those of the already known particles, in conflict with experimental data.

To avoid such constraints global supersymmetry needed to be promoted to a local symmetry. As a supersymmetry transformation is in a sense the square root of a translation, this entailed that a theory of local supersymmetry must also incorporate gravity. This theory was first constructed in the late 1970s, and was further generalized to make model building possible. The flip side of this was that the inclusion of gravity opened the Pandora's box of nonrenormalizability of the theory. Hopes that (extended) supergravity might be renormalizable soon vanished.

In parallel with the developments above, a part of the community resurrected the old idea of Kaluza and Klein of unifying gravity with the other gauge interactions. If one starts from a higher-dimensional theory of gravity and compactifies the theory to four dimensions, one ends with four-dimensional gravity plus extra gauge interactions. Although gravity in higher dimensions is more singular in the UV, physicists hoped that at least at the classical level one would get a theory that is very close to the Standard Model. Although progress was made, a stumbling block turned out to be obtaining a four-dimensional chiral spectrum of fermions as in the SM.

Although none of the directions above provided a final and successful theory, the ingredients were very interesting ideas that many felt would form a part of the ultimate theory.

1.2 The Case for String Theory

String theory has been the leading candidate over the past two decades for a theory that consistently unifies all fundamental forces of nature, including gravity. It gained popularity because it provides a theory that is UV finite.[3]

The basic characteristic of the theory is that its elementary constituents are extended strings rather than pointlike particles as in quantum field theory. This makes the theory much more complicated than QFT, but at the same time it imparts some unique properties.

One of the key ingredients of string theory is that it provides a finite theory of quantum gravity, at least in perturbation theory. To appreciate the difficulties with the quantization of Einstein gravity, we look at a single-graviton exchange between two particles (Fig. 1.1a)). Then, the amplitude is proportional to E^2/M_P^2, where E is the energy of the process and M_P is the Planck mass, $M_P \sim 10^{19}$ GeV. It is related to the Newton constant as

$$M_P^2 = \frac{1}{16\pi G_N}. \tag{1.2.1}$$

[3] Although there is no rigorous proof to all orders that the theory is UV finite, there are several all-orders arguments as well as rigorous results at low-loop order. In closed string theory, amplitudes must be carefully defined via analytic continuation, standard in S-matrix theory. When open strings are present, there are divergences. However, they are interpreted as IR divergences (due to the exchange of massless states) in the dual closed string channel. They are subtracted in the "Wilsonian" S-matrix elements.

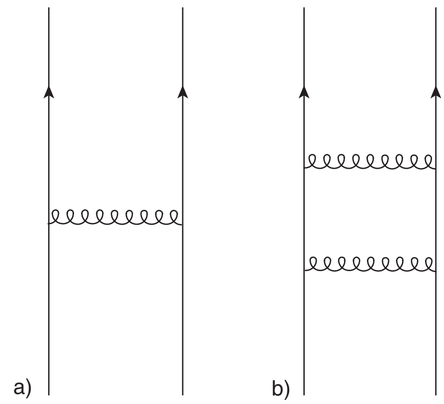

Figure 1.1 Gravitational interaction between two particles via graviton exchange.

Therefore, the gravitational interaction is irrelevant in the IR ($E \ll M_P$) but strongly relevant in the UV. In particular, this implies that the two-graviton exchange diagram (Fig. 1.1b)) is proportional to the dimensionless ratio

$$\frac{E^2}{M_P^4}\int_0^{\Lambda} d\tilde{E}\;\tilde{E} \sim \frac{\Lambda^2 E^2}{M_P^4}, \tag{1.2.2}$$

where E is the outgoing particle energy and $\tilde{E}$ is the internal particle energy. This is strongly UV divergent. It is known that Einstein gravity coupled to matter is nonrenormalizable in perturbation theory. Supersymmetry makes the UV divergence softer but the nonrenormalizability persists.

There are two ways out of this:

- There is a nontrivial UV fixed point that governs the high-energy behavior of quantum gravity. To date, no credible example of this possibility has been offered.
- There is new physics at $E \sim M_P$ (or even lower) and Einstein gravity is the IR limit of a more general theory, valid at and beyond the Planck scale. You could consider the analogous situation with the Fermi theory of weak interactions. There, a nonrenormalizable current-current interaction with similar problems occurred, but today we know that this is the IR limit of the standard weak interaction mediated by the $W^{\pm}$ and Z^0 gauge bosons. So far, there is no consistent field theory that can make sense at energies beyond M_P and contains gravity. Good reviews of the ultraviolet problems of Einstein gravity can be found in [1,2].

