Recent developments in superstring theory

JOHN H. SCHWARZ
California Institute of Technology, Pasadena, CA 91125

Contributed by John H. Schwarz, December 10, 1997

ABSTRACT There have been many remarkable developments in our understanding of superstring theory in the past few years, a period that has been described as “the second superstring revolution.” In particular, what once appeared to be five distinct theories are now recognized to be different manifestations of a single (unique) underlying theory. Some of the evidence for this, based on dualities and the appearance of an eleventh dimension, is presented. Also, a specific proposal for the underlying theory, called “Matrix Theory,” is described. The presentation is intended primarily for the benefit of nonexperts.

This paper presents a brief overview of some of the advances in understanding superstring theory that have been achieved in the last few years. It is aimed at scientists who are not experts in string theory, but who are interested in hearing about recent developments. Where possible, the references cite review papers rather than original sources.

String theories that have a symmetry relating bosons and fermions, called “supersymmetry,” are called “superstring” theories. Until a few years ago, it appeared that there are five distinct consistent superstring theories, each one requiring ten dimensions (nine space and one time) but differing in other respects. It also appeared that they are the only mathematically consistent quantum theories containing gravitation.† It is now clear that they are better viewed as five special points in the manifold (or “moduli space”) of consistent solutions (or “quantum vacua”) of a single underlying theory. Moreover, another special limit corresponds to a sixth consistent quantum vacuum, this one having Lorentz invariance in eleven dimensions (ten space and one time). Even though a fully satisfactory formulation of the underlying theory remains to be completed, it is already clear that this theory is unique—it contains no arbitrary adjustable parameters. This gives a philosophically satisfying picture: there is a unique theory that can give rise to a number of consistent quantum solutions and that contains gravitation. (When there seemed to be five such theories, we found that disturbing.)

The number of quantum solutions is quite unclear at the present time; the number could be very large. At the very least it contains the six solutions mentioned above, and probably many more in which some of the spatial dimensions form a compact manifold, so that the number of apparent dimensions is reduced. The hope, of course, is that solutions with four uncompactified dimensions and other realistic features are sufficiently scarce that the theory is very predictive. However, there is much that still needs to be understood before we can extract reliable detailed predictions. This is not an enterprise for someone who requires a rapid payoff.

The plan for this report is to sketch in Section 1 where things stood after the “first superstring revolution” (1984–1985) and then to describe the recent developments (the “second superstring revolution”) and their implications in the subsequent sections. A detailed survey of the subject, as it was understood 10 years ago, can be found in ref. 1.

Some of the evidence that supports the new picture is reviewed in Section 2. The second superstring revolution is characterized by the discovery of various nonperturbative properties of superstring theory. Some of these are “dualities” that explain the equivalences among the five superstring theories. Another important feature that appears nonperturbatively (i.e., is not apparent when the theory is studied by means of a power series expansion in a coupling constant) is the occurrence of p-dimensional excitations, called p-branes. Their properties are under good mathematical control when they preserve some of the underlying supersymmetry. The maximally supersymmetric p-branes that occur when there are 10 or 11 uncompactified dimensions are surveyed in Section 3. For detailed reviews of the material in Sections 2 and 3, see refs. 2–7.

Section 4 describes how suitably constructed brane configurations can be used to derive, and make more geometrical, some of the nonperturbative properties of supersymmetric gauge theories that have been discovered in recent years. Section 5 presents evidence for the existence of new nongravitational quantum theories in six dimensions. They are likely to play an important auxiliary role in understanding the gravitational theory. Finally, in Section 6, the Matrix Theory proposal, which is a candidate for a nonperturbative description of M theory in a certain class of backgrounds, is sketched. Matrix Theory has been reviewed recently in refs. 8 and 9. It is a rapidly developing subject, which appears likely to be a major focus of research in the next couple of years.

It is not possible in a survey of this size and scope to describe all of the interesting results that have been obtained recently. Among the omitted topics are applications to particle physics phenomenology and to cosmology. Both of these subject areas have seen some progress, but nothing that would appear dramatic to a nonexpert. I hope that these topics will be the main emphasis in some future survey. A recent result that certainly is dramatic, but which will be mentioned only briefly in this survey, is the demonstration that in superstring theory black hole thermodynamics has statistical mechanical underpinnings. Specifically, starting with ref. 10, it has become possible to compute black hole entropy (in a wide variety of cases) by counting microscopic quantum states. For recent reviews of this subject see refs. 11–13.

Abbreviations: vev, vacuum expected value; BFSS, Banks, Fischler, Shenker, and Susskind; BPS, Bogomolny, Prasad, and Sommerfield; EM, electric–magnetic.

†There is a string theory without fermions or supersymmetry, called the “bosonic string”. However, as best we can tell, the only consistent quantum solutions of this theory are ones without gravitation.

‡Since 1991, papers in this field have been posted in the hep-th. archive of the Los Alamos e-print archives. They can be accessed and downloaded easily from http://xxx.lanl.gov. Archive numbers are included in the reference list.
For two other surveys of recent developments in string theory, see refs. 14 and 15.

**Section 1. Historical Setting and Background**

Major advances in understanding of the physical world have been achieved during the past century by focusing on apparent contradictions between well-established theoretical structures. In each case the reconciliation required a better theory, often involving radical new concepts and striking experimental predictions. Four major advances of this type are indicated in Fig. 1 (16). These advances were the discoveries of special relativity, quantum mechanics, general relativity, and quantum field theory. This was quite an achievement for one century, but there is one fundamental contradiction that still needs to be resolved, namely the clash between general relativity and quantum field theory. Many theoretical physicists are convinced that superstring theory will provide the answer.

There are various problems that arise when one attempts to combine general relativity and quantum field theory. The field theorist would point to the breakdown of renormalizability—the fact that short-distance singularities become so severe that the usual methods for dealing with them no longer work. By replacing point-like particles with one-dimensional extended strings, as the fundamental objects, superstring theory certainly overcomes the problem of perturbative nonrenormalizability. A relativist might point to a different set of problems, including the issue of how to understand the causal structure of space-time when the metric has quantum-mechanical excitations. There are also a host of problems associated with black holes such as the fundamental origin of their thermodynamic properties and an apparent loss of quantum coherence. The latter, if true, would imply a breakdown in the basic structure of quantum mechanics. The relativist’s set of issues cannot be addressed properly in a perturbative setup, but the recent discoveries are leading to nonperturbative understandings that should help in addressing them. Most string theorists expect that the theory will provide satisfying resolutions of these problems without any revision in the basic structure of quantum mechanics. Indeed, there are indications that someday quantum mechanics will be viewed as an implication of (or at least a necessary ingredient of) superstring theory.