Strings provide a theory that induces new physics at the string scale M_s which in perturbation theory (the string coupling g_s being weak) is much lower than the Planck scale M_P. It is still true, however, that the string perturbation theory becomes uncontrollable when the energies approach the Planck scale.

There are two important reasons why closed string theory does not have UV divergences. One is the fact that string dynamics and interactions are inextricably linked to the geometry of two-dimensional surfaces. For example, for closed strings, by decomposing the string

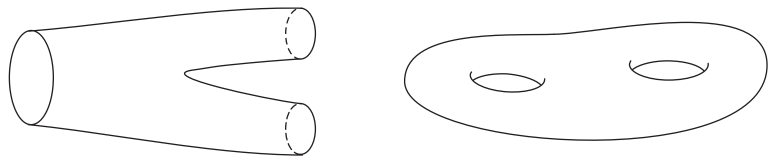

Figure 1.2 Open and closed Riemann surfaces.

Feynman diagrams, it is obvious that there is essentially a universal three-point interaction in the theory. This interaction is dictated by the two-dimensional geometry of closed Riemann surfaces as is obvious from figure 1.2. The other related reason is the presence of an infinite tower of excitations with masses in multiples of the string scale. Their interactions are carefully tuned to become soft at distances larger than the string length ℓ_s but still longer than the Planck length ℓ_P.

For open strings the situation is subtler. There, UV divergences are present, but are interpreted as IR closed-string divergences in the dual closed-string channel. This UV-IR open-closed string duality is at the heart of many of the recent developments in the field.

Another key ingredient of string theory is that it unifies gravity with gauge interactions. It does this in several different ways. The simplest is via the traditional KK approach. Superstring theory typically is defined in ten dimensions. Standard four-dimensional vacua can be obtained via compactification on a six-dimensional compact manifold. However, gauge symmetry can also arise from D-branes that sometimes are part of the vacuum (as in orientifolds). There is even gauge symmetry coming from a nongeometrical part of the theory as happens in the heterotic string. The unified origin of gravity and gauge symmetry extends even further to other interactions. For example, the Yukawa interactions, crucial for giving mass to the SM particles, are also intimately related to the gauge interactions.

Unlike earlier Kaluza-Klein approaches to unification, string theory is capable of providing, upon appropriate compactifications, chiral matter in four dimensions. This happens via a subtle interplay between anomaly-related interactions and the process of compactification.

Another characteristic ingredient of string theory is that the presence of space-time fermions in the theory implies the appearance of space-time supersymmetry at least at high energies. Supersymmetry, consequently, is an important ingredient of the theory.

There are good reasons to believe that the theory is unique, although there are many possible vacua that could be stable. Recent understanding of nonperturbative dualities has strengthened the belief in the uniqueness of this string theory structure. It has also pointed out new corners in the overall theory, which many call M-theory, that are still uncharted.

Despite all this, string theory, after thirty-five years of research, has many important questions still unanswered. Physicists feel that the fundamental definition of the theory is not known. How it fits the real world in detail is also not known presently.

It may be that string theory will turn into an "intellectual classical black hole." It may also be that it is the correct description of physics at short distances. Time and experiment will show.

1.3 A Stringy Historical Perspective

In the 1960s, physicists tried to make sense of a large amount of experimental data relevant to the strong interaction. There were lots of particles (or "resonances") and the situation could best be described as chaotic. Some regularities were observed:

Almost linear Regge behavior. It was noticed that a relatively large number of resonances could be nicely put on (almost) straight lines by plotting their mass versus their spin

$$m^2 = \frac{J}{\alpha'} + \alpha_0, \tag{1.3.1}$$

with $\alpha' \sim 1\ \text{GeV}^{-2}$; this relation was checked up to $J = 11/2$.

s-t duality. If we consider a scattering amplitude of two hadrons → two hadrons $(1, 2 \to 3, 4)$, then it can be described by the Mandelstam invariants

$$s = -(p_1 + p_2)^2, \quad t = -(p_2 + p_3)^2, \quad u = -(p_1 + p_3)^2, \tag{1.3.2}$$

with $s + t + u = \sum_i m_i^2$. We are using a metric with signature $(-+++)$. Such an amplitude depends on the flavor quantum numbers of hadrons (for example SU(3)). Consider the flavor part, which is cyclically symmetric in flavor space. For the full amplitude to be symmetric, it must also be cyclically symmetric in the momenta p_i. This symmetry amounts to the interchange $t \leftrightarrow s$. Thus, the amplitude should satisfy $A(s, t) = A(t, s)$. Consider a t-channel contribution due to the exchange of a spin-J particle of mass M. Then, at high energy

$$A_J(s, t) \sim \frac{(-s)^J}{t - M^2}. \tag{1.3.3}$$

Thus, this partial amplitude increases with s and its behavior becomes worse for large values of J. If one sews amplitudes of this form together to make a loop amplitude, then there are uncontrollable UV divergences for $J > 1$. Any finite sum of amplitudes of the form (1.3.3) has this bad UV behavior. Moreover, such a finite sum has no s-channel poles. However, if one allows an infinite number of terms then it is conceivable that the UV behavior might be different.