When a new theoretical edifice is proposed, it is very desirable to identify distinctive testable experimental predictions. In the case of superstring theory there have been no detailed computations of the properties of elementary particles or the structure of the universe that are convincing, though many valiant attempts have been made. In my opinion, success in such enterprises requires a better understanding of the theory than has been achieved as yet. It is very difficult to assess whether this level of understanding is just around the corner or whether it will take many decades and several more revolutions. In the absence of this kind of confirmation, we can point to three general “predictions” of superstring theory that are very encouraging. The first is the existence of gravitation, approximated at low energies by general relativity. No other quantum theory can claim to have done this (and I suspect that no other ever will). The second is the fact that superstring vacua generally include Yang–Mills gauge theories such as those that make up the “standard model” of elementary particles. The third general prediction, not yet confirmed experimentally, is the existence of supersymmetry at low energies (the electroweak scale).

Supersymmetry (at “low energy”) implies that every known elementary particle should have a supersymmetry partner, with a mass that is about 100 to 1000 times the mass of a proton. There is no direct sighting of any of these particles as yet, though there are a number of indirect observational indications that support a belief in their existence. Once any one of them is discovered and identified, it will imply the existence of the rest of them, and set the agenda for experimental particle physics for several decades to come—maybe even make a compelling case for a new improved Superconducting Supercollider (SSC) project. Some of these particles could show up this century at the large electron–positron collider (LEP) machine at the European Center for Nuclear Research (CERN) or the tevatron collider at Fermilab. Otherwise, we must await the large hadron collider (LHC) at CERN (a proton collider with about 40% of the energy that was planned for the SSC), whose completion is scheduled for 2005.

The history of string theory is very fascinating, with many bizarre twists and turns. It has not yet received the attention it deserves from historians of science or popular science writers. Here we will settle for a very quick sketch. The subject arose in the late 1960s in an attempt to describe strong nuclear forces. In 1971 it was discovered that the inclusion of fermions requires world-sheet supersymmetry (17, 18). This led to the development of space-time supersymmetry, which was eventually recognized to be a generic feature of consistent string theories (hence the name “superstrings”). This was a quite active subject for about 5 years, but it encountered serious theoretical difficulties in describing the strong nuclear forces, and quantum chromodynamics (QCD) came along as a convincing theory of the strong interaction. As a result the subject went into decline and was abandoned by all but a few diehards for over a decade. In 1974 two of the diehards (Joel Scherk and J.) proposed that the problems of string theory could be turned into virtues if it were used as a framework for realizing Einstein’s old dream of “unification,” rather than as a theory of hadrons and strong nuclear forces (19). Specifically, we pointed out that it would provide a perturbatively finite theory that incorporates general relativity. One implication of this change in viewpoint was that the characteristic size of a string became the Planck length \( L_{PL} = \left( \frac{G}{c^3} \right)^{1/2} \approx 10^{-33} \text{ cm} \), some 20 orders of magnitude smaller than previously envisaged. This is the natural length scale in a theory that combines gravitation (characterized by Newton’s constant \( G \)) in a relativistic (\( c \) is the speed of light) and quantum mechanical (\( \hbar \) is Planck’s constant divided by \( 2\pi \)) setting. (More refined analyses lead to a string scale \( L_s \) that is about two orders of magnitude larger than the Planck length.) In any case, experiments at existing accelerators cannot resolve distances shorter than about \( 10^{-16} \text{ cm} \), which explains why the point–particle approximation of ordinary quantum field theories is so successful.

In 1984–1985 there were a series of discoveries (20–22) that convinced many theorists that superstring theory is a very promising approach to unification. Almost overnight, the subject was transformed from an intellectual backwater to one of the most active areas of theoretical physics, which it has remained ever since. By the time the dust settled, it was clear that there are five different superstring theories, each requiring ten dimensions (nine space and one time), and that each has a consistent perturbation expansion. The perturbation expansions are power series expansions in powers of a coupling constant (or, equiva-
lently, of Planck’s constant) like those that are customarily used to carry out computations in quantum field theory. The five theories, about which I will say more later, are denoted type I, type IIA, type IIB, $E_8 \times E_8$ heterotic (HE, for short), and $SO(32)$ heterotic (HO, for short). The type II theories have two supersymmetries in the ten-dimensional sense, while the other three have just one. The type I theory is special in that it is based on unoriented open and closed strings, whereas the other four are based on oriented closed strings. Type I strings can break, whereas the other four are unbreakable. The IIA theory is nonchiral (i.e., it is parity conserving), and the other four are chiral (parity violating).

A string’s space-time history is described by functions $X^\mu(\sigma, \tau)$, which map the string’s two-dimensional “world sheet” $(\sigma, \tau)$ into space-time $X^\mu$. There are also other world-sheet fields that describe other degrees of freedom, such as those associated with supersymmetry and gauge symmetries. Surprisingly, classical string theory dynamics is described by a conformally invariant two-dimensional quantum field theory. What distinguishes one-dimensional strings from higher-dimensional analogs is the fact that this two-dimensional theory is renormalizable. [Objects with $p$ dimensions, called “$p$-branes,” have a $(p + 1)$-dimensional world volume. In this language, a string is a 1-brane.] Perturbative quantum string theory can be formulated by the Feynman sum-over-histories method. This amounts to associating a genus $h$ Riemann surface (a closed and orientable two-dimensional manifold with $h$ handles) to an $h$-loop string theory Feynman diagram. The attractive features of this approach are that there is just one diagram at each order $h$ of the perturbation expansion, and that each diagram represents an elegant (though complicated) finite-dimensional integral that is ultraviolet finite. In other words, they do not give rise to the severe short-distance singularities that plague other attempts to incorporate general relativity in a quantum field theory. The main drawback of this approach is that it gives no insight into how to go beyond perturbation theory.