A proposal for such a dual amplitude was made by Veneziano [3],

$$A(s, t) = \frac{\Gamma(-\alpha(s))\Gamma(-\alpha(t))}{\Gamma(-\alpha(s) - \alpha(t))}, \tag{1.3.4}$$

where Γ is the standard Euler Γ-function and

$$\alpha(s) = \alpha(0) + \alpha' s. \tag{1.3.5}$$

By using the standard properties of the Γ-function, it can be checked that the amplitude (1.3.4) has an infinite number of s, t-channel poles:

$$A(s,t) = -\sum_{n=0}^{\infty} \frac{(\alpha(s)+1)\cdots(\alpha(s)+n)}{n!} \frac{1}{\alpha(t)-n}. \tag{1.3.6}$$

In this expansion, the $s \leftrightarrow t$ interchange symmetry of (1.3.4) is not manifest. The poles in (1.3.6) correspond to the exchange of an infinite number of particles of mass $M^2 = \frac{(n-\alpha(0))}{\alpha'}$ and high spins. It can also be checked that the high-energy behavior of the Veneziano amplitude is softer than any local quantum field theory amplitude, and the infinite number of poles is crucial for this.

It was subsequently realized by Nambu, Goto, Nielsen, and Susskind that such amplitudes came out of theories of relativistic strings. However, string theories had several shortcomings in explaining the dynamics of strong interactions.

- All of them seemed to predict a particle with negative mass squared, the tachyon.
- Several of them seemed to contain a massless spin-2 particle that was impossible to get rid of.
- All of them seemed to require a space-time dimension of 26 in order not to break Lorentz invariance at the quantum level.
- They contained only bosons.

At the same time, experimental data from SLAC showed that at even higher energies hadrons have a pointlike structure; this opened the way for quantum chromodynamics as the correct theory that describes strong interactions.

However, some work continued in the context of "dual models" and in the mid-1970s several interesting breakthroughs were made.

- It was understood by Neveu, Schwarz, and Ramond how to include space-time fermions in string theory.
- Gliozzi, Scherk, and Olive also understood how to get rid of the omnipresent tachyon. In the process, the constructed theory gained space-time supersymmetry.
- Scherk and Schwarz, and independently Yoneya, proposed that closed string theory, always having a massless spin-2 particle, naturally describes gravity and that the scale α' should be related to the Planck scale. Moreover, the theory can be defined in four dimensions using the Kaluza-Klein idea, namely, considering the extra dimensions to be compact and small.

However, the new big impetus for string theory came in 1984. After a general analysis of gauge and gravitational anomalies [4], it was realized that anomaly-free theories in higher dimensions are very restricted. Green and Schwarz showed in [5] that open superstrings in ten dimensions are anomaly-free if the gauge group is O(32). $E_8 \times E_8$ was also anomaly-free but could not appear in open string theory. In [6] it was shown that another supersymmetric string exists in ten dimensions, a hybrid of the superstring and the bosonic string, which can realize the $E_8 \times E_8$ or O(32) gauge symmetries.

Since the early 1980s, the field of string theory has been continuously developing. There was much heterotic model building in the late 1980s; the matrix model approach to two-dimensional string theory was developed in the early 1990s, followed by the study of stringy black holes. In the mid-1990s, nonperturbative dualities between different supersymmetric string theories were uncovered. This development gave rise to the hope that the theory is unique, and led to the name M-theory. D-branes were discovered and studied. They turned out to be crucial for the construction of controllable models for the identification of black-hole microstates and the microscopic explanation of the black-hole entropy. This, moreover, led to the formulation of AdS/CFT correspondence and its generalizations.

String theory is a continuously evolving subject and this book gives only a brief introduction to some of the best understood topics.

1.4 Conventions

Unless otherwise stated we use natural units in which $\hbar = c = 1$. The string length ℓ_s is kept explicitly throughout the book. It is related to the Regge slope α' and the string (mass) scale M_s by

$$\ell_s = \sqrt{\alpha'} = \frac{1}{M_s}.$$

In the literature, most of the time $\alpha' = 2$ in closed-string theory, $\alpha' = 1/2$ in open-string theory, and sometimes $\alpha' = 1$ in CFT.

The fundamental string tension is $T = \frac{1}{2\pi\ell_s^2}$. We denote by T_p the D_p-brane tension, by T_{M_2,M_5} the respective M_2- and M_5-brane tensions, and by $\tilde{T}_5$ the NS_5-brane tension.

We use X^μ for the space-time coordinates of the string and x^μ for their zero modes. By convention, the left-moving part of the string is the holomorphic part, with conformal dimensions $(\Delta, 0)$. The right-moving part is the antiholomorphic part with dimensions $(0, \bar{\Delta})$. The right-moving part is taken as the nonsupersymmetric side of the heterotic string.

$F_{L,R}$ is the world-sheet (left-moving or right-moving) fermion number. The operator that we use is $(-1)^{F_{L,R}}$. In the NS sector, it counts the number of fermion oscillators modulo 2. Its action is explained in sections 4.12 on page 71 and 7.7.1 on page 174. It should be distinguished from the "space-time fermion number" operators $\mathbf{F}_{L,R}$. $\mathbf{F}_L = 0$ in the left-moving NS sector and 1 in the left-moving Ramond sector. A similar definition holds for $\mathbf{F}_R$. Note that, in the heterotic string, $\mathbf{F}_L$ is indeed the space-time fermion number. In a type-II string the space-time fermion number is $\mathbf{F}_L + \mathbf{F}_R$ modulo 2.

We are using the "mostly plus" convention for the signature of the space-time metric. Our curvature conventions are such that the n-sphere S^n has positive scalar curvature. When there is no risk of confusion, we use for the volume element of the metric $\sqrt{-\det g} \leftrightarrow \sqrt{\det g} \leftrightarrow \sqrt{g}$.

Our conventions on the two-dimensional geometry are spelled out in appendix A, on page 503. Those on differential forms, the ϵ-density, the related E-tensor, and the Hodge dual are found in appendix B on page 505.

We use a unified notation for extended supersymmetry in diverse dimensions. Generically, n supersymmetries in d dimensions are denoted by $\mathcal{N} = n_d$. In two, six, and ten dimensions, because of the existence of Majorana-Weyl spinors, we may define extended supersymmetries with p left-handed and q right-handed MW supercharges. In this case, we use the notation $\mathcal{N} = (p, q)_d$. However, sometimes, even in such dimensions, if the chirality of the supersymmetry is not important for our purposes, we might still use the $\mathcal{N} = n_d$ notation.

For symplectic groups we use the notation Sp(2N) with Sp(2) ~ SU(2). For the groups SO(2n), the subscript $\pm$ on the spinor indicates its eigenvalue $\pm$ 1 under the appropriate generalization of γ^5.

We also call the heterotic string based on the $\mathrm{Spin}(32)/\mathbb{Z}_2$ lattice "the O(32) heterotic string," for simplicity.

Bibliography

The guide to the associated literature presented in this book has been compiled with pedagogy as its main motivation. This book is intended as a textbook, and an appropriately chosen bibliography is a crucial complement. Review articles have been favored here but also original papers when they are deemed to have pedagogical value.

Most of the papers and reviews after 1991 have appeared in the electronic physics archives, are referred to as "arXiv:hep-th/*yymmnnn*," and are available from the central archive site http://arXiv.org/ and its mirrors worldwide.

There are several books and lecture notes on string theory. The first benchmark is that by Green, Schwarz, and Witten (GSW) [7]. It is a reference two-volume set. It summarizes the older literature on string theory, and describes in detail string compactifications up to the mid-1980s. The best and most detailed exposition of the Green-Schwarz approach to the superstring is presented here in detail. There is a balance between covariant and light-cone methods used in the quantization.

The second benchmark is Polchinski's two-volume set [8]. It focuses on the modern approach to string theory via covariant quantization and the use of CFT methods. It also contains a description of D-branes and their roles in nonperturbative string dualities.

There are several other books with different characteristics. Johnson's recent book [9] provides an in-depth look at D-branes and their effects in string theory, while at the same time providing an introduction to the basics. Ortin's book [10] provides a coherent and in-depth exposition of geometric aspects of the theory and its many interesting classical solutions. Szabo's book [11] is a short introduction to string theory (120 pages in all) which covers the very basics. Last but not least is the recent book of Zwiebach [12], which is written at a more introductory level and is addressed to advanced undergraduates. It discusses several interesting subjects in string theory at an accessible level.

There have been several other books and good reviews over the years which are limited by their scope or date of appearance. There is, however, some merit in consulting them since they may have other advantages, like the in-depth description of special subjects. These are [14–21]. We should also note the relatively short but up-to-date introduction to string theory by Danielsson in [21].

Marolf's resource letter [22] is an excellent source of various articles and reviews on string theory. It includes a wide spectrum of sources, from popular science to specialized reviews. The review article of Seiberg and Schwarz [23] provides a useful overview of the field, its achievements, and its goals.