To have a chance of being realistic, the six extra space dimensions must somehow curl up into a tiny geometrical space, whose size is presumably comparable to the string scale $\alpha'$. Though its size in some directions might differ significantly from that in other directions. Because space-time geometry is determined dynamically (as in general relativity) only geometries that satisfy the dynamical equations are allowed. The HE string theory, compactified on a particular kind of six-dimensional space called a Calabi–Yau (CY) manifold, has many qualitative features at low energies that resemble the standard model of elementary particles. In particular, the low-mass amplitudes occur in suitable representations of a plausible unifying gauge group. Moreover these representations occur with a multiplicity whose number is controlled by the topology of the CY manifold. Thus, at least in relatively simple examples, a CY manifold with Euler number equal to $\pm 6$ is required to account for the observed three families of quarks and leptons. These successes have been achieved in a perturbative framework and are necessarily qualitative at best, because nonperturbative phenomena are essential to an understanding of supersymmetry breaking and other important matters of detail.

Section 2. Superstring Dualities and M Theory

The second superstring revolution (1994—?) has brought nonperturbative string physics within reach. The key discoveries were the recognition of amazing and surprising “dualities.” They taught us that what we viewed previously as five theories is in fact five different perturbative expansions of a single underlying theory about five different points in the moduli space of consistent quantum vacua! It is now clear that there is a unique theory, though it may allow many different vacua. For example, a sixth special vacuum involves an eleven-dimensional Minkowski space-time. Another lesson we have learned is that, nonperturbatively, objects of more than one dimension (membranes and higher “$p$-branes”) play a central role. In most respects they appear to be on an equal footing with strings, but there is one big exception: a perturbation expansion cannot be based on $p$-branes with $p > 1$. The reason for this will become clear later.

A schematic representation of the relationship between the five superstring vacua in ten dimensions and the eleven-dimensional vacuum, characterized by eleven-dimensional supergravity at low energy, is given in Fig. 2. The idea is that there is some large moduli space of consistent vacua of a single underlying theory—denoted by M here. The six limiting points, represented as circles, are special in the sense that they are the ones with (super) Poincaré invariance in ten or eleven dimensions. The letters on the edges refer to the type of transformation relating a pair of limiting points. The numbers 16 or 32 refer to the number of unbroken supersymmetries. In ten dimensions the minimal spinor is Majorana–Weyl (MW) and has 16 real components, so the conserved supersymmetry charges (or “supercharges”) correspond to just one MW spinor in three cases (type I, HE, and HO). Type II superstrings have two MW supercharges, with opposite chirality in the IIA case and the same chirality in the IIB case. In eleven dimensions the minimal spinor is Majorana with 32 real components.

Three kinds of dualities, called S, T, and U, have been identified. It can sometimes happen that theory A at strong coupling is equivalent to theory B at weak coupling, in which case they are said to be $S$ dual. Similarly, if theory A compactified on a space of large volume is equivalent to theory B compactified on a space of small volume, then they are called $T$ dual. Combining these ideas, if theory A compactified on a space of large (or small) volume is equivalent to theory B at strong (or weak) coupling, they are called $U$ dual. If theories A and B are the same, then the duality becomes a self-duality, and it can be viewed as a (gauge) symmetry. T duality, unlike S or U duality, holds perturbatively, because the coupling constant is not involved in the transformation, and therefore it was discovered between the string revolutions.

The basic idea of T duality (for a recent discussion see ref. 23) can be illustrated by considering a compact dimension consisting of a circle of radius $R$. In this case there are two kinds of excitations to consider. The first, which is not special to string theory, is due to the quantization of the momentum along the circle. These “Kaluza–Klein” momentum excitations contribute $(n/R)^2$ to the energy squared, where $n$ is an integer. The second kind are winding-mode excitations, which arise due to a closed string getting caught on the topology of space and winding $m$ times around the circular dimension. They are special to string theory, though there are higher-dimensional analogs. Letting $T = (2\pi L^{-1})$ denote the fundamental string tension (energy per unit length), the contribution of a winding mode to the energy squared is $(2\pi m R T)^2$. T duality exchanges these two kinds of excitations by mapping $m \leftrightarrow n$ and $R \leftrightarrow L^{-1}/R$. This is part of an exact map.
between a T-dual pair of theories A and B. One implication is that usual geometric concepts break down at short distances, and classical geometry is replaced by “quantum geometry,” which is described mathematically by two-dimensional conformal field theory. It also suggests a generalization of the Heisenberg uncertainty principle according to which the best possible spatial resolution $\Delta x$ is bounded below not only by the reciprocal of the momentum spread, $\Delta p$, but also by the string size, which grows with energy. This is the best one can do with the fundamental strings. However, by probing with certain nonperturbative objects called D-branes, which will be discussed later, it is possible to do better and measure distances all the way down to the Planck scale.

Two pairs of ten-dimensional superstring theories are T-dual when compactified on a circle: the IIA and IIB theories and the HE and HO theories. The two edges of Fig. 2 labeled $T$ connect vacua related by T duality. For example, if the IIA theory is compactified on a circle of radius $R_A$ leaving nine noncompact dimensions, this is equivalent to compactifying the IIB theory on a circle of radius $R_B = (m/L_A)^{-1}$, where $m = 1/L_A$ is the characteristic string mass scale. The T duality relating the two heterotic theories (HE and HO) is essentially the same, though there are additional technical details in this case. These two dualities reduce the number of (apparently) distinct superstring theories from five to three. The point is that the two members of each pair are continuously connected by varying the compactification radius from 0 to infinity. The compactification radius arises as the vacuum expected value (vev) of a scalar field, which is one of the massless modes of the string, and is a modulus of the theory. Therefore, varying this radius is a motion in the moduli space of quantum vacua.

Suppose now that a pair of theories (A and B) are S dual. This means that if $f_A(g_A)$ denotes any physical observable of theory A and $g_A$ denotes the coupling constant, then there is a corresponding physical observable $f_B(g_B)$ in theory B such that $f_B(g_B) = f_A(1/g_A)$. This duality, whose recognition was the first step in the current revolution (24–28), relates one theory at weak coupling to the other at strong coupling. It generalizes the electric–magnetic symmetry of Maxwell theory. The point is that because the Dirac quantization condition implies that the basic unit of magnetic charge is inversely proportional to the unit of electric charge, their interchange amounts to an inversion of the charge (which is the coupling constant). Electric–magnetic symmetry is broken in nature, because the basic unit of electric charge is much smaller than the self-dual value. It could, however, be a symmetry of the underlying equations. S duality relates the type I theory to the HO theory and the IIB theory to itself. This explains the strong coupling behavior of those three theories. For each of the five superstring theories, the string coupling constant $g_s$ is given by a vev of a massless scalar field $\phi$ called the “dilaton.” (The precise relation is $g_s = e^{\phi_0}$.) Therefore varying the strength of the string coupling also corresponds to a motion in the moduli space of quantum vacua.

The edge connecting the HO vacuum and the type I vacuum is labeled by $S$ in the diagram, because these two vacua are related by S duality. Specifically, denoting the two string coupling constants by $g_{\text{E}_6}$ and $g_{\text{SO}(32)}^{(SO(32))}$, the relation is $g_s \rightarrow g_{\text{E}_6}^{(SO(32))} = 1$. In other words, the vevs of the two dilatons satisfy $(\Phi^{(SO(32))} + \Phi^{(E_6)}) = 0$, and the edge connecting the HO and I points in Fig. 2 represents a continuation from weak coupling to strong coupling in one description, which is weak coupling in the other one. It had been known for a long time that the two vacua have the same gauge symmetry [$SO(32)$] and the same supersymmetry, but it was unclear how they could be equivalent, because type I strings and heterotic strings are very different. It is now understood that heterotic strings appear as nonperturbative excitations in the type I description. The converse is not quite true, because type I strings disintegrate at strong coupling. The link labeled $\Omega$ in Fig. 2 connects the type IIB and type I vacua by an “orientifold projection,” which will not be explained here.

The understanding of how the IIA and HE theories behave at strong coupling, which is by now well established, came as quite a surprise. As discussed in more detail below, in each of these cases there is an eleventh dimension whose size $R$ becomes large at strong string coupling $g_s$, the scaling law being $R \sim e^{\phi_0}$. In the IIA case the eleventh dimension is a circle, whereas in the HE case it is a line interval (so that the eleven-dimensional space-time has two ten-dimensional boundaries). The strong coupling limit of either of these theories gives an eleven-dimensional Minkowski space-time. The eleven-dimensional description of the underlying theory is called “M theory.” As yet, it is less well understood than the five ten-dimensional string theories.

The eleven-dimensional vacuum, including eleven-dimensional supergravity, is characterized by a single scale—the eleven-dimensional Planck scale $m_\text{P}$. It is proportional to $G_5^{-1/2}$, where $G_5$ is the eleven-dimensional Newton constant. The connection to type IIA theory is obtained by taking one of the ten spatial dimensions to be a circle ($S^1$ in the diagram) of radius $R$. Type IIA string theory in ten dimensions has a dimensionless coupling constant $g_s$, which is given by the vev of $e^\phi$, where $\phi$ is the dilaton field—a massless scalar field belonging to the IIA supergravity multiplet. In addition, the IIA theory has a mass scale, $m_\text{v}$, whose square gives the tension of the fundamental IIA string. In the units $h = c = 1$, which are used here, $m_\text{v}$ is the reciprocal of $L_s$, the string length scale introduced earlier. The relationship between the parameters of the eleven-dimensional and IIA descriptions is given by

$$m_\text{v}^2 = R m_\text{P}^3$$

$$g_s = R m_\text{v}$$

Numerical factors (such as $2\pi$) are not important for present purposes and have been dropped. The significance of these equations will emerge later. However, one point can be made immediately. The conventional perturbative analysis of the IIA theory is an expansion in powers of $g_s$ with $m_\text{v}$ fixed. The second relation implies that this is an expansion about $R = 0$, which accounts for the fact that the eleven-dimensional interpretation was not evident in studies of perturbative string theory. The radius $R$ is a modulus—the vev of a massless scalar field with a flat potential. One gets from the IIA point to the eleven-dimensional point by continuing this vev from zero to infinity. This is the meaning of the edge of Fig. 2 labeled $S^1$.

The relationship between the $E_6 \times E_6$ heterotic string vacuum (denoted HE) and eleven dimensions is very similar. The difference is that the compact spatial dimension is a line interval (denoted I in the diagram) instead of a circle. The same relations in Eqs. 1 and 2 apply in this case. This compactification leads to an eleven-dimensional space-time that is a slab with two parallel ten-dimensional faces. One set of $E_6$ gauge fields is confined to each face, whereas the gravitational fields reside in the bulk. One of the important discoveries in the first superstring revolution was the existence of a mechanism that cancels quantum mechanical anomalies in the Yang–Mills gauge symmetry for the special case of $SO(32)$ and $E_6 \times E_6$ gauge groups. There is a nice generalization of the ten-dimensional anomaly cancellation mechanism to this eleven-dimensional setting (29). It works only for $E_6$ gauge groups.

Section 3. p-branes

Supersymmetry algebras with central charges admit “short representations,” the existence of which is crucial for testing conjectured nonperturbative properties of theories that previously were defined only perturbatively. Schematically, when a state carries a central charge $Q$, the supersymmetry algebra implies that its mass is bounded below ($M \geq |Q|$). Moreover, when the state is “BPS saturated,” i.e., $M = |Q|$, the representation theory changes, and a state can belong to a short representation of the supersymmetry algebra. This phenomenon is already familiar for
the case of Poincaré symmetry in four dimensions, which allows a massless photon to have just two polarizations (a short representation), whereas a massive spin one boson must have three polarizations.

This BPS saturation property arises not only for point particles, characterized by a mass $M$, but for extended objects with $p$ spatial dimensions, called $p$-branes. In this case the central charge is a rank $p$ tensor. At first sight, this might seem to be in conflict with the Coleman–Mandula theorem, which forbids finite tensorial central charges. However, the $p$-branes carry a finite charge per unit volume, so that the total charge is infinite for a BPS $p$-brane that is an infinite hyperplane, and there is no contradiction. The BPS saturation condition in this case implies that the tension (or mass per unit volume) of the $p$-brane equals the charge density. Another way of viewing BPS $p$-branes is as solitons that preserve some of the supersymmetry of the underlying theory.

The theories in question (I will focus on the ones with 32 supercharges) are approximated at low energy by supergravity theories that contain various antisymmetric tensor gauge fields. Supercharges are approximated at low energy by supergravity some of the supersymmetry of the underlying theory. The coefficient $\alpha$ derives from the boundary conditions assigned to the ends of the world-volume degrees of freedom required by supersymmetry. Theories in question (I will focus on the ones with 32 supercharges) are approximated at low energy by supergravity theories that contain various antisymmetric tensor gauge fields. They are conveniently represented by differential forms

$$A_\mu = A_{\mu_1 \mu_2} \ldots A_{\mu_p} dx^{\mu_1} \wedge dx^{\mu_2} \wedge \ldots \wedge dx^{\mu_p}. \quad [3]$$

In this notation, the corresponding gauge-invariant field strength is given by an $(n + 1)$-form $F_{p+1} = dA_p$ plus possible additional terms. Fields of this type are a natural generalization of the Maxwell field, which corresponds to the case $n = 1$. A type II or eleven-dimensional supergravity theory with such a gauge field has two kinds of BPS $p$-brane solutions, which preserve one-half of the supersymmetry. One, which can be called “electric,” has $p = n - 1$. The other, called “magnetic,” has $p = D - n - 3$, where $D$ is the space-time dimension (ten or eleven for the cases considered here). As a check, note that for Maxwell theory in four dimensions electric and magnetic excitations are both 0-branes (point particles).

A hyperplane with $p$ spatial dimensions in a space-time with $D - 1$ spatial dimensions can be surrounded by a sphere $S^{D-p-2}$. If $A$ is a $(p + 1)$-form potential for which a $p$-brane is the source, the electric charge $Q_E$ of the $p$-brane is given by a straightforward generalization of Gauss’s law:

$$Q_E \sim \int_{S^{D-p-2}} *F. \quad [4]$$

where $S^{D-p-2}$ is a sphere surrounding the $p$-brane and $*F$ is the Hodge dual of the $(p + 2)$-form field strength $F$. Similarly, a dual $(D - p - 4)$-brane has magnetic charge given by

$$Q_M \sim \int_{S^{p+2}} F. \quad [5]$$

The Dirac quantization condition, for electric and magnetic 0-branes in $D = 4$, has a straightforward generalization to a $p$-brane and a dual $(D - p - 4)$-brane in $D$ dimensions; namely, the product $Q_E Q_M$ is $2\pi$ times an integer.

An approximate description of the classical dynamics of a “thin” $p$-brane is given by an action that is a generalized Nambu-Goto formula $S_p = T_p \cdot V_{p+1} \ldots$. Here $V_{p+1}$ is just the invariant space-time volume of the embedded $p$-brane, generalizing the invariant length of the world-line of a point particle or the area of the world-sheet of a string. The dots represent terms involving other world-volume degrees of freedom required by supersymmetry. The coefficient $T_p$ is the $p$-brane tension—its universal mass per unit volume. Note that for $h = c = 1$ $T_p \sim (\text{mass})^{p+1}$.

Another source of insight into nonperturbative properties of superstring theory has arisen from the study of a special class of $p$-branes called Dirichlet $p$-branes (or D-branes for short). The name derives from the boundary conditions assigned to the ends of open strings. The usual open strings of the type I theory have Neumann boundary conditions at their ends, but T duality implies the existence of dual open strings with Dirichlet boundary conditions in the dimensions that are T-transformed. More generally, in type II theories, one can consider an open string with boundary conditions at the end given by $\sigma = 0$

$$\frac{\partial X^\mu}{\partial \sigma} = 0 \quad \mu = 0, 1, \ldots, p \quad [6]$$

$$X^\mu = X_0^\mu \quad \mu = p + 1, \ldots, 9 \quad [7]$$

and similar boundary conditions at the other end. At first sight this appears to break the Lorentz invariance of the theory, which is paradoxical. The resolution of the paradox is that strings end on a $p$-dimensional dynamical object—a D-brane. D-branes have been studied for a number of years, but their significance was clarified by Polchinski recently (30). They are important because it is possible to study the excitations of the brane using the renormalizable two-dimensional quantum field theory of the open string world sheet instead of the nonrenormalizable world-volume theory of the D-brane itself. In this way it becomes possible to compute nonperturbative phenomena by using perturbative methods! Many (but not all) of the previously identified $p$-branes are D-branes. Others are related to D-branes by duality symmetries so that they can also be brought under mathematical control.

D-branes have found many interesting applications, but the most remarkable of these concerns the study of black holes. Strominger and Vafa (10) (and subsequently many others—for reviews see refs. 11–13) have shown that D-brane techniques can be used to count the quantum microstates associated to classical black hole configurations. The simplest case, which was studied first, is static extremal charged black holes in five dimensions. Strominger and Vafa showed that for large values of the charges the entropy (defined by $S = \log N$, where $N$ is the number of quantum states the system can be in) agrees with the Bekenstein–Hawking prediction (1/4 the area of the event horizon). This result has been generalized to black holes in four dimensions as well as to ones that are near extremal (and radiate correctly) or rotating. In my opinion, this is a truly dramatic advance. It has not yet been proved that black holes do not give rise to a loss of quantum coherence and hence a breakdown of quantum mechanics, as Hawking has suggested, but I expect that result to follow in due course.

Altogether, superstring theories in ten dimensions have three distinct classes of $p$-branes. These are distinguished by how the tension $T_p$ depends on the string coupling constant $g_s$. A “fundamental” $p$-brane has $T_p \sim (m_s)^{p+1}$, with no dependence on $g_s$. Such $p$-branes occur only for $p = 1$—the fundamental strings. Because these are the only objects that survive at $g_s = 0$, they are the only ones that can be used as the fundamental degrees of freedom in a perturbative description. A second class of $p$-branes, called “solitonic,” have $T_p \sim (m_s)^{p+1}/g_s$. These occur only for $p = 5$, the five-branes that are the magnetic duals of the fundamental strings. This dependence on the coupling constant is familiar from field theory. A good example is the mass of an’t Hooft–Polyakov monopole in gauge theory. The third class are the Dirichlet $p$-branes (or D$p$-branes), which have $T_p \sim (m_s)^{p+1}/g_s$. This behavior, intermediate between fundamental and solitonic, was not previously encountered in field theory. In ten-dimensional type II theories D$p$-branes occur for all $p \leq 9$—even values in the IIA case and odd ones in the IIB case. They are all interrelated by T dualities; moreover, the electric–magnetic (EM) dual of a D$p$-brane is a D$p'$-brane with $p' = 6 - p$. D-branes are very important, and so I will have more to say about them later.

Eleven-dimensional supergravity contains a three-form potential $A_3$. Therefore, according to the rules given earlier, there are two basic kinds of $p$-branes—the M2-brane (also known as the “supermembrane”) and the M5-brane. These are EM duals of one another. Because the only parameter of the eleven-
dimensional vacuum is the Planck mass $m_p$, their tensions are determined by dimensional analysis, up to numerical coefficients, to be $T_{M2} = (m_p)^3$ and $T_{M5} = (m_p)^6$.

We can use the relation between the eleven-dimensional theory compactified on a circle of radius $R$ and the IIA theory in ten dimensions to deduce the tensions of certain IIA $p$-branes. Starting with the M2-brane, we can either allow one of its dimensions to wrap the circular dimension, leaving a string in the remaining dimensions, or we can simply embed it in the non-compact dimensions, where it is then still viewed as a 2-brane. In the latter case, the tension remains $m_p$. Using Eqs. 1 and 2, we can recast this as $T = (m_p)^3/g_s$, which we recognize as the tension of the D2-brane of the type IIA theory. On the other hand, the wrapped M2-brane leaves a string of tension $T = m_p/R = m_p^2$. Thus we see that Eq. 1 reflects the fact that, from an eleven-dimensional perspective, a fundamental IIA string is actually a wrapped M2-brane. Starting with the M5-brane, we can carry out analogous calculations. If the M5-brane is not wrapped we obtain a 5-brane of tension $T = m_p^5/m_p^5/g_s$, which is the correct relation for the solitonic 5-brane (usually called the NS5-brane). If the M5-brane is wrapped on the circle, one is left in ten dimensions with a 4-brane of tension $T = m_p^4R = m_p^2$. This has the correct tension to be identified as a D4-brane. In other words, from an eleven-dimensional perspective, the D4-brane is actually a wrapped M5-brane. This fact will prove to be important in the next section.

There are a couple basic facts about D-branes in type II superstring theories that should be pointed out. First, because they can be understood in the weak coupling limit (which makes them heavy) as surfaces on which fundamental type II strings can end, the dynamics of D-branes at weak coupling can be deduced from that of fundamental strings by using perturbative methods. Another basic fact is that because a type II string carries a conserved charge that couples to a two-form potential, the end of a string must carry a point charge, which gives rise to electric flux of a Maxwell field. This implies that the world-volume theory of a D-brane contains a $U(1)$ gauge field. In fact, for strong fields that vary slowly the world-volume theory of the D-brane is actually a nonlinear theory of the Born–Infeld type. The $U(1)$ gauge field can be regarded as arising as the lowest excitation mode of an open string with both of its ends attached to the D-brane.

Consider now $k$ parallel $D_p$-branes, which are well approximated by $(p + 1)$-dimensional hyperplanes in ten-dimensional space-time. In this case, the two ends of an open string can be attached to two different branes. The lowest mode of a string connecting the $i$th and $j$th D-brane is a gauge field that carries $i$ and $j$ type electric charges at the corresponding ends. These gauge fields, together with the ones associated with the individual branes, fill out the $k^2$ states in the adjoint representation of a $U(k)$ group, and give rise to a $U(k)$ gauge theory in $p + 1$ dimensions. Classically, this gauge theory can be constructed as the dimensional reduction of $U(k)$ super Yang–Mills theory in ten dimensions. The separations of the D-branes are given by the vevs of scalar fields, which break the $U(k)$ gauge group to a subgroup by the Higgs mechanism. This is one of many examples in which a mechanism that once appeared to be just a mathematical abstraction has acquired a concrete physical/geometric realization. For $p = 3$, these gauge theories have a straightforward quantum interpretation, but for $p > 3$ the gauge theories are nonrenormalizable. I will return to this issue in Section 5.

Section 4. Brane-Configuration Constructions of Supersymmetric Gauge Theories

In the last section we saw that a collection of $k$ parallel D-branes gives a supersymmetric $U(k)$ gauge theory. The unbroken supersymmetry in this case is maximal (16 conserved supercharges). In this section I describe more complicated brane configurations, which break additional supersymmetries, and give supersymmetric gauge theories in four dimensions with a richer structure. This is an active subject, which can be approached in several different ways. Here I will set for two examples in one particular approach. (For a different approach see ref. 31.)

The first example (32) is a configuration of NS5-branes and D4-branes in type IIA theory depicted in Fig. 3. This configuration gives rise to an $SU(N_c)$ gauge theory in four dimensions with $N = 2$ supersymmetry (8 conserved supercharges). To explain why, one must first describe the geometry. All of the branes are embedded in ten dimensions so as to completely fill the dimensions that will be identified as the four-dimensional space-time with coordinates $x^6, x^7, x^8, x^9$. In addition, the NS5-branes also fill the $x^6$ and $x^9$ dimensions, which are represented by the vertical direction in the figure, and they have fixed values of $x^6, x^7, x^8, x^9$. The D4-branes, on the other hand, have a specified extension in the $x^6$ direction, depicted horizontally in the figure, and they have fixed values of $x^6, x^7, x^8, x^9$. The idea is that the gauge theory lives on the $N_c$ D4-branes, which are suspended between the NS5-branes. The $x^6$ extension of these D4-branes becomes negligible for energies $E < \ll 1/L$, where $L$ is the separation between the NS5-branes. In this limit the five-dimensional theory on the D4-branes is effectively four dimensional. In addition, there are $N_f$ semi-infinite D4-branes, which result in $N_f$ hypermultiplets (supersymmetry multiplets that contain only scalar and spinor fields) belonging to the fundamental representation of the $SU(N_c)$ gauge group. These states arise as the lowest modes of open strings connecting the two types of D4-branes depicted in the figure. The presence of the NS5-branes is responsible for breaking the supersymmetry from $N = 4$ to $N = 2$.

This picture is valid at weak coupling, because the gauge coupling constant $g_{YM}$ is given by $g_{YM}^2 = g_s/(L_m)$, and the IIA picture is valid for small $g_s$. Substituting Eq. 2, we see that $g_{YM}^2 = R/L$, where $R$ is the radius of a circular eleventh dimension. So far, the description of the geometry omits consideration of this eleventh dimension, but by taking it into account we can see what happens to the gauge theory when $g_{YM}$ is not small and quantum effects become important. The key step is to recall that a D4-brane is actually an M5-brane wrapped around the circular eleventh dimension. Thus, reinterpreted as a brane configuration embedded in eleven dimensions, the entire brane configuration corresponds to a single smooth M5-brane! The junctions are now smoothed out in a way that can be made quite explicit. The correct configuration is one that is a stable static solution of the M5-brane equation of motion, which degenerates to the IIA configuration described in the limit $R \to 0$. There is a simple method, based on complex analysis, for finding such solutions. If some of the dimensions of the embedding space are described as a complex manifold, with a specific choice of complex structure, then the brane configuration is a stable static solution if some of its spatial dimensions are described as a complex manifold that is embedded holomorphically. In the example at hand, the relevant dimensions are two dimensions of the M5-brane, which are embedded in the four spatial dimensions denoted $x^6, x^7, x^8, x^{10}$, where $x^{10}$ is the circular eleventh dimension. A complex structure is specified by choosing as holomorphic coordinates $v = x^6 + ix^{10}$ and $t = \exp[i(x^6 + ix^{10})/R]$, which is single-valued. Then a holomorphically em-

![Fig. 3. Brane configuration for an $N = 2$ 4d gauge theory.](image-url)
bedded submanifold of one complex dimension (or two real dimensions) is specified by a holomorphic equation of the form $F(t, v) = 0$. The appropriate choice of $F$ is a polynomial in $t$ and $v$ with coefficients that correspond in a simple way to the positions of the NS5-branes and D4-branes. (For further details see ref. 32.) This two-dimensional submanifold is precisely the Seiberg–Witten Riemann surface (or “curve”) that characterizes the exact nonperturbative low-energy effective action of the gauge theory. When first discovered (33, 34), this curve was introduced as an auxiliary mathematical construct with no evident geometric significance. We now see that the Seiberg–Witten solution to the $SU(N_C)$ gauge theory with $N = 2$ supersymmetry and $N_F$ fundamental representation hypermultiplets is encoded in an M5-brane with four of its six dimensions giving the space time and the other two giving the Seiberg–Witten curve! This simple picture makes the exact nonperturbative low energy quantum physics of a wide class of $N = 2$ gauge theories almost trivial to work out by entirely classical reasoning. The construction can be generalized to various other gauge groups and representations by considering more elaborate brane configurations. A class of examples of special interest are theories that have superconformal symmetry. Such theories are free from the usual ultraviolet divergences of quantum field theory.

The brane configuration described above can be modified to describe certain $N = 1$ supersymmetric gauge theories. One way to achieve this is to rotate one of the two NS5-branes so that it fills the dimensions $x^1, x^2$ and has fixed $x^3, x^6$ coordinates. When this is done the $N_C$ D4-branes running between the NS5-branes are forced to be coincident. The rotation breaks the supersymmetry from $N = 2$ to $N = 1$. One of the remarkable discoveries of Seiberg is that an $N = 1$ supersymmetric gauge theory with gauge group $SU(N_C)$ and $N_F \geq N_C$ flavors is equivalent in the infrared to an $SU(N_C - N_F)$ gauge theory with a certain matter content (35). This duality can be realized geometrically in the brane configuration picture by smoothly deforming the picture so as to move one NS 5-brane to the other side of the other one (36–38). Such a move certainly changes the exact quantum vacuum described by the configuration. However, the parameters involved are irrelevant in the infrared limit, so one achieves a simple understanding of Seiberg duality.

Section 5. New Nongravitational Six-Dimensional Quantum Theories

We have seen that it is interesting and worthwhile to consider the world volume theory of a collection of coincident or nearly coincident branes. For such a theory to be regarded in isolation in a consistent way, it is necessary to define a limit in which the brane degrees of freedom decouple from those of the surrounding space-time “bulk.” Such a limit was implicitly involved in the discussion of the preceding section. (This involves some subtleties, which I did not address.) In this section I wish to consider the six-dimensional world-volume theory that lives on a set of (nearly) coincident 5-branes. If one can define a limit in which the degrees of freedom of the world-volume theory decouple from those of the bulk, but still remain self-interacting, then we will have defined a consistent nontrivial six-dimensional quantum theory (39). (The only assumption that underlies this is that M theory/superstring theory is a well-defined quantum theory.) The six-dimensional quantum theories that are obtained this way do not contain gravity. The existence of consistent quantum theories without gravity in dimensions greater than four came as quite a surprise to many people.

As a first example consider k parallel M5-branes embedded in flat eleven-dimensional space-time. This neglects their effect on the geometry, which is consistent in the limit that will be considered. The only parameters are the eleven-dimensional Planck mass $m_p$ and the brane separations $L_i$. In eleven dimensions an M2-brane is allowed to terminate on an M5-brane. Therefore, a pair of M5 branes can have an M2-brane connect them. When the separation $L_i$ becomes small, this M2-brane is well approximated by a string of tension $T_s = L_s/m_p^2$. The limit that gives decoupling of the bulk degrees of freedom is $m_p \rightarrow \infty$. By letting the separations approach zero at the same time, this limit can be carried out holding the string tensions $T_s$ fixed. In the limit one obtains a chiral six-dimensional quantum theory with $(2,0)$ supersymmetry containing $k$ massless tensor supermultiplets and a spectrum of strings with tensions $T_s$. There are five massless scalars associated to each brane (parametrizing their transverse excitations). They are coordinates for the moduli space of the resulting theory, which is $(R^4)^k/S_k$. The permutation group $S_k$ is due to quantum statistics for identical branes. String tensions depend on position in moduli space, and specific ones approach zero at its singularities.

A closely related construction is to consider k parallel NS 5-branes in the IIA theory. The difference in this case is that one of the transverse directions (parametrized by one of the five scalars) is the circular eleventh dimension. In carrying out the decoupling limit one can send the radius R to zero at the same time, holding the fundamental type IIA string tension $T = m_s^2 = m_p^2 R$ fixed. The resulting decoupled six-dimensional theory contains this string in addition to the ones described above. It becomes bound to the NS5-branes in the limit, as the amplitude to come free vanishes in the limit $g_s \rightarrow 0$. The resulting theory has the moduli space $(R^4 \times S^2)/S_k$. This theory contains fundamental strings and has a chiral extended supersymmetry, features that are analogous to type IIB superstring theory in ten dimensions. However, it is actually a class of nongravitational theories (labeled by $k$) in six dimensions. Because of the analogy some authors refer to this class of theories as iib string theories. Six-dimensional nongravitational analogs of type IIA string theory, denoted iia string theories, are obtained by means of a similar decoupling limit applied to a set of parallel NS5-branes in IIB theory. These iia and iib string theories are related by T duality. Explicitly, compactifying one spatial dimension on a circle of radius $R_0$ or $R_s$, the theories (with given $k$) become equivalent for the identification $m_p^2 R_0 R_s = 1$. This feature is directly inherited from the corresponding property of the IIA and IIB theories.

There are various generalizations of these theories that will not be described here. There are also six-dimensional nongravitational counterparts of the two ten-dimensional heterotic theories. These have chiral $(1,0)$ supersymmetry. In the notation of Fig. 2, they could be referred to as he and ho theories. They, too, are related by T duality. Although the constructions make us confident about the existence and certain general properties of these theories, they are not very well understood. The ten-dimensional string theories have been studied for many years, whereas these six-dimensional string theories are only beginning to be analyzed. Like their ten-dimensional counterparts, the fact that they have T dualities implies that they are not conventional quantum field theories.

Section 6. The Matrix Theory Proposal

The discovery of string dualities and the connection to eleven dimensions has taught us a great deal about nonperturbative properties of superstring theories, but it does not constitute a complete nonperturbative formulation of the theory. In October 1996, Banks, Fischler, Shenker, and Susskind (BFSS) made a specific conjecture for a complete nonperturbative definition of the theory in eleven uncompactified dimensions called “Matrix Theory” (9, 40). In this approach, as we will see, other compactification geometries are only beginning to be analyzed. Like their ten-dimensional counterparts, the fact that they have T dualities implies that they are not conventional quantum field theories.
One of the p-branes that has not been discussed yet is the D0-brane of type IIA theory in ten dimensions. Being a D-brane, its mass is \( M = m_0/g_s \). Using Eq. 2, one sees that \( M = 1/R \), which means that it can be understood as the first Kaluza–Klein excitation of the eleven-dimensional supergravity multiplet on the circular eleventh dimension. In fact, this is a good way of understanding (and remembering) Eq. 2. Like all the type II D-branes it is a BPS state that preserves half of the supersymmetry, so one has good mathematical control. From the eleven-dimensional viewpoint it can be viewed as a wave going around the eleventh dimension with a single quantum of momentum. Higher Kaluza–Klein excitations with \( M = n/R \) are also BPS states. From the IIA viewpoint these are bound states of N D0-branes with zero binding energy.

By the prescription given in Section 3, the dynamics of N D0-branes is described by the dimensional reduction of \( U(N) \) super Yang–Mills theory in ten dimensions to one time dimension only. When this is done, the spatial coordinates of the super Yang–Mills theory in ten dimensions to one time dimension

\[ \frac{N}{5} \]

letting D0-branes in the infinite momentum frame (IMF). This entails \( N \) D0-branes with zero binding energy.

The BFSS conjecture is that this IMF frame

\[ \text{N} \]

is the circular eleventh dimension. In fact, this is a good way of understanding (and remembering) Eq. 2. This means that it can be understood as the first Kaluza–Klein dimensional viewpoint it can be viewed as a wave going around the circular eleventh dimension. In fact, this is a good way of understanding (and remembering) Eq. 2. This means that it can be understood as the first Kaluza–Klein way of describing the BFSS conjecture.

The techniques involved here are reminiscent of those developed in connection with the parton model of hadrons in the late 1960s. The BFSS conjecture is that this IMF frame \( N \rightarrow \infty \) limit of the D0-brane system constitutes an exact nonperturbative description of the eleven-dimensional quantum theory. The \( N \rightarrow \infty \) limit is awkward, to say the least, for testing this conjecture. A stronger version of the conjecture, due to Susskind, is applicable to finite \( N \).

It asserts that the IMF D0-brane system, with fixed \( N \), provides an exact nonperturbative description of the eleven-dimensional theory compactified on a light-like circle with \( N \) units of (null) momentum along the circle.

One of the first issues to be addressed was how this conjecture should be generalized when additional dimensions are compact, specifically if they form an \( n \)-torus. The reason that this is a nontrivial problem is that open strings connecting pairs of D-branes can lie along many topologically distinct geodesics. It turns out that all these modes can be taken into account very elegantly by replacing the one-dimensional quantum theory of the D0-branes by an \( (n+1) \)-dimensional quantum theory, where the \( n \) spatial dimensions lie on the dual torus. The extra dimensions precisely account for all the possible stretched open strings. This picture has had some immediate successes. For example, it nicely accounts for all the duality symmetries for various values of \( n \). However, \( (n+1) \)-dimensional super Yang–Mills theory is nonrenormalizable for \( n > 3 \), so this description of the theory is certainly incomplete in those cases. The new theories described in Section 4 provide natural candidates when \( n = 4 \) or \( 5 \), but when \( n > 5 \) there are no theories of this type, and so we seem to be stuck (42, 47). One of its intriguing features is that it seems to give “noncommutative geometry” (48) a natural home in string theory (49).

Section 7. Concluding Remarks

Perturbative superstring theory was largely understood in the 1980s. Now we are rapidly learning about many of its remarkable nonperturbative properties. In particular, Matrix Theory is a very interesting proposal for defining M theory/superstring theory nonperturbatively. Whether it is precisely correct, or needs to be modified, is very much up in the air at the present time. However, even if it is right, it does not seem to be useful for defining vacua with more than five compact dimensions. This fact is very intriguing, because this is precisely what is required to describe the world that we observe. It may be that a somewhat different approach, such as the one offered recently in ref. 50, is required.

Despite all the progress that has taken place in our understanding of superstring theory, there are many important questions whose answers are still unknown. In fact, it is not even clear how many more important discoveries still remain to be made before it will be possible to answer the ultimate question that we are striving to answer—Why does the universe behave the way it does? Short of that, we have some other pretty big questions: What is the best way to formulate the theory? How and why is supersymmetry broken? Why is the cosmological constant so small (or zero)? How is a realistic vacuum chosen? What are the cosmological implications of the hypothesis that inflationary predictions can we make? I remain optimistic that we are closing in on the correct theory and that the coming decades will bring progress on some of these challenging questions.

This work was supported in part by the U.S. Department of Energy under Grant DE-FG03-92ER40701.